WP2 Developing the initial model
D2.2 Report on the legal analysis

Final version – 29 February 2016

Funded by the Horizon 2020
Framework Programme of the
European Union

D2.2 Report on the legal analysis

Work package	WP2 Developing the initial model
Lead author	Helena Ursic (LEIDEN)
Contributing author(s)	Bart Custers (LEIDEN), Michel Olmedo (ROOTER)
Due date	M13 (February 2016)
Date	29 February 2016
Version	1.0

Type	Report
Dissemination level	Public

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 645244. This document does reflect the authors view only. The European Commission is not responsible for any use that may be made of the information the document contains.

Document history

Version	Date	Author(s)	Notes
0.1	26 January 2016	Helena Ursic (LEIDEN)	First consolidated draft
0.2	29 January 2016	Helena Ursic (LEIDEN)	Consolidated draft – second version
0.3	3 February 2016	Michel Olmedo (ROOTER), Bart Custers (LEIDEN)	Version ready for internal review
0.4	26 February 2016	Helena Ursic (LEIDEN)	Finalisation
1.0	29 February 2016	Daniel Bachlechner (FRAUNHOFER)	Final check

EuDEco in a nutshell

EuDEco assists European science and industry in understanding and exploiting the potentials of data reuse in the context of big and open data. The aim is to establish a self-sustaining data market and thereby increase the competitiveness of Europe. To be able to extract the benefits of data reuse, it is crucial to know how to understand the underlying economic, societal, legal, and technological framework conditions and challenges to build useful applications and services. Despite the amount of activities in this domain, an effort is missing to develop use cases and business models that are economically viable, legally certain and taking societal needs and concerns into account. EuDEco will accomplish this by leveraging the engagement of other projects conducting pilots on data reuse as well as by the engagement of external experts and stakeholders. EuDEco moves beyond the classical approaches by applying the approach of complex adaptive systems to model the data economy in order to identify value networks, use cases and business models for data reuse. In the course of the project we develop and refine the data economy model in several steps further by case studies on previous pilots on data reuse, by in-depth analysis from legal, socio-economic and technological points of view, and by extensive tests of use cases and business models with other projects. Therefore, it will analyse framework conditions relevant and challenges related to data reuse and the emergence of a self-sustaining data market. Finally, EuDEco will deliver a model of the data economy including viable use cases and business models as well as suggestions and recommendations addressing the main legal, contractual, societal and technological concerns and challenges such as contractual framework or data protection. Above that, EuDEco will develop an observatory for policy makers enabling them to track the development of the data economy.

Disclaimer

Table of contents

List of tables

List of abbreviations

APEC	Asian Pacific Economic Cooperation
API	Application programming interface
BCR	Binding corporate rules
CAS	Complex adaptive system
CC	Creative Commons
CCTV	Closed-circuit television
CJEU	Court of Justice of the EU
DPA	Data protection authority

EC	European Commission
EDPS	European Data Protection Supervisor
EU	European Union
GDPR	The General Data Protection Regulation
ICO	Information Commissioner's Office
ICT	Information and communication technology
IoT	Internet of Things
IP	Intellectual property
IPR	Intellectual property right
ISO	International Organization for Standardization
IT	Information technology
LAPSI	The European Thematic Network on Legal Aspects of PSI
NIS	Network and Information Security
OECD	Organisation for Economic Co-operation and Development
PSI	Public sector information
SaaS	Software as a Service
SME	Small and medium-sized enterprises
TRIPS	Agreement on Trade-Related Aspects of Intellectual Property Rights
TTIP	The Transatlantic Trade and Investment Partnership
UGC	User generated content
UK	United Kingdom
US	United States
WIPO	The World Intellectual Property Organization

Executive summary

D2.2 comprises the in-depth legal analysis of the initial heuristic model, focused on addressing, in more detail, the main legal concerns for data reusers in the European data economy. This detailed analysis of the legal propositions presented in D2.1 is supplemented by an analysis of the technological, societal and economical propositions included in D2.1.

The first step in the analysis is to clarify the two main cornerstones of the project: the suitability of EU legislation as the basis for the legal study and the adequacy of the complex adaptive system (CAS) methodology for a research focused on data reuse. The suitability of EU legislation as basic legal framework for the project comes from various reasons, but mainly due to the focus of the project being set on the European data economy and the characteristics and the elements that define this legislation, in terms of uniformity and interpretation, which result in a mostly homogeneous legal framework. The adequacy of the CAS approach to EuDEco is based on the fact that both the data economy and the law itself can be perceived as a CAS, thus generating the need of a detailed analysis on each of the actors that are part of the CAS to understand their interrelations and obtain solid conclusions.

The second step of the analysis contains a detailed overview of each of the legal disciplines that affect data reuse from an EU perspective, but including some notes on the international perspectives on the matter, when available. Each of the sections includes a table with the main pieces of legislation of each legal discipline, a detailed analysis of its implications for data reuse and a checklist highlighting these implications. They are supplemented with another set of tables that illustrate the relation between each legal area of knowledge and the propositions provided by the technical, societal and economical experts of EuDEco.

The section on data protection rights begins this step of the analysis with a description of what should be understood as “personal data” and the legal obligations that tie the hands of those who deal with this kind of data, which is the case of many data reusers. A description of all the roles that are present in the data management chain, from data subjects to data reusers are as well depicted and explained in this section. This section also includes a detailed and extended list of the data protection principles that data reusers will have to keep in mind when dealing with personal data in the EU and the rights of data subjects regarding their data.

The section on human rights follows this first section and provides a thorough overview of how these fundamental rights relate to data in general and data reuse in particular. The focus is set on the right to privacy and the right to non-discrimination, as they are the ones that most directly relate to data. Other human rights that must be borne in mind by data reusers can be the right to liberty, justice or dignity. Considering that these rights are provided for on all human rights conventions and treaties applicable within the EU, it is key for data reusers to respect them.

The next section covers the interaction between data reuse and intellectual property rights. In this regard, the main forms of intellectual property are addressed in a sub-section of their own (copyrights, trademarks, patents and databases). Furthermore, this analysis is complemented by an overview of other related figures such as unfair competition and trade secrets. This section relates to the technological and societal perspectives in the most direct way, as it includes restrictions and limitations of use in relation to data, but also in relation to the software, algorithms and brands that are used by data users and reusers alike, helping to shape business models and boost (or discourage) development.

Contracting is the main concern of the following section and it is quite closely related to intellectual property, in the sense that also influences the relation between data holders, users and reusers, thus being instrumental in the development of data reuse-based business models. The most direct consequence of the need to establish contracts (or licenses) in order to have access to the data or to use a computer program leads to existing of power struggles between those who own the means and tools and those who need them, making it imperative to be aware of the contracting rules that are in place in the EU.

The last sections cover Public Sector Information and Cybersecurity, which complete the circle of related legal areas. These sections focus on non-personal information gathered by public bodies and the security measures that will have to be implemented to comply with the proposed Network and Information Security Directive.

The Deliverable ends with a set of conclusions drawn from the analysis of the sections as a whole. It is considered that a flexible environment, where this complex legal framework can adapt to the reality of the market, will be key in developing a sustainable and compliant data economy.

1 Introduction

1.1 Purpose and scope

The objective of the D2.2 is to provide a thorough, clear and concise overview of the legal requirements, aimed exclusively at the actors in the European data economy. Getting a solid understanding of legal limitations will help us guarantee the legality of the EuDEco model. The deliverable is based on the Task 2.2, which involves the following three steps: First, the legal framework from D1.2 will be concretized by exploring the requirements and compliance issues more in depth, in particular those that the previous deliverables identified as onerous, e.g. the principle of purpose limitation. Whereas D1.2 already thoroughly analysed the relevant legal landscape, the added value of D2.2 will be a concise approach, which better suits the needs of the data economy model. Second, we will carefully examine relevant technological and socio-economic issues as outlined in D2.1 that might have been neglected in our previous legal analyses. The results will help us indicate which legal requirements or areas of law may require further investigation. Moreover, this will help us better align with the work in other working streams and contribute to the model in a more meaningful and practical way. Third, this deliverable will also consider possible future adaptations in the law and assess the impact of the dynamics in the legal system on the data economy model. European businesses face challenges due to the rapidly changing regulatory environment, with numerous legislative proposals and amendments to the existing law. Our model will only be able to offer a meaningful guidance if it manages to encompass that changing legal reality.

1.2 Structure of the document

This deliverable is split into 11 chapters. Chapter 1 gives an introduction to our work and explains our scientific approach. Chapter 2 defines the subject matter, Chapter 3 explains why our analysis is based on the EU law and Chapter 4 describes the essentials of the complex adaptive systems (CAS) theory and how it is incorporated in our report. Chapter 5 to 10 address the areas of data protection law, privacy and human rights regulation, intellectual property law, big data contracting, public sector regulation and cybersecurity law, respectively. Chapter 11 concludes.

1.3 Relationships to other deliverables

The deliverable is based upon D2.1, D1.3 and D1.2. D1.2 provides an overview of the legal framework, D1.3 lists a number of use cases, while D2.1 describes the legal perspective of the first heuristic model. D2.2 builds on the previous research efforts, however, by comparing legal propositions with the technological and socio-economic framework and describing those relations through the lens of the CAS theory, it makes a step forward to a more detailed data economy model. An important improvement in D2.2 is that it also considers international regulations and assesses their impact on the EU. Furthermore, D2.2 traces all the major legal developments on the EU level and updates the framework for the data economy accordingly.

1.4 Methodology

Since the subject matter of this research deals with issues that have their basis amongst others in law, economy, sociology and technology, it is necessary to adopt a multi-disciplinary approach when it comes to attaining the research goals. This in turn establishes the need for deploying various and complementary research methods.

1. The first and most essential method is systematic *desk research of available scientific literature.* This thorough analysis is based on databases and journals available via public institutions and online open sources. The focus is on the recent scholarship, whereas less recent papers are also considered if widely recognized as fundamental literature. The literature analysis is systematic, which means it will base on a pre-defined approach consisting of reliable sources from the world's most widely-used databases.[1] When the report deals with recent issues that have not yet been addressed in a published contribution, a limited number of non-scientific and journalistic sources is taken in consideration. The focus is on well-reputed web blogs[2], world's leading newspapers[3] and news portals[4].
2. *The case-law analysis* is conducted with the help of the official European Union (EU) law database[5] and the published jurisprudence available via the Court of Justice of the EU (CJEU) database[6]. Occasionally, the case law analysis will depart from the chosen standpoint and draw some attention to national specifics that might have an impact on the EU policy. Due to strong co-dependency between the EU regulation and Member States' national laws, this is unavoidable.
3. *The case study* method is only used to a limited extend, mainly to draw attention to some practical dilemmas.
4. Last but not least, the *comparative analysis* between technological, socio-economic and legal propositions is used throughout the deliverable to draw parallels between the three frameworks and to identify ways in which they interact. The CAS theory is used as a frame for this analysis of relations and interdependencies within the system.

[1] The current selection of the sources and databases includes Leiden University Catalogue, Social Science Research Network, Elsevier's Scopus and Web of Science by Thomson Reuters.
[2] E.g. Hunton Privacy Blog (https://www.huntonprivacyblog.com/), Fieldfisher Privacy Blog (http://privacylawblog.fieldfisher.com/), IAPP Privacy Perspectives (https://iapp.org/news/privacy-perspectives).
[3] E.g. The Economist, The Guardian, The New York Times, The Financial Times.
[4] E.g. BBC.com, CNN.com, Volkskrant.nl.
[5] http://eur-lex.europa.eu/homepage.html
[6] http://curia.europa.eu/jcms/jcms/j_6/

2 Subject matter

A prerequisite for analysing the data reuse landscape from the legal point of view is addressing the question 'what is data reuse'.

In her book about the value of data in the research domain, Christine L. Borgman admits that the greatest difficulty in assessing practices for data reuse is the lack of agreement on what constitutes reuse. In turn, reuse depends on what is meant by use of data or other forms of information. Information seeking needs and uses are long-standing and thorny problems in information science. No satisfactory definition of information use applies across disciplines so the lack of agreement on use or reuse of data is unsurprising.[7]

In Europe, data reuse was defined in Article 2 of The Directive 2003/98/EC on the re-use of public sector information (PSI Directive)[8] as "the use by persons or legal entities of documents held by public sector bodies, for commercial or non-commercial purposes other than the initial purpose within the public task for which the documents were produced." This definition only applies to public sector information and may not be appropriate in the private domain, although it may serve as a foundation. However, given the myriad of business models and the ever-evolving data-driven economy, the definition from the PSI directive may be too scarce and may not encompasses all different types in which data reuse can manifest.

Mayer-Schönberger and Cukier consider data reuse in the private sector and see it as the primary way to unleash big data's potential value.[9] Schneider defines data reuse as secondary use of data that follows initial collection and use. In his opinion data reuse is what bothers citizens most, when they think about their personal data.[10]

Based on the above, we can conclude that the term data reuse in its broadest sense suggests that there is initial (primary) use of data and subsequent (secondary) use of data, i.e., the reuse of data. However, data reuse does not necessarily need to be perceived as a two-step process. It can be split into more categories, for instance, by distinguishing data recycling, data repurposing and data recontextualisation based on the proximity between the reuse's objective and the initial purpose of data use.[11]

[7] CL Borgman, *Big Data, Little Data, No Data*: *Scholarship in the Networked World* (MIT Press: 2015), p. 214.
[7] V Mayer-Schönberger and K Cukier, *Big Data: A Revolution That Will Transform How We Live, Work and Think* (New York: Houghton Mifflin Harcout 2013).
[8] [2003] OJ L 345/90
[9] *Ibidem*.
[10] B Schneier, Risks of data reuse, Schneider on security - online blog, June 28, 2007. <https://www.schneier.com/blog/archives/2007/06/risks_of_data_r.html> accessed 23 January 2016.
[11] BHM Custers and H Ursic, 'Big data and data reuse: a taxonomy of data reuse for balancing big data benefits and personal data protection' (2016) *International Data Privacy Law*, forthcoming.

Data Recycling is the most simplified form of data reuse. An actor is using the same data in the same way for more than once. A typical example may be a health insurance company that collects patient data in order to have a proper client database used for billing the insurance premiums that are due and to reimburse medicines, treatments and therapies. When they use a client's address for sending them a bill, they will do this monthly, quarterly or annually. In that sense, they periodically reuse the address more than once for the same purpose.

In case the same insurance company starts using the data to assess risks of patients in order to determine risk-based insurance premiums (e.g., higher premiums for people at risk or showing unhealthy behaviour like smoking, not exercising, etc. and lower premiums for people at low risks showing healthy behaviour),[12] they are reusing the data for a different purpose. This is a form of data reuse that we classify as *data repurposing*.[13] For data repurposing typically stricter legal rules apply, e.g., a requirement to obtain an (additional) informed consent of the data subjects.

When the health insurance company in the examples above starts selling the data, other companies may also make use of the data, for instance, for marketing their products to particular target groups. This is a form of data reuse that we classify as *data recontextualisation*. In such cases data can be reused in a (sometimes completely) different context. This may cause issues of contextual integrity, since data may have a different meaning or may be interpreted differently in another context.[14] This is where the legal discussion often transforms into ethical considerations.

It is also possible to distinguish data reuse from a data subject's perspective. The sub-categories that we recognise are data sharing, data portability and data blocking (the right to be forgotten). In the data economy, where personal data is increasingly used as a commodity and individuals are often degraded to the level of an object, understanding the ways in which they might impact data reuse, seems to become increasingly useful.

3 The justification for choosing the EU law as the basis for our legal analysis

There are several reasons why the EU law was chosen as the basis of our legal analysis.

[12] For more detail on risk profiling, see, for instance, M Hildebrandt and S Gutwirth, *Profiling the European Citizen* (Heidelberg: Springer, 2008); BE Harcourt*, Against Prediction: Profiling, Policing and Punishing in an Actuarial Age* (Chicago: University of Chicago Press, 2007); F Schauer, *Profiles, Probabilities and Stereotypes* (Cambridge MA: Harvard University Press, 2003); TZ Zarsky, 'Mine Your Own Business!' (2003) 5 Yale Journal of Law and Technology 57; B Custers, *The Power of Knowledge* (Nijmegen: Wolf Legal Publishers, 2004).

[13] Loshin makes the distinction between data reuse and data repurposing, see D Loshin, *The Practitioner's Guide to Data Quality Improvement* (Burlington MA: Morgan Kaufmann OMG Press, 2011).

[14] See, for instance, H Nissenbaum, 'Privacy as Contextual Integrity' (2004) 79 Washington Law Review 1, 119-158.

Firstly, this project examines the European data economy with a special aim to create a model that reflects the reality as well as indicate the future directions. In the process of creating the model, the EuDEco project partners have been using the CAS approach. This method will be also employed to describe the legal framework. As has been noticed by some authors who wrote about the law and CAS, the goal of a theory of law's complexity is not to work around the complexity of the legal system, but to immerse lawyers and legal institutions in it by making the invisible hands of law visible.[15] In other words, the point of the theory is not to search for differences in the systems and examine national specifics, but to understand the high-level components and the forces that guide the development and the application of the law. We follow this line of reasoning by limiting our scope on the EU law, which, as we believe, justly reflects legal trends and driving forces in the EU society.

Secondly, although the EU is a union of sovereign states with their own national laws, the rules on the EU level are common to all Member States and act as a reflection of the EU consensus on adequate legal standards. This approach has been confirmed by the doctrine of direct and indirect effect of the EU legislation. Through this doctrine, the CJEU has established that the EU regulations are directly applicable and should be interpreted coherently throughout the union.[16] EU directives do not require such unification, however, the CJEU has often emphasized the importance of the interpretation of national law in accordance with the EU rules.[17] Under limited conditions, it has even allowed their direct effect. [18]

Thirdly, and most importantly, while it is clear that the European market is legally, economically and culturally fragmented, the general perception and political tendency is to see it as a single market. It is believed that more harmonised legal provisions would also grant more protection to the citizens as well as reduce the administration cost for the European businesses. The proposed data protection regulation, which advocates unified standards and more collaboration between the Member States, supports this idea.

4 The CAS approach

To explain the current and the future direction in the European data economy, EuDEco uses the CAS theory. The CAS theory studies the systems comprised of a macroscopic, heterogeneous set of autonomous agents interacting and adapting in response to one another and to external environment inputs. It emerged primarily from the physical sciences in the 1980s and later spread to economics, ecology, sociology and law. One of the pioneers in the application of the CAS theory to legal systems, J. B. Ruhl, believes that the legal system is also one social world in which invisible hands, similar to those

[15] JB Ruhl, 'Law's Complexity - A Primer', FSU College of Law, Public Law Research Paper No. 313.
[16] *Flaminio Costa v E.N.E.L.,* C-6/64, 15 July 1964.
[17] See for example *Melloni*, C-399/11, 26 February 2013
[18] See for example *Francovich v Italy*, *C-6/90, 19 November 1990* for the vertical direct effect, and *Kücükdeveci* v Swedex *GmbH & Co KG*, C-555/07, 19 February 2010 for the horizontal direct effect.

known from the economic or social science, are at work. By making a parallel to the natural science, he showed that all the ingredients and properties of complex adaptive systems also exist in legal systems. [19]

There is no standard definition of a complex adaptive system, however, based on the CAS literature, it is possible to extract a number of typical features, which can, if taken together, serve as a solid description of the system.[20]

Table 1 CAS theory principles

CAS theory principles	**Explanation**
Emergence and aggregation (system property)	as system scope grows, system behavior emerges from the aggregation of network causal chains which cannot be explained by examining any isolated part of the system
Self-organized structure (system property)	as system scale grows, the system tends to organize around a set of deep structural rules that lend stability to system behavior, by establishing the levels of influence that each agent will have in the system
Adaptive resistance and resilience capacity (system property)	as a result of these internal behaviors, the system as a whole proves resistant to environmental perturbations and resilient at returning to or near its self-organized critical state following a perturbation
Adaptive resistance and resilience capacity (system property)	the agents interact with and adapt to each other according to deterministic rules (illusion of free will)
Nonlinear relationships (agent property)	the agent interaction rules do not produce behavior that is in continuously proportionate relationships over time; sharp tipping points and discontinuities frequently occur
Critical states (system property)	notwithstanding deep stable structure traits, dynamic qualities of the system (nonlinear relationships, network feedback) lean toward change at the "surface" of the system, so that the system evolves under a "stable disequilibrium" set of behaviors, sometimes near or "on the edge of" the chaotic
Phase transitions (system property)	if pushed too far from its self-organized critical state, however, either by a massive perturbation or by constant pressure from less severe perturbations, the CAS could "tip" in a nonlinear and potentially irreversible move into a new set of behaviours
Path dependence (system property)	the next state of the system depends on the information that has flowed through the system in all prior states
Network connectivity of feedback (agent property)	there is high connectivity, or feedback, between agents, parts, and scales of the system, creating a network of nodes and channels through which information (energy, money, food) flows

[19] *Supra* 15.
[20] See for instance *supra* 15.

Power law event distributions (system property)	the distribution of the "size" of events in the system does not exhibit a binomial normal distribution, but rather takes on asymptotic properties with many "small" events and very few "large" events
Heterogeneity (agent property)	complex adaptive systems consist of a number of different classes of autonomous agents

EuDEco understands the European data economy as a CAS. The law can be seen either as a framework within this larger CAS, or a CAS itself. This may quickly overcomplicate the relationships between the two. For the need of this deliverable we perceive law as the framework and the European data economy as the CAS trapped within the boundaries of legal requirements.[21] The fact that the law is only seen as a framework and not as a CAS itself does not diminish the level of its complexity and adaptivity.

Our discourse on the complexities in law will be centred around the list of leading questions proposed by Ruhl[22]:

- What patterns exist in the distribution and organization of legal systems?
- Are these patterns uniquely determined by local conditions or are they historically and spatially contingent?
- How do legal systems become assembled over social time?
- How does evolution shape legal system properties?
- What are the relationships between legal system structure and functioning?
- Does evolution of legal systems increase resiliency or lead to criticality? Does it lead to the edge of chaos?

These CAS aspects will be addressed briefly at the end of each chapter. Instead of going into detail of each characteristic, we will only use them as a frame for the analysis of relations and interdependencies within the system.

[21] Z Kunbei and AHJ Schmidt, 'Thinking of data protection law's subject matter as a complex adaptive system: A heuristic display', 31 Computer Law & Security Review 2, pp. 201-220.
[22] *Supra* 15.

5 Data protection requirements and their role in the model

5.1 Data protection – overview of the legal framework

Table 2 Overview of the EU data protection law

EU DATA PROTECTION LAW	
Primary EU law	European Convention on Human Rights (Article 8) Charter of the fundamental rights of the EU (Arts. 7 and 8) Treaty on the functioning of the EU (Article 16)
Secondary EU law	Personal Data Protection Directive Regulation concerning the protection of individuals with regard to the processing of personal data by Community institutions and bodies E-privacy Directive General Data Protection Regulation (draft proposal from December 16, 2015)
National legislation	28 Member States' national legislations

This section focuses on the provision of the Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (Data Protection Directive or, shortly, the DPD). As the directive encompasses the constitutional requirements set forth in the EU Charter of fundamental rights and European Council's Convention 108[23], they will not be examined in further detail. Similarly, we will conduct no systematic analysis of the EU national laws.[24] However, we will use them occasionally as a reference to illustrate some practical issues.[25]

The EU legislator is currently discussing a draft EU data protection regulation, which will replace the DPD. The legislative process is running to an end and it is expected that the regulation will be adopted in

[23] Convention for the protection of individuals with regard to automatic processing of personal data, Council of Europe, 1981.
[24] See the reasons explained above.
[25] German law has been one of the most strict data protection regulation, in particular as regard s data reuse. For example, the principle of purpose specification has been interpreted very restrictively in Germany. <http://ec.europa.eu/justice/policies/privacy/docs/studies/new_privacy_challenges/final_report_country_report_A4_germany.pdf> accessed 23 January 2016.

2016. After that it will take another two years for it to be enforced. New rules will therefore be used only from 2018 on. The new regulation is based on the existing directive, which means that the system of data protection will not be modified tremendously. That said, there are still a couple of changes that will require special attention of data reusers. Where applicable, we refer to them by using the latest publicly available draft of the data protection regulation.[26]

5.1.1 What does "personal data" stand for?

Most companies that base their business model on data reuse have exhibited particular interest in personal data. While this type of data is easy to monetize, e.g., in marketing purposes,[27] it is also subjected to strict regulations[28] and often triggers a lively ethical debate.

The European data economy includes a myriad of different business models, many of which employ personal data processing. The latter may vary in intensity and scope. Some of the players on the market only perform personal data analytics as a side activity,[29] while others use the data throughout their value chain and are closely involved in its collection, storage and transfers.[30] Hence, different entities will have different obligations towards the data. What is common to everyone is that the usage of personal data will always trigger the applicability of data protection regulations. As the first step, they will therefore need to get a solid understanding of the concept of personal data.

The definition of personal data according to the DPD (Article 2) contains four main elements:

- "any information"
- "relating to"
- "an identified or identifiable"
- "natural person"

Each of the elements can be further explained. It is beyond the scope of this deliverable to explore the definition in more detail.[31] Generally speaking, the EU data protection authorities tend to adopt a wide

[26] Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM (2012) 11 final, 25 January 2012. The latest unofficial version of the GDPR proposal available <http://static.ow.ly/docs/Regulation_consolidated_text_EN_47uW.pdf> accessed 23 January 2016.

[27] FJ Zuiderveen Borgesius, Personal Data Processing for Behavioural Targeting: Which Legal Basis? (November 26, 2015) International Data Privacy Law.

[28] See for example P De Filippi , Big data, big responsibilities (2014) 3 *Internet Policy Review* 1.

[29] For example, retail sector can greatly benefit from the big data analytics, especially if it is used within the supply chain process to boost profitability and to "shape consumer price perception". McKinsey, CMAC - Creating competitive advantage from big data, June 2012.

[30] Acxiom Corporation is a marketing technology and services company with offices in the United States, Europe, Asia, and South America. Acxiom offers marketing and information management services, including multichannel marketing, addressable advertising, and database management. Acxiom collects, analyses, and parses customer and business information for clients, helping them to target advertising campaigns, score leads, and more. <http://www.acxiom.com/about-acxiom/> accessed 23 January 2016.

[31] See Article 29 Working Party (2007), Opinion 4/2007 on the concept of personal data (No. WP 136).

definition of personal data. For instance, many of them consider IP addresses personal data, although this might seem at odds with the definition of personal data set forth in the DPD.[32] It has been argued, however, that the broader interpretation is indispensable to adequately respond the challenge of extensive online data collection and processing, particularly in relation to behaviour advertising.[33] It is also important to note that Member States are free to opt for a more benevolent legislation e.g., they can extend the scope of data protection to also include legal entities.

At this point, we would like to stress the concept of identifiability and its relevance for the data-driven economy, particularly due to the increasing importance of big data analytics. While identifiable data can be processed in all phases of a big data analytics cycle, in practice this does not always happen, since companies may try to circumvent data protection rules by anonymising their data and processing only non-identifiable information.[34] Bypassing the data protection rules can be problematic for two reasons: first, an absolute anonymization is never possible,[35] and second, even anonymised data could have some undesirable consequences for personal privacy.[36]

Some categories of personal data require special protection. These are the so-called sensitive data, which include, for instance, information about a person's ethnical origins, religion, health or political views. As a general rule, processing of data related to health is prohibited, however the DPD allows for certain exceptions. For the processing of health data, the explicit consent of the data subject is the most relevant exception.[37] Furthermore, Article 8(4) allows Member States to lay down, for reasons of substantial public interest, exemptions in addition to those mentioned above either by national law or by decision of the supervisory authority. This provision may legitimize the processing of sensitive data for the purposes of medical scientific research and government statistics provided that all these exceptions are in line with the safeguards set forth in the DPD.[38]

The concept of sensitive data has faced some criticism. It has been argued that the group of characteristics that fall under the definition is chosen arbitrary and that many other types of data can also reveal very sensitive information about an individual.[39] Moreover, in the data economy, longitudinal

[32] For a critical analysis of the issue see Zwenne, Diluted Privacy Law, inaugural lecture, April 2013, p. 5. <https://zwenneblog.weblog.leidenuniv.nl/files/2013/09/G-J.-Zwenne-Diluted-Privacy-Law-inaugural-lecture-Leiden-12-April-2013-ENG.pdf> accessed 23 January 2016.
[33] FJ Zuiderveen Borgesius, 'Online Price Discrimination and Data Protection Law' (August 28, 2015) Amsterdam Law School Research Paper No. 2015-32.
[34] M Oostveen, Working Paper on Big Data, presented at the NILG PhD Forum in Amsterdam, November 2015.
[35] P Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization', U of Colorado Law Legal Studies Research Paper No. 9-12.
[36] B Custers: *The Power of Knowledge* (Wolf Legal Publishers, 2004). We will further address those shortcomings in Chapter 6.
[37] N Purtova, E Kosta and BJ Koops, 'Laws and Regulation for Digital Health' in SA Fricker, C Thuemmier and A Gavras (Eds) Requirements Engineering for Digital Health (Springer, 2014), pp. 47-75.
[38] Article 29 Working Party (2007), Opinion 4/2007 on the concept of personal data (No. WP136).
[39] Report of the CEPS digital forum, April 2013, p. 42.<http://www.ivir.nl/publicaties/download/1350> accessed 23 January 2016.

and combined data sets – also called a comprehensive digital identity – play a significant role. This data is not especially protected, although it can be very revealing about individual circumstances.[40]

The data protection regulation proposal indicates that the direction of negotiations for the final text is going towards an expansive definition of personal data capturing cookies, IP addresses, web beacons and other tracking technologies when used to track an individual.[41] The proposal also introduces a number of new categories of sensitive personal data, e.g. genetic data.[42]

International perspective:

There are some important differences in the definition of personal data in the EU and in the US. A notable example is the concept of financial data, which is listed under sensitive data in the US, yet not in Europe.[43] *Another striking issue is the conflicting approach to privacy in public space, for instance in the social media in the US, as soon as personal information is shared publicly, it is no more protected (third party doctrine).*[44] *In the EU, however, this information will remain under the shelter of personal data protection law.*

Table 3 Compliance checklist for data protection law (a)

Compliance checklist
• Do you process any personal data (personal data is any information about an identified or identifiable individual)? It can be as little as a name or contact detail and information can be easily linked to other information in your possession in order to identify an individual. • If you do process personal data, which categories of data subjects will have their personal data processed – employees, third parties, website users, costumers etc.? Bear in mind that different rules might apply to different groups of data subjects. • What type of personal data will be processed? Does this data relate to health, labour relations, political opinions, racial origin, criminal history, sexuality, biometric or genetic characteristics? Processing personal data with these specific characteristics normally triggers stricter requirements.

[40] *Ibidem*.
[41] Olswang, EU Data Protection Reform: Where are we – and what can you do to prepare? (n.d.)
[42] *Ibidem*.
[43] Although the perception of an average citizen is that his or her financial data is indeed sensitive.
[44] Prof. Francesca Bignami, Professor of Law, George Washington University, Washington DC, presentation at the NILG conference, 13 November 2015, Amsterdam.

5.1.2 Which reusers should comply with the EU data protection law?

At the outset, it needs to be mentioned that the DPD does not cover all data processing operations. Most noticeably, reusing criminal databases for the purposes of public security and processing of data by a natural person in the course of purely personal or household activities fall outside the boundaries of the EU law (e.g. reusing the data collected by a CCTV that someone installs in his own apartment).

As for the rest, the DPD provides two legal bases for its application in the EU Member States. According to Article 4 EU law should apply whenever

a) a data controller is established in the EU and processes personal data in the context of the activities of the establishment (e.g., an Austrian SMEs collects the data of their costumers) and
b) a non-EU data controller uses its equipment in an EU Member State without being established on the European territory (e.g., an Australian company sells cars with a GPS system, which is connected to the company's servers in Australia).

All data reusers based in the EU will thus have to adhere to the EU rules and in some cases the same legal framework will also apply to reusers that are located outside the EU. Which national law applies, will be determined by the location of the processing.[45]

It should be noted that in the landmark case *Google v. Spain*, the Court of Justice of the EU (CJEU) interpreted the "establishment of a controller" in the way that also covers its subsidiary/branch, even though the latter takes no processing decisions but only contributes to the controller's commercial activity (which, however, constitutes data use and reuse). The recent case law thus tends to expand the scope of EU law. Even more directly, the recent decision of the French DPA (CNIL) advocates a global applicability of the right to be forgotten.[46] While the idea of the global implication of data protection law was rejected by the CJEU in the landmark decision *Lindquist,* there is still no clear boundaries set for the EU data protection rules.[47]

Also important to consider, the proposed data protection regulation extends the applicability requirements to all entities that offer services and goods to the citizens located in the EU or that monitor their behaviour in the EU (Article 3 of the proposal). The exact wording and the effects of the new law remain to be seen, but it has been suggested that the regulation will most probably increase the number of organisations that fall within the scope of data protection law. Factors such as offering languages and currencies generally used in one or more Member States with the possibility of ordering

[45] Article 29 Working Party (2010), Opinion 8/2010 on applicable law for additional guidance (No. WP179).
[46] News published on the CNIL's website interpreting the CJEU's decision in Google v Spain <http://www.cnil.fr/english/news-and-events/news/article/right-to-delisting-google-informal-appeal-rejected/ > accessed 23 January 2016.
[47] See for example Kuner, Extraterritoriality and International Data Transfers in EU Data Protection Law, Oxford working papers, August 2015.

in that language, or mentioning customers in the EU would make it more likely that the controller will be held to be offering goods or services to the EU citizens.[48]

International perspective:

The way in which the EU law is applicable has some direct consequences for the global data economy. In the recent years, the EU tried to mitigate the global reach of the Internet and its impact on citizens' privacy by extending the scope of data protection law. It is possible that the upcoming Transatlantic Trade and Investment Agreement (TTIP) will also have some consequences to the controllers and processors globally, but the final outcome remains to be seen.

Table 4 Compliance checklist for data protection law (b)

Compliance checklist
• Is the controller established in the EU ("being established" refers to very diverse types of activities and any presence of a controller in the EU, even its subsidiary/branch which takes no processing decisions but only contributes to the controller's commercial activity)? • Does the controller use equipment in the EU (e.g. an Australian company sells cars with a GPS system, which is connected to the company's servers in Australia)?

5.1.3 Reusers as controllers and processors – what are their obligations?

The EU law splits those that process data into two big groups. Based on the level of their autonomy in relation to data processing, they are considered either controllers or processors. The line between the two groups is thin and, since the data economy is known for its diversity, it will not always be clear who is a controller and who is a processor.[49] The final decision should be based on the actual relation and not on a (potential) contract. In most cases, however, reusers will be considered data controllers and therefore they will be subject to all obligations provided by data protection laws.[50]

A data controller is a 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 (Article 2 of the DPD).

The main obligations/duties of a data controller are the following:

- Adherence to privacy principles

[48] *Supra* 41, p. 4.
[49] <http://www.huntonfiles.com/files/webupload/CIPL_Safe_Harbor_3.08.pdf> accessed 23 January 2016.
[50] LAPSI. (2012). *Policy Recommendation N. 4: Privacy and Personal Data Protection*.

- Observance of data subject rights, which include the rights of the data subject to information, access, rectification, erasure and blocking, and to object to the processing of personal data are all framed in such a way as to create obligations for the controller.[51]
- Adherence to information requirements (concretised by a privacy notice)
- Security obligations
- Notification requirement

In the current data economy, there are often more controllers involved in the same data processing. If they genuinely share their obligations, they should be considered joint controllers. An example of joint controllers are online publishers and ad-network providers who closely cooperate in order to target online consumers with relevant ads and customize their advertisements with the help of personal data. The status of joint controllers also implies that they carry the same responsibilities and liabilities.

Although "data controller" has an independent meaning under the EU law, its identification from a data protection perspective will be interconnected in practice with the civil, administrative or criminal law rules providing for the allocation of responsibilities or sanctions to which a legal or a natural person can be subject.[52]

A data processor is a natural or legal person, public authority, agency or any other body, which processes personal data on behalf of the controller (Article 2 of the DPD). In terms of data reuse, a processor would typically be an external entity specialized in data analysis to which the controller outsources certain tasks. In the rapidly changing digital economy, we can notice that many times the relationship between the controller and processor shifts on the side of the former. For example, a European SME as controller is much weaker than a global processor such as Dropbox. The negotiating power will be on the side of the latter, which can affect the freedom of contracting in the mandate between the controller and processor.

The main obligations/duties of a data processor are the following (Article 17 of the DPD):

- Security in line with the national law that applies to a processor
- Technical and organisational measures in order to protect the data in line with the contract signed with the data controller

If the processor engages a sub-processor, the latter should adhere to the same requirements.

[51] Article 29 Working Party (2010), Opinion 1/2010 on the concepts of "controller" and "processor" (No. WP 169).
[52] *Idem*.

International perspective:

The concepts of data controller and processors are inherent to the EU law and are unknown, for instance, in the US legal system. The differences in the legal terminology have to be handled carefully when using the instruments that enable international data transfers such as contractual clauses or safe harbour.[53]

Table 5 Compliance checklist for data protection law (c)

Compliance checklist
• Do you identify yourself as data controller (i.e. the one that determines the purposes and means of the processing of personal data)? If you do, you need to adhere to the requirements listed above. • Do you identify yourself as data processor (i.e. the one that processes data on behalf of another person or company who is considered controller)? If you do, this means you will also have some responsibilities under data protection law and you should consider the relevant requirements carefully. • Are you processing data together with another party? You should keep in mind that you may be a joint controller. If this is the case, you may share legal responsibility with the co-controllers.

5.1.4 Data protection principles that every reuser has to observe

Data privacy principles serve as guidance for data controllers and processors to handle personal data in a legitimate and responsible way. In practical terms, it is highly recommended to implement privacy principles into business processes to make sure they are observed always when personal data is processed. Privacy principles aim to establish boundaries to data processing and are designed to offer a balanced approach. This seems to be of great significance for data-intense business models where data reuse is part of everyday processes.

Lawful processing

A legal basis is the initial and critical point of every data processing. The DPD recognizes as valid the following five options (Article 7):

(a) the data subject has unambiguously given his consent to process his or her data; or
(b) processing is necessary for the performance of 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; or

[53] *Supra* 44.

processing is necessary for compliance with a legal obligation to which the controller is subject; or

(c) processing is necessary in order to protect the vital interests of the data subject; or

(d) processing 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 or in a third party to whom the data are disclosed; or

(e) processing is necessary for the purposes of the legitimate interests pursued by the controller or by the third party or parties to whom the data are disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject.

In cases of data reuse, many of the legal bases above will hardly be used in practice. For instance, it is difficult to imagine how data reuse could be a requirement to perform a contract (b). As Article 29 Working party notices, Article 7 (b) must be interpreted strictly and 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. For example, the provision is not a suitable legal ground for building a profile of the user's tastes and lifestyle choices based on a clickstream on a website and the items purchased. This is because the data controller has not been contracted to carry out profiling, but rather to deliver particular goods and services.[54]

Also, it is very unlikely that reuse would be legitimated by a data subject's vital interest or general public interest. Both provisions suggest that they have limited application. First, the phrase 'vital interest' appears to limit the application of this ground to questions of life and death. Second, the general public interest refers to public tasks that are assigned to an official authority or that are imposed on a private part by a public body.

Thus, a consent (a) or a data controller's legitimate interest (f) will probably be used as a legal basis. But even then, it will not always be easy for a commercial reuser to justify a processing. A valid secondary consent is difficult to receive, especially when a considerable amount of time has passed since the initial consent was gained. Legitimate interest of a commercial performer (probably closely related to its business goals) will suffice if it outweighs the importance of the right to data protection. As stressed by the researchers in the LAPSI project, data protection is considered a fundamental right, hence a reuser will usually have a hard time proving that its interest wins over privacy.

Article 29 WP takes a more balanced approach in its opinion. It states that when interpreting the scope of Article 7 (f), it is necessary to ensure the flexibility for data controllers for situations where there is no undue impact on data subjects. However, it is important that data subjects are provided with sufficient legal certainty and guarantees which prevents misuses of this open-ended provision.[55]

[54] See *supra* 41.

[55] Article 29 Working Party (2014), Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC (No. WP 217), p. 10.

Although the WP provided an extended guidance to interpretation of Article 7(f), the dilemma whether commercial organisations are capable of performing the balancing task, remains. Following the landmark case *Google v Spain*, Google was required to start processing requests for the right to be forgotten. In those cases, a balance between two fundamental rights, freedom of expression and right to privacy has to be found. Google has been arguing this should not be its task as Google is an objective party with no preference for either solution. A similar challenge can be seen in the case of 7 (f). However, there is an important distinction. In those cases, a commercial party would indeed have a preference for one specific output of the balancing test. Thus, it will often be challenging to find a balanced solution.

Fair processing

Article 10 of the DPD guarantees the right to information to all data subjects whose data is being processed. This information has to be given in an intelligible form including the details of purposes of processing, the categories of data concerned, the data undergoing processing, the recipients or categories of recipients to whom the data are disclosed, and any available information about the source and logic involved in any automatic processing of data. It is the data controllers' and processors' responsibility to ensure data subjects have actually received all necessary information. They usually fulfil the requirement by making use of their privacy policies or statements that are publicly available, most often on the Internet. Providing the information in a clear and transparent way is understood by policy makers as a reflection of the concept of "fair processing".[56]

When personal data is transferred to a third party and reused, the right to information, which is derived from the openness principle, does not cease to apply. On the contrary, at this point it becomes even more important that the data subject is fully informed about the activities in which he or she is indirectly involved (through his/her own data). There are two options how to ensure data subject's awareness. First, data reuse activities that might happen in the future are described and communicated to the data subject before personal data is collected. Second, the data subject renews consent every time before the data is reused for a new purpose, based on the information communicated through the updated privacy policy. Both tactics prove to be difficult to apply. In the first case, it is hard to predict all the purposes for data reuse that may appear in the future. In the second case, it is almost impossible to get in touch with all data subjects and to secure their valid consent.[57]

[56] Ustaran et al., European privacy, Law and Practice for Data protection professionals, p. 106.
[57] See for example B Schermer, BHM Custers and S van der Hof, 'The Crisis of Consent: How Stronger Legal Protection May Lead to Weaker Consent in Data Protection' (2014) 6 Ethics and Information Technology 3. The authors establish that privacy policies nowadays contain information overload, absent a meaningful choice for the users which leads to the situation where data subjects no more make informed decision but simply consent whenever they are asked to do so. Not only are data subjects unaware how their information and under what conditions will be processed, they also lack some basic understanding of whether their data can be and will be reused.

The aim of the principle of fairness is to establish the right balance between the right of the individual to have control over his or her data and the flexibility required for businesses to develop and innovate and make best use of the vast amount of data generated online and offline. As pointed out by the UK Information Commissioner's Office (ICO), the processing of big data can challenge the principle of fairness, which is closely related to the reasonable expectations of privacy that data subjects may have.[58] These expectations can be exceeded by data reuse. An example would be the purchase of data from a social media provider by a data broker. A user may not be aware how his or her data is shared nor may he or she expect such a trade. It has been proposed that a well-designed and workable mechanism for opt-out could mitigate the tensions in cases where the legitimate interests of a controller and the interests of data subjects are at stake.[59]

Conveying adequate information to an individual not only indicates fairness of processing, but it is also an indispensable source of transparency and individual involvement. Only after receiving clear information the data subject is able to invoke his or her core rights such as right to access, erase and object.

Purpose specification

The purpose specification is probably the most significant principle for data reusers. It states that the purposes for which personal data are collected should be specified and the data may only be used for these purposes.[60] In other words, the data can only be used for a purpose which is compatible with the one for which it was collected.

Controllers have to determine these purposes before the processing of data starts.[61] The chosen legal basis will only be valid for one specific purpose. For instance, consent will only be valid in cases of data use and reuse as they were communicated by the controller at the moment when the data was collected.

In practice, it is unlikely that all possible reuses can be defined or predicted in advance. This can be frustrating for data reusers, as they might feel that the possibilities in which they can exploit the collected data have been disproportionately restricted. Admittedly, a wide range of reuse activities can be covered by choosing an open formulation of the initial purpose. However, this can be seen as circumventing the intention of the legislator and processing based on it can be considered illegitimate.

The principle of purpose specification leads to the biggest challenges for data reusers. At the same time, it is a defence against excessive data use, profiling and analytics. Given the increasing prevalence of

[58] ICO, Big data and data protection 20140728 Version: 1.0 <https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdf > accessed 23 January 2016.
[59] 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' from 19 November 2014.
[60] Article 6(b) of the DPD.
[61] Article 29 Working Party (2013), Opinion 03/2013 on purpose limitation, WP 203, p. 15.

these practices, data protection could be sought much more in regulating the decision-making stage than in regulating the data collection and data processing stages.[62]

New regulation seems to shine some light in that direction (Article 6, para. 3a). The Council's amendments have improved the provision on purpose limitation with a more detailed guidance for data reusers. In case of further processing the judgement whether the processing is compatible should be based on the following criteria: (a) any link between the purposes for which the data have been collected and the purposes of the intended further processing; (b) the context in which the data have been collected; (c) the nature of the personal data; (d) the possible consequences of the intended further processing for data subjects; (e) the existence of appropriate safeguards.[63]

Careful observance of the purpose limitation has been stressed by the EDPS as one of the key decisions of accountable organisations. Also, the EDPS emphasized the importance of the context, in which data is reused. Reusers have to consider whether data initially used in one context can be legitimately used in another context.[64]

Data quality

This principle requires data controllers to observe that data remains relevant, not excessive in relation to the purpose and kept no longer than necessary for the processing. The latter requirement is known as principle of data minimisation, and it has been increasingly contested by economic developments and social practices. Most obviously, data minimisation seems at odds with an information-rich society, which collects vast amounts of data because they might prove useful in the future.[65] In some sectors, such as medicine or pharma, using a vast amount of data is critical. To overcome this challenge, the ICO advises the reusers to carefully explain the goal of data accumulation and to use anonymization techniques as much as possible.[66] By no means the regulators think that data minimisation is not applicable in the big data world. On the contrary, the majority of them has already made clear that this is a principle that should remain unchanged.

Security and confidentiality of personal data

The DPD also gives some guidelines in terms of security and confidentiality of data. Precautions should be taken against risks of loss, unauthorized access, destruction etc. of personal data.[67]

[62] Bert-Jaap Koops, 'The trouble with European data protection law', (2014) 4 International Data Protection Law 4, pp. 250-261.
[63] *Supra* 26 (Article 6, para. 3a).
[64] *Supra* 59, p. 16.
[65] *Supra* 39, p.45.
[66] ICO, Big Data & Data Protection, 2014. <https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdf> accessed 23 January 2016.
[67] The upcoming NIS directive will impose additional security-related requirements. See Chapter 10 for more detail.

International perspective:

The principles of data protection are based on the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108), which was adopted by the Council of Europe in 1981. This treaty is still valid,[68] *although it has been superseded by later covenants. Nevertheless, the principle of fair processing of personal data from the convention has remained the basis of the EU data protection directive as well as the national laws. Also many other countries around the world follow the same principles. Contractual mechanisms, such as contractual clauses, that are used for the transfers of data between the EU and third states, reflect those common ideas.*

Table 6 Compliance checklist for data protection law (d)

Compliance checklist
• Regularly check whether the personal data being processed is accurate and kept up-to-date. • Make sure that you have a good overview of how the personal data is stored (hard copy file, local database, regional database, global database). • Be aware of where the databases are located. • Monitor how long the data is retained and how it is destroyed. • Are you making any new use of data that you own or that you receive from elsewhere? If yes, check the nature of the data and consider possible legal restrictions to its reuse. • It is recommended that there is a privacy policy/notice in place, which covers data processing that you intend to undertake. • On what legal basis do you process data? If you use consent forms, maintain a registry.

5.1.5 Data subject rights and data reuse

Working alongside the data controllers' obligations to notify the relevant data protection authority and to provide data subjects with certain information are the data subjects' rights[69]:

- The right to access
- The right to object
- The right to rectification, erasure and blocking of the data
- The right not to be subjected to solely automated decisions

[68] It currently has 47 ratifications.
[69] Edoardo Ustaran et al., *European Privacy: Law and Practice for Data Protection Professionals* (International Association of Privacy Professionals, 2012), p. 125.

Data subject rights are designed to place a data subject in a more equal position towards data controllers and processors. As Levin notes, in the face of technological developments those rights are becoming increasingly important.[70]

However, according to Koops, technological developments are at the same time causing that data rights have been diluting. In his opinion, the exercise of data subject rights is highly theoretical: *"Yes, you can be informed, if you know where to look and how to read (but who knows, looks, and reads?). Yes, you can request controllers to let you know what data they process, if you know that you have such a right in the first place (but which controller really understands and seriously complies with all such requests, particularly if exercised on an above-incidental scale?). Yes, you can request correction or erasure, if you know whom to ask (but how are you ever going to reach everyone in the chain, or mosaic, or swamp, of interconnected data processing?). There are simply too many ifs and buts to make data subject rights meaningful in practice.*"

Koops continues by asserting that not even the firmest believers in informational self-determination can claim that they actually know which of their data are being processed in what ways by data controllers, or that they have effective control on most data-processing operations they are subjected to.[71]

Particularly in a data economy where data is reused frequently, these rights are often difficult to invoke. This does not mean, however, that they do not apply. As data reusers are also considered data controllers, they should abide with the same data protection obligations, including the provisions on data subject rights – right to access to data, right to rectification, right to object to processing, right to erase/block the data.[72]

5.1.5.1 The right to access

By invoking the right to access a data subject gets the information whether or not his or her personal data are being processed. This information must be given in an intelligible form and needs to include: purposes of processing, the categories of data concerned, the data undergoing processing, the recipients or categories of recipients to whom the data are disclosed, any available information about the source and logic involved in any automatic processing of data (Article 12a of the DPD).

The directive allows that national legislations define the meaning of "reasonable intervals" and "without excessive delay or expense". This resulted in variations across Member States.[73]

The right of an individual to receive confirmation that information relating to him or her is being processed is generally understood to mean that controllers are required to respond to every request, even if the response is to deny that data is being processed.[74]

[70] *Idem*, p. 126.
[71] *Supra* 62.
[72] *Supra* 50.
[73] *Supra* 56, p. 126.

The scope of information that needs to be conveyed to a data subject is not defined in the DPD, although that might be the case in national legislations. The Recital 38 of the DPD gives some guidance in this regard: *"Whereas, if the processing of data is to be fair, the data subject must be in a position to learn of the existence of a processing operation and, where data are collected from him, must be given accurate and full information, bearing in mind the circumstances of the collection."*

Knowing the logic of processing is often essential to decide whether one should invoke the right to object or erase. According to the directive, the logic should be *at least* revealed in the "*… case of the automated decisions which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc."*

Automated decisions are decisions based on mathematical algorithms or automatic search such as calculating credit scores. If they produce legal effects that concern an individual or affect him/her significantly, they should always be accessed.

Infringing trade secrets by requesting information about the logic behind an automated decision can be a challenge. Policymakers need to navigate intellectual property and privacy rights skilfully.[75] The DPD only briefly mentions this issue in Recital 41:*"… whereas this right (i.e. the right to access) must not adversely affect trade secrets or intellectual property and in particular the copyright protecting the software; whereas these considerations must not, however, result in the data subject being refused all information."*

Obtaining the communication is interpreted as receiving a copy.[76] However, as stressed above, the information needs to be accurate and full, but also intelligible. An illustrative case is Max Schrems' request to access his personal information processed by Google. As a response, he received over 1200 pages long file about the data that is processed about him.[77] While the overload of information can be seen as camouflaging meaningful data, it can also indicate the struggle of data controllers to appropriately address the applications. To help better handle access requests, the UK information commissioner has issued a useful guidance that can help controllers to appropriately react to data subject requests.[78]

There are numerous exceptions to the right to access, as the majority of Member States have taken the opportunity to include additional restrictions in their national laws. For instance, the German law expressly provides that the information should not be disclosed when the interest of a trade secret protection outweighs the interests of a data subject.

---

[74] *Idem*, p. 127.
[75] F Pasquale, 'Great Bargaining for Big Data: The emerging Law of Health Information', 72 Maryland Law Review 3, p.1.
[76] *Supra* 56, p. 127.
[77] <http://www.forbes.com/sites/kashmirhill/2012/02/07/the-austrian-thorn-in-facebooks-side/> accessed 23 January 2016.
[78] ICO, Subject access code of practice, Dealing with requests from individuals for personal information <https://ico.org.uk/media/for-organisations/documents/1065/subject-access-code-of-practice.pdf > accessed 23 January 2016.

5.1.5.2 The right to rectification, erasure and blocking of the data

If the data is incomplete, inaccurate or has been handled inappropriately by a controller, a data subject has a right to obtain rectification, erasure or blocking of such data processing (Article 12b of the DPD). Blocking of data will not be allowed if the processing complies with legal requirements.

As Levin explains, the data subject's right to obtain corrective action is formulated in a number of ways across the Member States. In the UK and Germany, the right is considerably restricted, as individuals must apply to the court for an order to rectify, block, erase and destroy the inaccurate data. In terms of possible remedies, the UK approach is similarly limiting, allowing the order for rectification, deletion or blocking only if the data subject experiences damage and if there is substantial risk for further non-compliance.[79]

In Slovenia, the law is more benevolent, judicial intervention is only required if the initial request is neglected by the controller. In addition, the Slovenian law only gives the controller 15 days to respond to an individual request and demands no judicial intervention.[80]

The second part of the right to rectification and erasure relates to cases where data was shared with third parties. In those cases, the controller has to notify the third parties whom the data was transferred to about the data subject's request, unless this would involve disproportionate efforts.[81]

The wording of the requirement in the directive is laid back. However, some of the Member States have not even implemented it (e.g., Greece).[82] In Slovenia, on the other hand, the requirement includes a stronger diction and excludes the notification duty only if the effort would be disproportional or would require excessive time.

Invoking data subject rights in the data economy can be very challenging. For example, retail companies use Twitter's APIs to bundle all tweets that contain a specific keyword, e.g., Tesco or Walmart, to analyse what the consumer preferences are. If a Twitter user later requests to delete his (publicly posted) tweets (in other words, if he returns them back to anonymity) from the reasons that Article 12 recognizes as valid, Twitter is responsible to help the user make this decision effective. Hence, it has to ensure that all third parties are informed about the user's move and request them to delete the tweet from their databases accordingly. This is seen as a burden for many third parties such as data analytics companies, which, as they have told us,[83] usually do not adhere to this policy. They have found out they

[79] *Supra* 56, p. 133.
[80] Slovenian Personal Data Protection Act (ZVOP-1). <https://www.coe.int/t/dghl/standardsetting/dataprotection/National%20laws/SLOVENIA_DP_LAW.pdf> accessed 23 January 2016.
[81] European Union Agency for Fundamental Rights & Council of Europe, *Handbook on European data protection law*, (Luxembourg, 2014), p. 72-75.
[82] *Supra* 56, p.133.
[83] The discussion with 3rdPLACE founders took place in Madrid, 18.6.2015, during the EuDEco workshop. It has to be pointed out, though, that our discussion focused on Twitter's requests for deletion that does not necessarily base on the right to

were better off breaching the law as well as Twitter's terms, as there has been no control in place and no serious threat has been imposed by data protection authorities (obviously also no economic incentive exists for Twitter to demand compliance). This shows that despite of the twofold protection, the contractual and the statutory one, data subject rights are difficult to invoke when the data is reused.

5.1.5.3 The right to object

The right to object consists of two parts. First, Article 14 of the DPD provides that the data subject should be granted that right at least for the case of processing based on the legitimate interest of a controller or processing that is necessary for the performance of public functions, the exercise of official authorities or a task carried out in public interest. The objection made by a data subject has to be compelling and legitimate, based on the circumstances of a specific situation and under the condition the national legislation provides no exception. If his or her objection is justified, the processing initiated by the controller may no longer involve those data.

The Netherlands, the UK, Ireland, Portugal, Slovenia and Germany have largely replicated the directive's diction allowing no additional options for a data subject to object to processing. On the other hand, Italy, Denmark, Austria and Luxembourg extended the right to basically all circumstances and included all legal basis for data processing.

The DPD provides no further guidance on what "compelling and legitimate ground" means. The UK and Irish laws specify that by expounding that processing "*is causing or likely to cause substantial damage or stress"* to a data subject or to another person and that the damage or distress is unwarranted. This precise explanation results in a higher bar for data subject objection requests.[84]

Second, paragraph (b) of Article 14 grants a right to data subjects to object, on request and free of charge, to the processing of personal data relating to him or her which the controller anticipates being processed for the purposes of direct marketing, or to be informed before personal data are disclosed for the first time to third parties or used on their behalf for the purposes of direct marketing, and to be expressly offered the right to object free of charge to such disclosures or uses. Member States shall take the necessary measures to ensure that data subjects are aware of the existence of the right referred to in the first subparagraph of (b).

As direct marketing[85] is considered as one of the most severe interferences with data subject privacy, the bar for claiming objection is set on a lower level. Whatever the circumstance, the data subject

---

rectification as Twitter's terms and conditions allow a user to also delete personal twits that were dealt with accurately and appropriately.

[84] *Supra* 56, p. 135.

[85] Today, all marketing really is direct marketing. Modern brand and consumer relationships are now built on greater insight, heightened personalisation and ever-more direct sophisticated marketing. This is achieved through the intelligent collection and analysis of data that has become available as consumers spend more time connected via multiple devices. Obviously, such an approach involves a lot data reuse. <http://www.theguardian.com/media-network/marketing-agencies-association-partner-zone/2015/may/20/modern-direct-marketing-data-analysis> accessed 23 January 2016.

should be able to object to processing. In addition, Member States should guarantee that a data subject is informed and expressly offered the right to object before his data is disclosed or used by third parties for the purpose of direct marketing.

Article 14 is favourable to data controllers as it only asks them to grant the right to objection (an opt-out) and to handle those requests without requiring an opt-in by individual users. Usually, companies will implement a management tool to record and respond to requests. This will be of special importance in Member States where the national legislator set forth more detailed provisions e.g. regarding the time slot in which a company has to respond or regarding the form in which the requests have to be given.

In relation to direct marketing, every reuser should also consider requirements in the e-Privacy directive,[86] which contains a number of specific provisions regarding unsolicited communications. Article 13 of the E-Privacy Directive sets forth a basic rule of "opt-in" consent for "unsolicited communications": automated calling systems without human intervention (automatic calling machines), facsimile machines (fax) or electronic mail for the purposes of direct marketing.[87] Contrary to Article 14 of the DPD, which regulates general methods used in direct marketing, e-Privacy directive focuses on the most intrusive practices and requires a prior, explicit confirmation by a data subject. Moreover, according to Article 5(3) Member States shall ensure that storing information, or gaining access to information already stored, in the terminal equipment of a subscriber or user is only allowed on condition that the subscriber or user concerned has given his or her consent, having been provided with clear and comprehensive information about the purposes of the processing. These opt-in requirements can also be seen as a reflection of a data subject's right to block improper uses and the legislator's intention to grant them more control especially when dealing with intrusive marketing practices.

5.1.5.4 The right not to be subjected to solely automated decisions

Article 15, paragraph (a) obliges Member States to grant the right to every person not to be subject to a decision which produces legal effects concerning him or her, or affects him or her significantly and which is based solely on automated processing of data intended to evaluate certain personal aspects relating to him or her, such as his or her performance at work, creditworthiness, reliability, conduct, etc.

A fully automated decision is a decision that in no stage of processing includes any human intervention. Those decisions are prohibited as long as they significantly affect a data subject i.e. decision related to his or her employment. In fact, it is not very likely that someone is subjected to an automated decision, since it will almost always involve at least a tiny part of human intervention.

[86] *Supra* 56.
[87] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)

According to Bygrave[88], there are four cumulative conditions that must be satisfied in order for Article 15 (a) to be applicable:

1. A decision must be made;
2. The decision concerned must have legal or otherwise significant effects on the person whom the decision targets;
3. The decision must be based solely on automated data processing;
4. The data processed must be intended to evaluate certain personal aspects of the person who is targeted by the decision.

Paragraph 2 of Article 15 of the DPD allows for two exceptions when automated decisions are legitimate. First, those decisions are permitted as part of pre-contractual and contractual arrangements. Second, they are allowed when subscribed by law, which also provides for adequate safeguards.

It can be seen from the wording of the article that its objective is to protect a data subject from "privacy-invasive processing applications that apply subjective criteria" rather than intervene with established society-benefitting activities such as issuing a speeding ticket.[89]

It is important to note that Article 15 helps strengthen the right in Article 14(b) of data subjects to object to data on them being processed for the purposes of direct marketing and also contributes to Article 12(a) which provides data subjects with, *inter alia*, a right to "knowledge of the logic involved in any automated processing of data concerning him at least in the case of the automated decisions referred to in Article 15(1)".[90]

Finally, as Levin's argues, Article 15 is likely to become increasingly important, particularly given the trend toward the convergence in technologies, increasing amounts of data-linking to individuals and the widening of the concept of personal data to include less traditional identifiers such as IP addresses, biometrics and GPS data. However, she argues that the definition of automated processing will become blurred, as computers are becoming more sophisticated. For example, in behavioural marketing campaigns, data concerning personality traits and browsing habits of individuals is collected and automatically segmented into predetermined market segments. Assuming that the data collected is personal, would the act of determining the qualities of each market segment be sufficient to mean that this is not a fully automated system?[91]

[88] LA Bygrave, 'Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling' (2001) Computer Law & Security Report 17, pp. 17–24.
[89] *Supra 56.*
[90] *Supra* 56, p. 139.
[91] *Ibidem.*

5.1.5.5 Data protection rights – developments in the GDPR proposal

The EU data protection law has been subject of ongoing legislative negotiations. The process started in 2012 when the European Commission (EC) presented its first proposal of the EU regulation on data protection (General Data Protection Regulation or, shortly, GDPR).[92] The objective of the new law was to strengthen data protection and adapt it to the changed circumstances in a globalized and interconnected world. Based on the current status of the legislative process, the new law will be enforced in early 2018. Companies should start getting ready for the new requirements, as many of them will require adjustments in their data management and other business processes.

As Zanfir and Solove note, one of the major amendments that the GDPR will bring to the existing data privacy laws is the enhancement of the package of rights of the data subject (strengthening and detailing the existing ones, and introducing new ones).[93]

The GDPR proposal contains an amended setup of data subjects' rights. While the DPD grouped the rights in Article 12 and, somewhat confusingly, tagged them as rights to access, the regulation will take a more structured approach. Chapter III of the proposal splits the rights into different groups starting with the transparency requirement and the right to information. It then goes on to the right to erase, the right to be forgotten and the right to data portability. Interestingly, the right to information has clearly been made a part of the data subject rights bundle. Also, very significantly, the GDPR proposal now explicitly includes two new rights – the right to be forgotten and the right to data portability.

As regards the existing rights, there have not been many changes. Right of access is kept in a slightly amended form. Data controllers are granted some legal protection when access requests are unreasonable or excessive.

The regulation does not change the right to object to certain types of processing, except from the fact that it explicitly mentions profiling as falling under the definition of data processing. This right is not absolute and controllers' interests can outweigh the individuals'. However, there is an absolute right to object to processing for purposes of direct marketing – which in the proposed regulation also covers profiling as long as it is related to direct marketing.[94]

---

[92] On February 25, 2012, the Commission publicly revealed the draft regulation and published additional materials to support the proposed reform including its communication and impact assessments: http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm. See also: Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM(2012) 11 final, 25.1.2012.

[93] G Zanfir, 'The Right to Data Portability in the Context of the EU Data Protection Reform' (2012) *International Data Privacy Law*, p. 149. D Solove, 10 Implications of the New EU General Data Protection Regulation (GDPR), https://www.teachprivacy.com/new-eu-data-protection-regulation-gdpr/> accessed 23 January 2016.

[94] <http://www.twobirds.com/en/news/articles/2015/global/agreement-on-general-data-protection-regulation?utm_source=Concep%20Send&utm_medium=email&utm_campaign=Agreement%20on%20general%20data%20protection%20regulation_12/18/2015#Enhanced%20Individuals%27%20Rights> accessed 23 January 2016.

Like the DPD also the GDPR proposal restricts the ability for controllers to engage in entirely automated-decision-making if the decision could produce legal effects or if it significantly affects the individual. The individual has a right to object to such processing. Appropriate protections for the individual must also be put in place. If the processing is necessary to enter into or to perform a contract, then the individual will not have a right to object to the processing – but will nevertheless possess a right of human intervention and appeal. Automated decisions involving sensitive personal data are further restricted.

5.1.5.5.1 Data portability

As noted in the IVIR report,[95] the rule on data portability is new to the fabric of personal data protection and can therefore be considered a regulatory innovation. In addition, the right to data portability is a highly controversial issue, largely because it is not clear from the legislative proposal whether this is a 'lex social network' or concerns every other context, such as electricity providers and banks.

The first time that a requirement on data portability was included in data protection legislation was in 2012 when the EC kicked off the data protection reform by publishing its draft EU regulation on data protection. In March 2014, after intensive negotiations in the Parliament, the initial Commission's proposal was significantly amended and its initial sharpness was softened, the data portability requirement, however, remained the same.[96] After lengthy negotiations in the Council, the provision on data portability was changed again, this time incorporating Member States' objections and concerns.[97] During the trialogue negotiations, the EC, the Parliament and the Council reached a final agreement, in which the portability provision was drafted in the benefit of data subjects by allowing them " … to obtain that the data is transmitted directly from controller to controller where technically feasible".

In the latest version of the GDPR proposal, the wording of the data portability provision reads as follows: *"*The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured and commonly used and machine readable format and have the right to transmit those data to another controller without hindrance from the controller to which the data have been provided, where:

(a) the processing is based on consent pursuant to point (a) of Article 6(1) or point (a) of Article 9 (2) or on a contract pursuant to point (b) of Article 6 (1); and
(b) the processing is carried out by automated means*"*[98]

[95] *Supra* 39.
[96] More than 4000 amendments were added. O Proust, 'The EU General Data Protection Regulation is on its way…but when?' (2015, April 17) <http://privacylawblog.fieldfisher.com/2015/the-eu-dp-regulation-is-on-its-way-but-when> accessed 23 January 2016.
[97] More information on the Council's version is available via the following website: <http://www.consilium.europa.eu/en/press/press-releases/2015/06/15-jha-data-protection/> accessed 23 January 2016.
[98] <https://iapp.org/media/pdf/resource_center/2015_12_15-GDPR_final_outcome_trilogue_consolidated_text.pdf> accessed 23 January 2016.

From the above, we can see that the right to portability is split into two elements: firstly, the right to obtain a copy of that data for further use and secondly, the right for individuals to transmit their personal data from one provider to another.[99]

The article applies generally to all types of electronic processing including cloud computing, web services, smartphone apps, and other automated data processing systems. As we could read in the recitals, the idea of data portability was introduced due to alleged locks-in in case of the social networks,[100] but it is, given the open definition in Article 18, not limited to one specific market.

The right to data portability differs substantially from the right of access, although the latter can be seen as a predecessor. It goes over mere accessibility and insight into the data by emphasizing further data use. In other words, data portability transforms passive data subjects into active reusers and empowers them to take advantage of value-added services from third parties and lets individuals "share the wealth" created by big data.[101] Admittedly, data sharing also presupposes data subjects benefit from the fact they have exchanged their personal information with the service provider, however, data portability is the point when a data subject actively takes control over his or her own data and allocates it to the party where they can be reused in an economically more effective way.

Along with the right to be forgotten and the right to modify incorrect or outdated personal information stored in databases, data portability is a pillar of a stronger, more effective right to control over the processing of the data subject's personal data.[102]

From an individual's point of view, data portability is seen as a safeguard to his or her information self-determination by giving the individual actual control over his or her data stored in databases and having the freedom to choose the service provider for the storage and process of such data.[103] This is very similar to the objective of the right to be forgotten (see below). It has been argued that the increased data portability would decrease data protection, however, Zanfir believes that it would in fact enhance it provided that the regulator adopts adequate safeguards.[104]

[99] In the amended text of the European Parliament both aspects are joined in the same section of the article on the right to access, namely Article 15(2a). I Graef , J Verschakelen and P Valcke (2013) Putting the Right to Data Portability into a Competition Law Perspective. *The Journal of the Higher School of Economics,* Annual Review , pp. 53-63.
[100] While social networking sites like Facebook and Google+ offer users the possibility to obtain a copy of their data, there are still considerable limits on the direct transfer of personal information to other platforms. Moreover, social network providers do not allow third-party sites to directly acquire the user's information. For instance, Facebook blocks Google Chrome's extension for exporting friends. I Graef , J Verschakelen and P Valcke (2013) Putting the Right to Data Portability into a Competition Law Perspective. *The Journal of the Higher School of Economics,* Annual Review , pp. 53-63.
[101] 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: Brussels, March 2014).
[102] *Supra* 93.
[103] *Ibidem*.
[104] *Ibidem*.

The EC's definition of data portability emphasized the human side of the right in Recital 55, where it explained: *"To further strengthen the control over their own data and their right of access, data subjects should have the right, where personal data are processed by electronic means and in a structured and commonly used format, to obtain a copy of the data concerning them also in commonly used electronic format."* [105]

Reputation is another issue worth to consider in relation to data protection and data portability.[106] On websites like eBay or Airbnb concepts of identity and reputation play a major role. Reputation, as being part of one's personality, may be closely related to the free development of human personality and users' economic interests. Data portability has a strong link with that as it prevents excessive switching costs for the users of the services that rely on testimonials.[107]

However, data portability is also described as a key measure to avoid locks-in, opt for a more secure or more developed provider, and strengthen the competition on the market.[108] In addition, data portability also increases consumer protection: In particular, it can foster a more competitive market environment, by allowing customers more easily to switch providers (e.g., in the context of online banking or in case of energy suppliers in a smart grid environment). In Geradin's words, we could say it is the ability for people to *reuse* their data across devices and services.[109]

Furthermore, data portability can also contribute to the development of additional value - added services by third parties who may be able to access the customers' data at the request and based on the consent of the customers. This, again, may bring down barriers to entry to new markets that require access to personal data, and help create more competitive, less monopolistic market structures.

Data portability is becoming more relevant in the age of big data. Allowing data portability could enable businesses and individuals to maximise the benefits of big data in a more balanced and transparent way and may help redress the economic imbalance between controllers on one hand and individuals on the other. It could also let individuals benefit from the value created by the use of their personal data: it could allow them to use the data for their own purposes, or to license the data for further use to third parties, in exchange of additional services, or for cash value. Further, it could also help minimise unfair or discriminatory practices and reduce the risks of using inaccurate data for decision-making purposes.[110]

[105] Although Commissioner Almunia has also clearly acknowledged that data portability is also a measure of competition law. <http://europa.eu/rapid/press-release_SPEECH-12-860_en.htm> accessed 23 January 2016.
[106] *Supra* 93.
[107] <http://blogs.lse.ac.uk/mediapolicyproject/2014/04/16/data-portability-series-capitalising-on-the-market-for-interoperability/> accessed 23 January 2016.
[108] See also P Swire and Y Lagos, 'Why the Right to Data Portability Likely Reduces Consumer Welfare: Antitrust and Privacy Critique' (May 31, 2013), 72 Maryland Law Review 335.
[109] D Geradin (2014). Data portability and EU competition law, presentation at BITS conference, Brussels.
[110] *Supra* 100.

Besides the arguments related to human rights and antitrust, data portability should also be promoted due to its positive influence to data interoperability. Standardized data format would boost innovation and growth; however, many authors warn this is not easily achievable.[111] The contested points are the feasibility of implementing the portability of personal data and the extent to which this implies mandating electronic data exchange formats.[112]

5.1.5.5.2 The right to be forgotten

This right is a crystallisation of the more fundamental wish for 'control' over one's personal data and another safeguard of information self-determination.[113] Contrary to data portability, which is described above, the right to be forgotten does not imply data reuse but rather deals with blocking all further, secondary uses.

The right to be forgotten is a manifestation of the right to oblivion in the digital age. Originally, the right to oblivion was introduced to be invoked in cases where undesired public exposure is given to a person's past, as a shield against disproportionate intrusion by mainstream media (papers, news broadcasts, radio plays, etc.) into the private life of people who have entered into the public eye.[114] The right to be forgotten has no such tradition or connotation,[115] since it primarily aims at protecting an individual's *digital* reputation, but the fundamental interests it safeguards are similar.[116] As Werro establishes, the right to be forgotten ensures that someone can preclude others from identifying him or her in relation to his or her (criminal) past.[117] What is important is that Werro's definition focuses not so much on deletion of data, but rather on regulating (blocking) the *(re)use* of data.[118]

In 2012, when the EC came up with the data protection reform, it proposed the right to be forgotten as an independent right. This has been one of the most attention-grabbing parts of the EC's proposal, although it falls short from being a new legal concept, like the right to data portability.[119] The DPD from 1995 already included the principles underpinning the right to be forgotten,[120] but the proposed legal

[111] *Supra* 107.
[112] *Supra* 39, p. 68.
[113] J Ausloos, 'The ''Right to be Forgotten'' - Worth Remembering?' (2012) 28 Computer Law and Security Review, 2, 143-152.
[114] H Graux, J. Ausloos, J. and P Valcke, 'The Right to be Forgotten in the Internet Era' (2012) 11 ICRI Research Paper.
[115] *Idem*.
[116] G Finocchiaro and A Ricci, 'Quality of Information, the Right to Oblivion and Digital Reputation' in B Custers, T Calders, B Schermer and T Zarsky (Eds*.) Discrimination and Privacy in the Information Society, Data Mining and Profiling in Large Databases* (Berlin Heidelberg: Springer Verlag, 2013).
[117] F Werro, 'The Right to Inform v the Right to be Forgotten: A Transatlantic Clash' in AC Ciacchi, C Godt, P Rott and LJ Smith (eds), Haftungsbereich im dritten Millennium / Liability in the Third Millennium (Baden-Baden: Nomos, 2009) 291.
[118] BJ Koops, 'Forgetting footprints, shunning shadows' (2011) 8 *SCRIPTed* 3, 5.
[119] In the landmark case Case 131/12 *Google Spain v. AEPD and Mario Costeja Gonzales* issued on May 13 2014, the Court of Justice of the European Union made it clear that the right to be forgotten is encompassed in the fundamental right to privacy. See also C Kuner, 'The Court of Justice of the EU Judgment on Data Protection and Internet Search Engines' (2014) LSE Legal Studies Working Paper No. 3/2015.
[120] *De lege lata* the right to be forgotten has been reflected in the right to objection and deletion. Factsheet on the "Right to be forgotten" ruling C-131/12, European Commission (2014).

framework reaffirms and reshapes them to suit the modern information society better. After the Parliament's amendments, the provision was renamed into the right to erase. In the latest version of the GDPR proposal, the right to be forgotten is again explicated in an independent article (Article 17).

The diction of the right to be forgotten in the proposal has emphasized the importance of consent and the purpose limitation principle. Article 17 (1) explicitly allows data subjects to seek the deletion of data and block further reuse when the consent is withdrawn[121] or when the data is no longer necessary in relation to the purposes for which it was collected or otherwise processed. This is an important difference from the DPD, where deletion was only possible if processing conflicted with the legal rules. Article 17 (2) further proposes that the right to be forgotten should follow the data when the data controller has made it public (e.g., by publishing it on a website) or when publication is delegated to a third party. In the first scenario, the original controller only has to take 'all reasonable steps' to inform third parties about the data subject's request for erasure. In the second situation, the original controller will be considered responsible in any case.[122] This diction radically shifts the burden of proof – it is now for the data controller and not for the individual to prove that the data cannot be deleted because it is still needed or relevant.[123] Article 17 (3) provides for a number of exceptions to the general rule – if there are counter interests such as freedom of exception or various legal obligations, the right to be forgotten cannot be enforced.

According to Van Hoboken (2013) the added value of the updated provision for data subjects that want to see their data deleted is relatively minor.[124] Kuner (2012), on the contrary, believes the new provision is a significant one, in particular its reversed side, the duty of the controller to inform third parties about the data subject's request to erase data.[125] After realizing how little it took the CJEU to formulate the right to be forgotten from the existing provisions,[126] we agree with Van Hoboken that the new provision is anything but a revolution. However, we do acknowledge that the diction in the proposal is favourable to data subjects and represents a step forward to better data protection.

Despite this fresh approach to the right to erasure, many dilemmas have remained unsolved. The socio-technical context of big data implies that data processing is based on vague purpose definitions to allow unforeseen future uses and that data are increasingly used for secondary purposes. This fundamentally challenges not only the purpose-limitation principle itself but also the effectiveness of a right to be

[121] Article 7(3).
[122] *Ibidem*.
[123] Factsheet on the "Right to be forgotten" ruling C-131/12, European Commission (2014).
[124] J Van Hoboken, The Proposed Right to be Forgotten Seen from the Perspective of Our Right to Remember Freedom of Expression Safeguards in a Converging Information Environment (2013) <http://www.law.nyu.edu/sites/default/files/upload_documents/VanHoboken_RightTo%20Be%20Forgotten_Manuscript_2013.pdf> accessed 27 July 2015. See also Article 29 Working Party (2014), Opinion 05/2014 on Anonymization Techniques.
[125] C Kuner, The European Commission's Proposed Data Protection Regulation: A Copernican Revolution in European Data Protection Law (6 February 2012) Bloomberg BNA Privacy and Security Law Report, 1-15.
[126] Directive 95/46 and the EU Charter of fundamental rights.

forgotten.[127] For example, in an increasingly personalised Internet almost every bit of personal data can be argued to be relevant[128] and it will be hard to establish that the data should be forgotten on the ground of "no longer being necessary for the purpose for which it was initially collected".

Another question is who should do the balancing test. Should this be a commercial party such as Google? Google strongly opposes by stating they are an objective entity with no preference over each side of the scale – either this is the right of a data subject or the right to free speech and freedom of expression. Edward Lee argues that an independent agency should take over this challenging task.[129]

Table 7 Compliance checklist for data protection law (e)

Compliance checklist
• How can a data subject access his/her data? • Is there a procedure in place to enable data erasure or blocking? • Is there clear information about the purposes of data reuse, transfers to third parties etc. available online?

5.1.6 Data transfers

In the data economy, global transactions and transfers have a pivotal role. A data reuser can be much more effective, if it is free to transfer the data beyond the EU borders. However, the DPD imposes restrictions to data transfers outside the EU. Because of that, some authors describe it as a blocking provision.[130] According to the DPD, personal data can only be transferred outside the EU, if the third country ensures an adequate level of protection. There are different instruments available to manage the transfers:

- EC adequacy decisions (Andorra, Argentina, Australia, Canada, Switzerland, Faeroe Islands, Guernsey, Israel, Isle of Man, Jersey, US, New Zealand ad Uruguay). Here it should be noted that the EC's decision that concerned the transfers to the US has recently been revoked. Companies are no longer allowed to use this instrument and are advised to adopt alterative solutions e.g. standard clauses or binding corporate rules.[131] In the beginning of February 2016 the EC announced the agreement on the future framework for the EU-US data transfers, the so-called

[127] *Ibidem*.
[128] I Graef , J Verschakelen and P Valcke (2013) Putting the Right to Data Portability into a Competition Law Perspective. *The Journal of the Higher School of Economics,* Annual Review , pp. 53-63.
[129] E Lee, Recognizing Rights in Real Time: The Role of Google in the EU Right to Be Forgotten (July 28, 2015). UC Davis Law Review, Forthcoming; Chicago-Kent College of Law Research Paper No. 2015-13.
[130] *Supra* 44.
[131] Court of Justice of the European Union, Press release No 117/15 Luxembourg, 6 October 2015. <http://curia.europa.eu/jcms/upload/docs/application/pdf/2015-10/cp150117en.pdf> accessed 27 January 2015.

privacy shield, which will replace the annulled Safe harbour decision and help stabilize data transfers between the continents.[132]

- Consent of a data subject, a transfer is based on and agreed on in a contract, there is a prevailing public interest that legitimates the transfer, legitimate interest
- Safeguards: contractual clauses (as standardized set of clauses approved by the EC or as *ad hoc* contracts that are individually approved by national data protection authorities) or binding corporate rules

Only loading personal data on the Internet (where this data can be globally accessed) is not a data transfer.[133]

In the GDPR, the current system is broadly carried across, although there were also some amendments.[134] An improvement is a section on binding corporate rules (BCRs). BCRs stand for internal rules (such as a Code of Conduct) adopted by multinational companies, which define its global policy with regard to the international transfers of personal data within the same corporate group to entities located in countries, which do not provide an adequate level of protection.[135] The fact that this option is now largely explicated in Article 43 indicates the intention of the legislator to encourage the usage of BCRs as a viable instrument for managing international transfers.

International perspectives:

Francesca Bignami has stressed the fact that after the Schrems ruling[136]*, the data flows between the EU and the States have been blocked. Compared to another major system for data transfers, the APEC (*Asia-Pacific Economic Cooperation*) privacy framework*[137]*, the EU system is burdensome*[138]*. Bignami pointed out, for example, that the APEC agreement on data transfers cannot be aligned with the EU binding corporate rules framework. There are differences in the perception of privacy in the public space, for instance, on social media, and the US conception of harm as a trigger for a privacy claim. Different expectations cause frustration of the users and decrease the level of trust.*

[132] <http://europa.eu/rapid/press-release_IP-16-216_en.htm > accessed 23 February 2015.
[133] *Lindquist,* C-101/01, 6 November 2003.
[134] *Supra* 94.
[135] <http://ec.europa.eu/justice/data-protection/international-transfers/binding-corporate-rules/index_en.htm> accessed 27 January 2015.
[136] *Maximillian Schrems v Data Protection Commissioner,* C-362/14, 6 October 2015.
[137] Asia Pacific Economic Cooperation's (APEC), Privacy framework.
[138] *Supra* 44.

Table 8 Compliance checklist for data protection law (f)

Compliance checklist
• Will any personal data be made available outside the country of origin of the personal data? If yes, check whether the country to which the data is transferred can be considered adequate. If the country is not treated as adequate, pick one of the tools that enable such data transfers.

5.2 Data protection law from the perspective of the Eudeco reuse model

The first version of the EuDEco model was discussed in D2.1. This first version is a heuristic model consisting of a set of propositions from the legal, the socio-economic and the technological perspective. To further combine these different perspectives, the socio-economic and technological challenges are discussed from a legal perspective in this section. In D2.3 and D2.4, the legal propositions are also discussed from a socio-economic and technological perspective.

5.2.1 Socio-economic propositions

Table 9 Socio-economic propositions related to data protection law

Socio economic propositions	***Legal response – data protection law***
Data users need a possibility to assess the relevance of data	The right to access and transparency play an important role in assessing the relevance of data. The EDPS has observed that only few individuals exercise their rights in practice.[139] The right to data portability, as proposed in the proposal for the GDPR tends to fill the gap by giving the individuals more tangible rules to effectively invoke their rights and to be fairly compensated for their personal information.
Data users need to be able to determine whether the data were trustworthy (e.g. through certificates, open data documentation)	It has been stressed that creating trust is one of the main objectives of the EU data protection law. A firm that is considered untrustworthy will find it difficult or impossible to collect certain types of data, regardless of the value offered in exchange.[140] A harsher approach to data protection reflects the

[139] *Supra* 100.
[140] <https://hbr.org/2015/05/customer-data-designing-for-transparency-and-trust> accessed 27 January 2015.

	concern of policy makers for insufficient level of trust for the economy. Self-regulation can be one solution. The proposal for the GDPR, for example, includes the option to apply for a privacy seal, an international self-regulatory recognition of adequate data protection practice.
Best practices in experimental methods and in the storage, archiving, and dissemination of experimental data should be applied	Standards of information security (e.g. ISO standards) could be seen as a partial, but not holistic solution. As one possible approach to responsible big data analytics, the UK information commissioner describes IBM's ethical framework.
Standards or guidelines are needed to facilitate interoperability and reuse	There are no legal rules in place to facilitate interoperability and reuse. The right to data portability indicates the future direction toward easier reusability and interoperability of digital assets. Due to unclear wording of the legal text, its application can be limited.
The big data dilemma describes the clash between (1) the opportunity to benefit from the use of innovative devices and services, based on data collection and analysis, and (2) possible consequences, particularly in terms of preserving one's privacy and business confidentiality	Law should strike the right balance and new solutions should be adopted. The proposed GDPR is moving toward this direction. Technology can help accelerate the adoption of the mechanisms that would restore the balance e.g. by introducing new privacy by design and default methods.
Traditional data sources such as company databases and applications are now complemented by *non-traditional sources* such as social media or sensors embedded in physical world devices including mobile devices, smart meters, cars and industrial machines.	Should/does law treat the non-traditional sources differently? In principle, the law treats all personal data in the same way, unless the data is sensitive and therefore requires additional protection. Given the current indications by the legislators, it can be assumed this will remain the default rule also in the future.[141]

[141] See for example Article 29 Working Party (2014), Opinion 8/2014 on the Recent Developments on the Internet of Things (No. Wp 223).

5.2.2 Technological propositions

Table 10 Technological propositions related to data protection law

Technological propositions		***Legal response – Data protection requirements***
Scalability and data management		
Unknown needs		The principle of purpose limitation defers from storing a large amount of data for future, non-specified uses. This may cause difficulties with meeting the unknown needs that could not have been predicted in advance. Data quality aims to minimize business practices that lead to an accumulation of a large amount of conflicting data.
Conflicting data		
Security and privacy aspects		
Data breaches	A data breach is defined as any incident involving the loss or exposure of digital personal records.	Data protection law includes several paragraphs on data security. In case of data reuse, an important source is also the e-Privacy directive and the proposal for the NIS directive, which impose additional security requirements for the transfer of data via communication channels.
Data loss	Data stored in the cloud can be lost due to technical reasons such as accidental deletion by the cloud service provider, or a physical catastrophe such as a fire or earthquake, or if a company has encrypted data and loses the encryption key.	
Account hijacking	Account and service hijacking involves phishing, fraud and software vulnerabilities where attackers steal credentials and gain unauthorized access to servers.	

Insecure APIs	From authentication and access control to encryption and activity monitoring, these interfaces must be designed to protect against both accidental and malicious attempts to circumvent policy.	
Account misuse	In data driven and distributed environments, providers often require different approaches for handling security. Simple security errors such as too simple passwords or too similar passwords often lead to big negative	

5.2.3 CAS and data protection law

Technologic views

As shown above, data protection law is strongly dependent on the technological and socio-economic environment.

The rise of the Internet of Things (IoT) means that the collection of data is getting new dimensions. Repurposing of the data collected by multiple devices is both a business opportunity and a risk. The main challenge is the fact that there are plenty of players involved in the functioning of the IoT (device manufacturers, third party application developers, IoT platforms, reusers such as insurance companies) who should all pay attention to the data protection requirements.

In the era of the IoT, data subjects seem to be more vulnerable than ever before. The burning issues are the quality of their consent and the level of effective control over their data. Some authors have proposed the idea of "renewing consent", where new approval has to be given every few months.[142] This would ensure that data subjects have an additional opportunity to express their opinion on data processing. For data reusers, however, this could be a challenge, since their data sets may suddenly lose value.

Cloud computing is one of the drivers of the data economy. Companies that use data as their main asset can easily access large cloud-based databases, which reduces the cost of doing business. Despite these obvious advantages, cloud computing also raises some legal, especially privacy-related risks. As cloud data cannot be precisely located, this will challenge the principle of the territorial applicability of the law. Data protection laws were adopted in the pre-Internet era, when centralised and limited processing

[142] Article 29 Working Party (2011) Opinion 15/2011 on the definition of consent (No.WP 187). *Custers, B.H.M. (2016) Click here to consent forever; Expiry dates for informed consent, Big Data and Society, pp. 1-6. DOI: 10.1177/2053951715624935.*

was common.[143] The situation has dramatically changed in the recent years, and the consequences, such as the loss of meaningful protection, can already be witnessed. For instance, European users may store their data in the cloud of a provider without realizing they are not protected by the EU legal regime.

Moreover, cloud computing can also challenge the purpose limitation principle. As a typical cloud scenario may easily involve a larger number of subcontractors, the risk of processing personal data for further, incompatible purposes is high. To minimise this risk, the WP29 suggests that the contract between cloud providers and cloud clients includes adequate technical and organisational measures that mitigate the risk of illegitimate secondary processing.[144]

Data subject rights are another problematic area. For instance, one SaaS provider pointed out that, because users have direct access to and control over data, including any personal data, it should be unnecessary to require providers to grant users the right to access. However, this proposal contravenes the idea of data protection law, which protects data subjects regardless of the contractual relationship between a user and a service provider.[145]

The examples above show how rapidly changing economic and technological environments influence the legal system. On the one hand, the technology opens up new dilemmas and questions the traditional interpretations of legal requirements, on the other hand, it supports reusers in their mission to comply with the law, e.g. by introducing technical tools to manage data subject requests or by implementing privacy by design in the development stage of a new technology.

Technology can be a saviour or an enemy. A good example is data anonymization, a technical concept that helps protect personal information. However, in spite of anonymising data it is still possible to achieve privacy interfering consequences. The belief that big data reusers who only reuse anonymised data are not bound by any legal obligation is a common misconception.

Data subjects and socio-economic views

Data subject rights are seen as the shield against intrusive business practices in the data-driven economy. Transparency, which also strongly relates to data subject rights, should be a default characteristic of business models. The importance of trust should create new ways to restore the balance on the market and drive legal changes. For example, to secure a higher level of trust the Internet service provider's involvement in data transfers should be enhanced.

[143] P Van Eecke, Cloud computing – legal issues, DLA Piper Brussels. <http://www.isaca.org/Groups/Professional-English/cloud-computing/GroupDocuments/DLA_Cloud%20computing%20legal%20issues.pdf> accessed 29 January 2016.
[144] Article 29 Working Party (2012), Opinion 05/2012 on Cloud Computing (No. WP 196)
[145] W Kuan Hon, C Millard and I Walden, Negotiating cloud contracts: looking at clouds from both sides now (2012) *16 Stanford technology law review* 1.

6 Human rights requirements and their role in the model

6.1 Privacy and human rights – overview of the legal framework

Table 11 Overview of the legal framework for the EU privacy and non-discrimination law

EU PRIVACY and NON-DISCRIMINATION LAW	
Primary EU law	Charter of the fundamental rights of the EU (Arts. 7 and 8)
	Treaty on the functioning of the EU (Art 16)
	European Convention on Human Rights (Article 14)
	United Nations' documents related to human rights
Secondary EU law	General Data Protection Directive
	E-privacy Directive
	General Data Protection Regulation (draft proposal)
	Anti-discrimination directives

6.1.1 Privacy law

Personal data protection law is closely related to but not the same as privacy law. The right to privacy is guaranteed by several international documents, in the EU, most notably by Article 8 of the EU Charter of human rights. Privacy has several different aspects, such as spatial privacy (for instance, in your home), relational privacy (for instance, during phone calls), physical integrity (for instance, not to be touched without consent) and informational privacy (for instance, the use of personal data). Personal data protection law particularly focuses on informational privacy.

Data controllers increasingly seem to make use of aggregated, anonymized data. Anonymization is a process of turning data into a form, which does not identify individuals.[146] Non-identifiable data is no longer personal data, hence, data protection law does not apply anymore. When the legal regime for protection of personal data is considered too restrictive, data controllers are particularly keen to adopt that technical solution. Anonymized data can be as useful as personal data in many cases. A typical example may be a company that wants to personalize its marketing campaigns with the help of profiling. The use of personal data may be helpful to assess which people are potentially interested in particular products or services, but aggregated data on street level or neighbourhood level may be similarly useful

[146] Definition used by ICO: ICO, *Anonymisation: managing data protection risk, Code of practice*, <https://ico.org.uk/media/1061/anonymisation-code.pdf> 23 January 2016.

and cheaper to process (no consent procedures required, no too detailed selection procedures necessary).[147] The fact that such targeting is not completely accurate (false positives and false negatives may exist) does not matter significantly and the costs of a more accurate approach are not proportionate.

As explained above, anonymized data are not protected by the DPD. However, this does not mean no protection is required. Big data developments such as profiling, personalization and (de-)identification may affect (the right to) privacy, although there have been no personal data used. With the help of big data, characteristics of people who refused to provide consent to process their personal data may be predicted anyway. Big data predictions can be made with high accuracy; Kosinski, Stillwell, and Graepel showed how a range of highly sensitive personal characteristics, including sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances and parental separation can be predicted very accurately on the basis of Facebook likes.[148] Obviously, predicting missing values is also possible for people who (whether on purpose or not) provided false information.

In addition, big data is a *shared resource*. Given recent advances in data processing techniques, personal data is no longer strictly personal. Much like genetic data pertains to one individual, but also reveals information about other people sharing the same genes, personal data disclosed by one individual – when put through the big data algorithms – reveals information about and hence presents benefits and risks to others.

As it has become clear that it is not possible to establish with absolute certainty that an individual cannot be identified from a particular dataset in combination with other data that may exist elsewhere,[149] the EDPS has encouraged those who employ anonymization techniques to carefully use such techniques in combination with other safeguards. Anonymization cannot be achieved by just stripping a dataset of some directly identifying attributes but requires a much more prudent approach.[150]

Given the numerous aspects of the right to privacy, Article 8 right will often overlap with the protection provided by the data protection acts. Normally, if a data processing is compliant with the data protection acts, it is likely to be compliant with the human rights provisions. However, the Article 8 right is not limited to situations involving the processing of personal data. This means that some disclosures

[147] For more examples, see TZ Zarsky, ''Mine your own business!': Making the case for the implications of the data mining of personal information in the forum of public opinion' (2003) 5 *Yale Journal of Law and Technology* 1.
[148] M Kosinski, D Stillwell and T Graepel, 'Private traits and attributes are predictable from digital records of human behavior' (2013) 110 *Proceedings of the National Academy of Sciences,* 15.
[149] P Ohm, 'Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization' (2009) UCLA Law Review, 57, 1701.
[150] 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' 19 November 2014. p.15

of information that do not engage the DPA could still engage the broader human rights provision. For example, information about a large family group might not be personal data but its disclosure may well breach the privacy rights of the family.[151]

6.1.2 Non-discrimination law and other human rights

Big data also raises issues with regard to equal treatment and non-discrimination and related laws. As indicated above, big data may be useful for profiling purposes, but the results from profiling and other types of data analyses may turn out to be stigmatizing or discriminating. When selecting individuals or groups of people on particular characteristics, this may be unwanted or unjustified or both. Selecting for jobs, offering products and services to specific groups only, and some other decision-making is considered unethical and, in many countries, forbidden by (anti-discrimination) law when it takes place on the basis of gender, ethnic background, etc. When risk profiles constructed by companies, governments or researchers become 'public knowledge', this may also lead to stigmatization of particular groups. Discrimination and stigmatization on a large scale may also result in polarization of (different groups of) society.

Interestingly, research has shown that removing sensitive attributes (such as ethnicity, gender, etc.) from databases in order to prevent unethical or illegal discriminating results does not prevent finding such profiles.[152] There are several possible explanations for this. For instance, most data mining tools make predictions on the basis of training data. If the training data is biased towards particular groups or classes of objects, e.g., there is racial discrimination towards black people, the learned model will also show discriminatory behaviour towards that particular community. These are self-fulfilling prophecies. Another reason is that often other attributes than the sensitive attributes that were removed will still allow for the identification of the discriminated community.[153] For example, the ethnicity of a person might be strongly linked with the postal code of his residential area, leading to a classifier with indirect racial discriminatory behaviour based on postal code. The postal code then serves as a proxy for ethnicity. This is closely related to the ease of making predictions of missing attributes in big data settings: this can be done for both identifying data items (resulting in privacy issues) and for items like ethnicity, religion, gender, etc. (resulting in discrimination issues).

Big data can even put pressure on human dignity. Solove[154] argues that in the information society, the reputation of people is more and more constituted by the data that is disclosed about them. Such disclosure of personal data can be voluntary or involuntary. As a result, people are also increasingly

[151] *Supra* 144.
[152] F Kamiran, and T Calders, 'Classification without discrimination' in *Proceedings of the IEEE International Conference on Computer, Control and Communication* (IEEE-IC4, 2009).
[153] D Pedreshi, S Ruggieri and F Turini, 'Discrimination-aware data mining' in *Proceedings of the 14th ACM SIGKDD International conference on Knowledge discovery and data mining* (2008), pp. 560–568.
[154] DJ Solove, *The future of reputation: Gossip, rumor, and privacy on the internet*, (New Haven Conn. u.a.: Yale Univ. Press, 2007).

judged upon their digital representation (the digital person) rather than human beings of flesh and blood. An example of a relationship based on digital reputation, trust and economic dependence is sharing economy. According to a recent workshop at the FTC, the reputation could replace the regulation, if the sharing economy continues to grow in the future.[155] This may be particularly problematic when characteristics of digital identities are incorrect or incomplete. It may also be problematic when automated decisions (i.e., without further human interference) are made upon individuals based solely on their digital identity.[156] Practices like profiling can reinforce a tendency to regard persons as mere objects.[157] Another issue may be so-called chilling effects. This refers to the fact that people may alter their behaviour when they are aware that they are being monitored. Sometimes, for instance in cases of camera surveillance, the aim is precisely to make people behave 'better', but a more general effect may be that people behave more modest and reluctant overall, reducing their freedom of expression, liberty and other important human rights and values.

Finally, data reuse can even challenge liberty and justice. The lack of privacy in the data economy greatly increases the possibility of price discrimination and influences some basic postulates of the free market.[158]

Table 12 Compliance checklist for the EU privacy and non-discrimination law

Compliance checklist
• Anonymising data does not mean that all legal restrictions become superfluous. • Big data analytics can impair some other legal and ethical values, e.g., equality, liberty and dignity. Before taking a decision based on big data, it is therefore important to consider larger impacts including discriminatory or other unethical consequences.

[155] Federal Trade Commission, The "Sharing" Economy: Issues Facing Platforms, Participants, and Regulators , A Federal Trade Commission Workshop <https://www.ftc.gov/system/files/documents/public_events/636241/sharing_economy_workshop_announcement.pdf> accessed 23 January 2016.
[156] Note that EU personal data protection law prohibits automated decision-making that is solely based on automated processing of data. See Article 15 of the General Directive.
[157] LA Bygrave, *Data protection law: Approaching its rationale, logic and limits*. *Information law series: Vol. 10* (The Hague: Kluwer Law International, 2002).
[158] A Bernasek, *All you can pay* (Nation books New York, 2015), p. 328.

6.2 The EU privacy and non-discrimination law from the perspective of the EuDEco model

6.2.1 Assessment of the EU privacy and non-discrimination law requirements related socio-economic and technological propositions

Table 13 Socio-economic and technological propositions related to the EU privacy and non-discrimination law

Technological propositions		***Legal response – Privacy and HR requirements***
Noise accumulation	Estimation errors accumulate when a decision or prediction rule depends on a large number of such parameters. Such a noise accumulation effect is especially severe in high dimensions and may even dominate the true signals.	The analysis of the technological framework has revealed that errors were difficult to eliminate, which can *have an impact on personal data processing and consequently affect individuals*. Fake correlations, noise accumulation and other errors can negatively affect data reuse not only in relation to personal data but also to the anonymised data. As we have shown, in the data economy those consequences can be very serious and can even affect some basic human rights provisions. Anonymization should not be seen as an absolute answer to the challenges of privacy and human rights protection.
Sophistication	Big data solutions available today do not support concepts required to carry out specific clustering, classification or network analysis tasks	

Socio-economic propositions		***Legal response – Privacy and HR requirements***
The big data dilemma describes the clash between (1) the opportunity to benefit from the use of innovative devices and services, based on data collection and analysis, and (2) possible consequences, particularly in terms of preserving one's privacy and business confidentiality		Technology can help accelerate the adoption of the mechanisms that would restore the balance by introducing new privacy by design and default methods. That said, technology should not be seen as an absolute solution. In some cases, applying the latest technological solutions still does not resolve problems with identification or discrimination.

6.2.2 CAS

Technology can help protect fundamental rights, but technology alone is not the right solution. Other safeguards should be sought, for example, standards, code of conducts or legal regulation.

When data is anonymised and data protection law ceases to apply, ethical questions become even more significant. As the economy is nowadays building on trust and digital reputation, a prudent approach toward ethical question, especially in consumer relation, will become increasingly important. Those issues have to be considered at an early development stage of a project.

The big data dilemmas are strikingly complex and without assessing them from the perspective of the data economy as a CAS, it will be difficult if not impossible to find an adequate solution. In particular, technology and laws have to work together to find methods that fairly balance the interests of everyone involved in data reuse.

7 Intellectual property requirements and their role in the model

7.1 Intellectual Property Law – overview of the legal framework

Table 14 Overview of the legal framework for the EU intellectual property law

EU INTELLECTUAL PROPERTY LAW	
Primary EU law	Charter of the fundamental rights of the EU (Article 17.2)
	Treaty on the functioning of the EU (Arts. 118, 207)
Secondary EU law	InfoSoc Directive (Copyright Directive)
	Database Directive
	Software Directive
	Trademark Regulation
	Unitary Patent Regulation (entry into force pending)
	Trade Secrets Directive (proposal)
National legislation	

Intellectual property rights (IPRs) protect immaterial goods, which are mostly the product of a creative mental human activity in the industrial, scientific, literary and artistic fields.[159] Based on the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) from 1994, which is the fundamental, internationally accepted set of IP related principles, we can roughly distinguish eight groups of IP rights: copyrights (and related rights, for example *sui generis* database protection) which protect literary, artistic and scientific works; patents granted for inventions; trademarks used to identify the commercial source of goods or services; industrial designs, protecting the eye appeal of products; protection against unfair competition and trade secrets; integrated circuits topographies, geographical indications and plant varieties.[160]

Among all the IPRs, copyrights, database rights and trade secrets are most closely related to data. Patents can apply to computer implemented processes that manipulate and process data, but generally not in relation to data or software itself.[161] Trademarks can apply to data products (like indices), but

[159] A Kur and T Dreier, *European intellectual property law: Text, cases and materials* (Cheltenham: Elgar, 2013), p. 2.
[160] T Cook, *EU intellectual property law* (Oxford: Oxford Univ. Press, 2010).
[161] I.e, Patent EP1049993A1

again, generally not in relation to the actual data.[162] In line with this view, our analysis will focus on copyrights, the *sui generis* database right and trade secrets as it could be argued that these are the legal concepts that might influence data reuse activities in the EU most heavily, from an IP perspective.

The protection of personal data and IPRs will sometimes overlap. Personal data is defined as "any information relating to an identified or identifiable natural person". Taking this into account, copyright will rarely be a direct issue when dealing with data reuse in relation to personal data. However, personal data protection does not exclude the existence of some other IPRs, such as trademarks.

7.1.1 To what extent is data reuse affected by copyright law in relation to data?

Copyright is an important concern for data reusers. Any data analytics or data mining will often involve the wholesale copying of information or databases, all of which will be protected by IPRs in relevant jurisdictions.[163] Where data is not owned or licensed by the reusers, they will either need to abstain from using it or rely on an exception; otherwise they will risk violating a copyright.[164] In the online environment, copyright rears its ugly head mostly when talking about social media (Twitter, Facebook,...) and User Generated Content (UGC). UGC refers to "content generated by a non-professional user, without commercial purposes, direct or indirect and made available to the public or published through digital networks".[165]

UGC could manifest in three different ways:

- Completely original creations
- Adaptations or transformations of existing work
- Works which incorporate previously existing work

In line with the abovementioned, this UGC could comprise the vast majority of tweets and Facebook posts, should they be enough as to qualify for copyright protection.[166] According to the terms of service

[162] R Kemp, *Legal Aspects of Managing Big Data* (Kemp IT Law, 2014). Retrieved from http://www.kempitlaw.com/wp-content/uploads/2014/10/Legal-Aspects-of-Big-Data-White-Paper-v2-1-October-2014.pdf.
[163] R Graham and A Lewington, *The Big Data Explosion: A New Frontier in Digital Law* (SCL – the IT law community) <http://www.scl.org/site.aspx?i=ed31114> accessed 23 January 2016.
[164] *Obiter dictum*, it will not always be easy to determine who actually owns the data i.e. can claim rights in data. In the case of CERN, the particle physics research institute, run by multiple international partners, they were unable to resolve which partner in the consortium owns which data. This caused a mismatch in their expectations around IP generated by the scientific research. RECODE, *Legal and ethical issues in open access and data dissemination and preservation: Deliverable D3.1.* (2014), p. 12. Kemp believes those problems are related to the uncertain scope of IP rights and the fact that the law in this area will surely continue to develop in the coming years as big data gathers pace (*supra* 160). See also chapter 8 of this deliverable.
[165] "El Futuro de los Derchos de Autor y los Contenidos Generados por los Usuarios" (Rooter, 2011). http://www.rooteranalysis.es/documents/futuro_derechos_autor_contenidos_generados_usuarios_web_2.0.pdf
[166] In this sense, Copyright in Germany is governed by the *Kleine Münze* clause, which, in principle, states that not everything that is written is automatically copyright protected, and the criteria is not set on length, nor complexity, but on the creative

from Twitter (https://twitter.com/tos) and Facebook (https://www.facebook.com/legal/terms), the author of a post or picture shared on one of these social media platform remains the author and sole owner of the content, but grants a non-exclusive license to the service provider, in the case of original creations.

In case the work is derived from a pre-existing work, or if incorporates a pre-existing piece of work, the author remains the owner of the work, but bears the burden of proving that her post or picture does not infringe the rights of the author of the previous creation, baring as well the whole liability in case any claims are brought in relation to the new content. In relation to this, it will be important for data reusers to stay up to date with the claims received by their data providers, removing from their databases all data that has been removed from their data providers' databases.

A different approach is taken in most Common Law countries (except for the UK, which is highly harmonized with the rest of EU countries), where some data reuses could be covered under the fair use doctrine. According to the US law, "the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright" (17 US Code § 107).

There are four factors that must be measured in order to assess the applicability of the fair use doctrine:

- The purpose and character of the use, including whether such use is of a commercial nature or is for non-profit educational purposes.
- The nature of the copyrighted work.
- The amount and substantiality of the portion used in relation to the copyrighted work as a whole.
- The effect of the use upon the potential market for or value of the copyrighted work.

The fair use doctrine has been vital in deciding some data mining cases that could be important for data reusers, such as the Authors Guild, Inc. v. Google, Inc. case, or the Authors Guild, Inc. v. HathiTrust case, both focused on digitizing and usage of digitized books. In both of them, it was found that the unauthorised digitizing, as well as the search and usage of digitized books is copyright compliant under the fair use doctrine.

level and the identifiability of the personality of the author (e.g. "Good morning people!! :D How are things shakin'??" would not be copyright protected, whereas a short haiku or movie title may be protected, not based on length, but originality).

In the EU, some authors believe that the current setting is obsolete and that copyrights as we know them today (more precisely, the right of reproduction which is inherent to a copyright) should be transformed into a right to reuse. This would prevent the discrepancy between the social and legal norms that is most clearly seen on the Internet. For example, many activities (such as file sharing, remixing or creating mash-ups) are not perceived negatively by end-users even though they formally constitute a copyright infringement.[167] Hargreaves and Hugenholtz have proposed a simple amendment in the legal framework that would align the laws with the social reality: if a technical copy has no economic significance, it should not count as reproduction.[168]

Adapting copyright rules to the new reality is one of the priorities of the Juncker Commission. The European Parliament has recently adopted a non-legislative report on a copyright reform prepared by Pirate Party member Julia Reda. The report calls for an adaptation of the EU 2001 Copyright Directive to the digital market and establishes the basis for the upcoming copyright reform proposal by the EU commissioner for Digital Economy and Society.[169] In December 2015, the EC published the communication "Towards a modern, more European copyright framework", in which it emphasizes the need for higher harmonisation and adaptation of copyright rules to new technological realities. Among others, the communication includes a proposal for a simplified cross-border access to online content services and the regulation of online platforms, in particular news aggregators.[170] Despite all those proposed changes the reality is that copyright national barriers are still up and should be carefully considered by reusers.

7.1.2 To what extent is data reuse affected by copyright law in relation to software?

Copyright does not only protect books and music, but it also grants protection to computer programs through the national implementation of Directive 2009/24/CE on the legal protection of computer programs. Similar degrees of protection are granted all around the world.

The main problem with data reusers in relation to copyright and computer programs resides in the computer algorithms and software used to gather, compile, analyse, distribute or share the reused data.[171]

When speaking about software, there are two main ways to distribute it, either as proprietary/commercial/closed-source software or as open-source software, being crucial a correct

[167] Legal norms advocate for restricted use and reuse of information, while social norms advocate for the free circulation of knowledge on the Internet. Thus, many activities (such as the practices of file-sharing, remix or mash-ups) are not perceived negatively by end-users even though they constitute copyright infringement. P De Filippi and K Gracz, 'Resolving the crisis of copyright law in the digital environment: reforming the "copy-right" into a "reuse right"' (2012) 7th International Conference on the interaction of knowledge rights, data protection and communication, Helsinki, Finland.
[168] I Hargreaves and B Hugenholtz '*Copyright reform for growth and jobs'* (2013) Lisbon Council Policy Brief, 13/2013.
[169] <https://juliareda.eu/copyright-evaluation-report/> accessed 23 January 2016.
[170] COM (2015) 626 final, 9 December 2015.
[171] JP Montero, *Aproximación Jurídica y Económica al Big Data* (Tirant lo Blanch, 2015).

assessment of the terms of each license, to ensure that no copyright infringement is committed when using/implementing/embedding the software.

On the one hand, proprietary/commercial/closed-source software is easier to assess as, most of the time, a license will be subject to a prior payment of the software, existing different levels of license, depending on how the company wants to use that particular software. The main concern will be to secure the adequate license for the ends pursued in the project, to avoid possible later copyright related claims. A licence is nothing more than a contract between a licensor and licensee that defines the scope of activities a licensee may engage in with regard to the licensed database e.g., use the data solely for internal use, distribute limited segments to others, combine the database with other data, etc.[172] Terms and conditions or other forms of authorization may be deemed as equivalent of or as containing a licence.[173]

Many companies currently use software that is offered through open-source licenses, such as Creative Commons,[174] for which different standards of protection are applied. The most widespread Creative Commons[175] are divided as follows:

- **Attribution**: Is the most permissive license, allowing virtually any use of the software, as long as the author is recognised as such. All the following licenses require as well recognition of the author of the original work.
- **Share Alike**: This license requires the author to be recognised and the new work to be published under the same license. In terms of computer programs, where software is comprised of multiple programs, it would be required for that specific part to share the same license, but not the software as a whole.
- **No Derivatives**: In the case of this particular license, only the original work could be used, and always referencing the author as such. This does not preclude from the use of fragments of the work, as long as it is not modified and it is indicated.
- **Non-Commercial**: This license precludes from the use of the work with any commercial purposes, but allows personal and private use of the works. It does not require to share the new works with the same license, but it does require for the author to be acknowledged.
- **Non-Commercial Share Alike**: This license would allow only for the same uses as the "Non-Commercial" license, but it also requires for the derived works to be shared under the "Non-Commercial Share Alike" license.

[172] See *supra* 200, p. 91.
[173] LAPSI, *Policy Recommendation N. 2: The Interface Between the Protection of Commercial Secrecy and the Re-Use of Public Sector Information* (n.d.).
[174] <http://www.slideshare.net/North_Bridge/2015-future-of-open-source-study> accessed 23 January 2016.
[175] <https://creativecommons.org/licenses/?lang=en> accessed 23 January 2016.

- **Non-Commercial No Derivatives**: This is the most restrictive license provided by Creative Commons and it only allows for the works to be downloaded and shared, crediting the author. No modifications could be done to the original work.

There are other open licenses being used around the world, mostly relating to software, such as the Bouncy Castle license, the Jason Hunter OSS license, or the Mozilla Public License; each of them granting different degrees of permissions.

As a result of the great number of different closed and open-source licenses available, each license should be examined individually before deciding to use any specific software for gathering or processing data in order to avoid a possible copyright infraction that might imply a change in the technical or even basic structure foundation of the project.

7.1.3 To what extent is data reuse affected by database protection law?

The concept database, according to Article 10(2), Agreement on Trade-Related Aspects of Intellectual Property Rights, applies to "Compilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations".

Article 5 of the World Intellectual Property Organization (WIPO) Copyright Treaty, establishes that these databases shall be protected under copyright, inasmuch they can be considered intellectual creations or works. This protection granted to databases is not extended to its contents, but "shall be without prejudice to any copyright subsisting in the data or material itself".

But databases are also protected under a *sui generis* right, set forth by Directive 96/9/EC. The goal is to protect the content of databases that is not protected under copyright or data protection laws, but that amounts to a substantial investment, in time or money, for the collecting, verifying and presentation of the data (not the creation of the data themselves). Since its adoption by means of the Database directive, the *sui generis* right has received much criticism, including some negative feedback from the CJEU. Although the Court limited the scope of the database right in its judgement in Case C-203/02 *The British Horseracing Board Ltd and Others v William Hill Organization Ltd*[176]*, sui generis* right is still considered a barrier to data reusers. Hargeaves and Hugenholtz claim the right is especially obstructing data mining and big data analytics.[177] Despite the critics, the database right is still fully applicable and data reusers should consider it carefully to avoid breaching IP law.

Whatever the case, if the data reuser intends to extract data from a database, there is as well the chance of it being subject to an open license, which is applicable in the same terms as explained above. If there is not such a license available for the database, an agreement will have to be reached with the author of the database or an authorized third party, in order to avoid legal exposure from this front.

[176] [2004] ECR II-2905
[177] I Hargreaves and B Hugenholtz 'Copyright reform for growth and jobs' (2013) Lisbon Council Policy Brief, 13/2013.

7.1.4 To what extent is data reuse affected by trademark law?

Trademarks, under their current configuration in EU law, as well as most jurisdictions, are not directly relevant to the gathering and processing of data, necessary for data reuse.

Trademarks become relevant for data reuse when the results of data processing are made public. In those cases, the provisions of unfair competition (passing off, in some countries) determine the way in which data should be exploited. Issues tend to arise when marketing or selling the resulting product and reference is made to a company that has been part of the process to some extent.

To comply with trademark and unfair competition regulations, it is best to seek contractual consent in the data transfer agreement or through a later license if there is any intent to use any trademarks when marketing the final product, if there is any.

7.1.5 To what extent is data reuse affected by patent law?

Patents are granted to inventions that are proved to be novel, usable and industrially applicable, so the owner can exclude others from selling or offering the invention without consent, for a term of, typically, 20 years. According to the current state of Patent Law practice, patent law should not be a major issue for data reuse, mainly after the rejection of the proposal for a Directive of the European Parliament and of the Council on the patentability of computer-implemented inventions (Commission proposal COM (2002) 92).

As patentability of software in itself, mathematical methods and formulae (among others) is not permitted under Article 52(2) of the current 1973 European Patent Convention, the interference of patent rights on data reuse is limited, and only when related to some specific hardware solutions with implemented software.

At this moment in time, patentability of algorithms that solve technical difficulties in the management, exploitation or harvesting of data, when implemented in a computer, is the main source of difficulties for a data-based company; but it could be patented if it complied with the three requirements of a patent: novelty, involve an inventive step and being susceptible of industrial application; the point at which most data-related patents applications fail.

However, there is still uncertainty on whether the upcoming Unified Patent will bring along a wave of software patent applications and the criteria that will govern the Unified Patent Court on the matter.[178]

Patents are configured under the principle of territoriality, meaning that a patent can only be enforced in the countries on which it has been officially registered (or has, at least, applied for registration). In this sense, the law in countries outside of the EU is not so uncommon to find patented computer-

[178] <http://www.the-european.eu/story-4148/software-patents-and-the-european-unitary-patent-a-commercial-view.html http://www.theguardian.com/technology/2011/aug/22/european-unitary-patent-software-warning> accessed 23 January 2016.

implemented processes in the United States as early as 1968, with US Patent No. 3,380,029, relating to "data processing systems and particularly to a system for sorting large quantities of data or records".[179]

Following the idea of territoriality, even though there are quite a few US patents that provide or foresee data reuse functionalities,[180] these are not binding in the EU, so, unless there is a patent registered in the Member State on which the data wants to be reused, patents should not be an issue for the data reuser.

In the unlikely scenario of a software patent having been registered in the Member State in question, the data reuser will have to countersue the patent owner for the invalidity of the allegedly infringed patent and/or prove that his product does not infringe the patent rights under one of the two tests available:[181]

- Test of the equivalents:
    - Do the inventions have the same function?
    - Would a person, skilled in the state of the art in question, have realised at that point in time that the changes included would lead to the same result?
    - Would the skilled in the art person consider, in light of their claims, consider both patents equivalent?

- Triple identity test:
    - Do the inventions have the same function?
    - Do they perform it in the same way?
    - Do they achieve the same result?

7.1.6 Trade secrets

Another relevant aspect of Intellectual Property law concerned with data reuse is trade secrets. Even though the term is pretty much self-explanatory, the TRIPS Agreement sets forth the fundamental requirements for commercial information to be considered as trade secrets on article 39.2: the information must have been kept in secret, it must have commercial value because of its secrecy and reasonable steps must have been taken to ensure its secrecy.

[179] <http://ipwatchdog.com/patents/US3380029A.pdf> accessed 23 January 2016.
[180] Among others, US Patent No. 8700000 or US. Patent No 7206400.
[181] These two approaches may be considered under different names depending on the country on which enforcement of the patent is sought, but most countries recognize at least one of the two options. In Spain, for example, the test of equivalents is known as “Obviousness test”, “Catnic/Improver test” or “Inventive step test”; whereas the triple identity test is known as the “triple substantial identity test” or the “triple equivalency test”.

Contrary to the US federal legal system[182], there is no legislation on the EU level yet that focuses specifically on trade secrets**.** However, in November 2013, the EC proposed a draft directive that would align existing laws against the misappropriation of trade secrets across the EU.[183] The proposal has been recently handed over to the Parliament to continue the regular legislative procedure. Until the common EU approach is adopted, the area remains regulated by the applicable international and national legal acts. At the international level, trade secrets are addressed by the fundamental agreement in the area of IP law – TRIPS, which many Member States apply directly on their legislations in order to set the concept and basic requirements of a trade secret.[184]

The currently available draft of the Trade Secrets Directive proposal,[185] establishes in recital (8a) that a definition of trade secret should "cover business information, technological information and know-how where there is both a legitimate interest in keeping confidential and a legitimate expectation in the preservation of such confidentiality". Considering how wide the concept of trade secrets is, and that it is in line with the definition set forth on the TRIPS Agreement, there are two main ways through which it could affect data reuse: data itself and the algorithms/computer programs that are being used.

On the one hand, from the perspective of the algorithms and computer that could be used in an unauthorized way as a result of a trade secret infringement, it is important to keep in mind that, in the same sense already stated above in the copyright sections, algorithms and computer programs can be susceptible of ownership and license. If a company makes sufficient efforts to keep those hidden and they are of commercial value, the usage of those computer programs and algorithms without the corresponding license could amount to a trade secret infringement, aside from the corresponding copyright infringement, if applicable.

On the other hand, commercial data from a company can be deemed a trade secret as well. Examples of commercial-related information that may be deemed as trade secret (when sufficient secrecy-preserving measures are in place and the information has enough commercial value) include formulae, production methods, organisational methods, know-how or even details of the personnel contracts.

The general idea behind trade secrets protection is to avoid that commercially valuable information that is being secretly kept by a company, is stolen and used to obtain an unfair advantage. In this sense, data reusers must be firmly aware of the sources from which they obtain their software and data, in order to

[182] In the United States, a federal source of law, the Uniform Trade Secrets Act, played a vital role in harmonizing the legal protection of trade secrets across the different US states. In the EU this development has just begun. See for example K A Czapracka, Antitrust and Trade Secrets: The US and the EU Approach, (2007) 24 Santa Clara High Tech. L.J. 207.
[183] <http://ec.europa.eu/growth/industry/intellectual-property/trade-secrets/index_en.htm> accessed 23 January 2016.
[184] e.g.: In Spain this principle has been used repeatedly. Among others, please see SSAP Madrid 23-III-2012 (SAP M 4877/2012), 14-X-2011 (SAP M 15159/2011), 25-II-2011 (SAP M 1578/2011), Barcelona 25-VI-2013 (SAP B 7226/2013), 16-V-2012 (SAP B 9858\2012), 20-IV-2011 (SAP B 11291/2011), Vizcaya 9-II-2011 (SAP BI 1586/2011), Alicante 29-I-2010 (SAP A 471/2010) and Zaragoza 17-V-2010 (SAP Z 2152/2010).
[185] Doc. 15382/1/15 REV 1 <http://data.consilium.europa.eu/doc/document/ST-15382-2015-REV-1/en/pdf> accessed 29 January 2016.

avoid claims under the trade secrets regulations of their respective countries. If the proposal for a Trade Secrets Directive, sees the light, the current scenario will be greatly simplified. However, for the time being, each data reuser will have to be aware of her national legislation in order to avoid trade secret infringement claims.

Table 15 Compliance checklist for the EU intellectual property law

Compliance checklist
• Check existing patents for a data-reuse-based solution or process prior to any software implementation, to reduce the risk of patent claims. • Avoid trademark infringement related claims by careful exploitation of processing results. • Ensure that database licenses allow for data reuse and that warranties are in place to secure that all content of the database has been obtained legitimately. • Be aware that open-license is a possibility. • Most databases will be protected, either under copyright, sui generis right or trade secrets. • Ensure that all implementations foreseen (embedding, customisation,…) are compliant with that database licenses on open-source software. • Secure licenses on all non-open-source software used within the project. • Rely on internally developed software or customized software whenever possible. • If copyrightable material is being used, a notice and take down system, which would allow copyright holders to notify data reusers and database owners of an infringement should be implemented, when dealing with copyrighted (or copyrightable) data. • Data reusers must secure that all data has been subject to appropriate licenses. • Not all data can be treated in the same way (understand which data us personal/non-personal, copyrighted/non-copyrighted).

7.2 Intellectual Property law from the perspective of the EuDEco model

7.2.1 Assessment of IPRs in relation to socio-economic propositions and business models

Table 16 Socio-economic propositions and business models related to IPRs

Socio-economic challenges	***Legal response – intellectual property law***
Data users need to be able to determine whether the data were trustworthy (e.g. through certificates, open data documentation).	Trustworthiness of data does not only include the reliability or the veracity of content, but also the legality and legitimacy of the data, to ensure that any future use will not be interrupted by an Intellectual

	Property claim. It is key to provide copyright holders with a way to control the use of their work, as well as data reusers with a tool to assess the legality and legitimacy of any data used within a project, what could be achieved by the updating of copyright related laws to the new digital environment.
Best practices in experimental methods and in the storage, archiving, and dissemination of experimental data should be applied.	The control of copyright holders over their works should be strengthened, according to the current position of the EU Digital Economy Commissioner. This task proves to be difficult if interpreted as a closed mandatory solution, as the proliferation of open licenses in the software industry seems to prove. An opt-in/opt-out system could be ideal, giving legal basis to the current de facto structure of the digital society.
Traditional data sources such as company databases and applications are now complemented by *non-traditional sources* such as social media or sensors embedded in physical world devices including mobile devices, smart meters, cars and industrial machines.	The protection of databases within the EU covers both, databases comprised of copyrighted and non-copyrighted material. In this sense, the rights of database authors are sufficiently addressed by EU Law, even though a further development of exceptions might be needed, since the tightness of the system encourages data collection and use, but discourages data reuse.

7.2.2 CAS

Intellectual Property has always been in close relation with innovation, technology and society, due to its inherent relation to culture and knowledge. In this sense, not all legal disciplines have the same relevance in the societal, technological and economical aspects of data reuse, which is yet another piece of evidence underlining that the law as well as the data economy is a CAS.

A neutral approach towards the relation between intellectual property and technology would reveal that they could be both allies and enemies, depending mostly on societal and economical aspects. The clearest example would be peer-to-peer technologies, which could be perceived as economic boosters, enabling the instant communication and fast transmission of files through the Internet (as many console videogames use it), generally including anonymity protocols; or as a threat for the very same videogame industry, due to piracy through tools such as eMule or µTorrent.

In this sense, countries in which royalties are paid to authors for their works, prices for services are considered fair and good quality services are present, tend to choose legitimate methods to acquire copyrighted materials (streaming platforms, video-on-demand platforms…), rather than piracy[186]. For example, collective management societies (entities entrusted with the management of author's royalties for their works) could be perceived as enemies if they ask for a very high royalties rate, as well as a strict tax policy could discourage users from acquiring legitimate works and opting for piracy.[187]

Considering all of the above, it is clear that intellectual property is in itself a highly complex environment and that policymakers have to deal with many circumstances when regulating this legal area.

8 Ownership of data, big data contracting and their role in the model

8.1.1 Property in (personal) data

In economics, granting property rights is often suggested as a solution to the incentive problems related to free riding. The concept of ownership typically means *"to have legal title and full property rights to something*".[188] Data are an intangible asset; like other information-related goods, they can be reproduced and transferred at almost zero marginal costs. So in contrast to the concept of ownership of physical goods, where the owner typically has exclusive rights and control over the good – including, for instance, the freedom to destroy the good – this is not the case for intangibles such as data. For these types of goods, IPRs are typically suggested as the legal means to establish clear ownership.[189]

The belief that nobody can own data has been challenged, given that the access and use of big data is turning to be a key for future business success in many sectors.[190] In spite of the arguments against data ownership in the narrow (economistic) sense, data ownership has become a buzzword.[191]

Privatisation, propertarisation and commodification of data has been also addressed in the EUs Digital Agenda plans, as it is a significant issue for the European digital strategy, particularly in relation with personal data protection.[192]

---

[186] <http://www.musicbusinessworldwide.com/piracy-virtually-eliminated-norway/> accessed 23 January 2016.
[187] <http://elpais.com/elpais/2015/02/25/inenglish/1424880255_595664.html> accessed 23 January 2016.
[188] A Kocharov, 'Data Ownership and Access Rights in the European Food Safety Authority' (2009) 4 European Food & Feed Law Review 5, p.335. Data owner indicates the entity which has the legal rights in relation to the data and takes decisions on how to use, store, disclose and share them.
[189] Furthermore, technologies such as cryptography have dramatically reduced the costs of exclusion, and thus are often used as a means to protect data. OECD, 'Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by 'Big Data'' (2013) OECD Digital Economy Papers No. 222, OECD Publishing.
[190] <http://www.twobirds.com/~/media/pdfs/brochures/information-technology/big-data.pdf> accessed 23 January 2016.
[191] Particularly after the Commission revealed their plans related to further regulation of data ownership and liability.
[192] <http://ec.europa.eu/digital-agenda/en/open-data-0> accessed 23 January 2016.

The discussion on propertarisation of personal data is linked to the emergence of new business models in the Internet and social media, in which users get services for free, but "pay" with their personal data, often without their knowledge. Personal data is often described as the new oil and rivalrous good, which makes it more tangible and explains the relevance of the property discourse.

The idea of propertarisation of privacy rights emerged in the US in the 1970s and came to the EU in the 2000s.[193] Property in personal data refers to the entitlement to exclude the other from personal data by default. Put differently, the default entitlement in favour of the individual implies that there is no disclosure, collection or use of personal data by default. [194]

The propertarisation theory was not intended to be an end in itself, but to offer (mostly, to American authors) a theoretical framework for legal solutions that would strengthen personal rights, in particular data protection rights. Under Lessig's proposal, consumers who would hold the original property entitlements to their own personal information would be able to bargain with data users to determine when it would be advantageous to forfeit their privacy by selling their data.[195] Also based on the idea of propertarisation, Schwartz proposed a default rule requiring consumers to "opt in" to any use of their information or a "right of exit" from existing agreements to data processing, to "prevent […] initial bad bargains from having long-term consequences." Individuals always maintain the ultimate entitlement to their own personal data and may not forfeit their rights through a contract. In addition, Schwarz encouraged enforceable rights against third parties. Creating burdens that bind third parties is a typical characteristic of property.[196]

In the EU, Purtova fiercely argues that maintaining the status quo where no ownership in personal data is formally assigned equals assigning ownership to the information industry and leaving an individual defenceless in the face of corporate power eroding his autonomy, privacy and informational self-determination.[197] Following that reasoning, Purtova criticizes the EU data protection law's approach, e.g. restricting the scope of data subject's consent as the legal basis for data reuse.

Jakob, in contrast, believes that the EU data protection plan takes for granted that personal data has become akin to a commodity capable of changing hands, although the use of property-derived rights is particularly unusual and significant since a human-rights-based approach to privacy, which the EU

[193] Nadezha Purtova has published a few in-depth research person on the topic and have argued in favour of the idea of propertarization. She considers propertarization as a viable solution for the drawbacks of the current data protection law. N Purtova, *Property Rights in Personal Data: a European Perspective* (Alphen aan den Rijn: Kluwer Law International, 2011).
[194] N Purtova, 'Illusion of personal data as no one's property' (2015) 7 Law, Innovation and Technology 1.
[195] L Lessig, 'Privacy as property - Part V: Democratic Process and Non-public Politics', Social Research, Spring 2002. <http://www.englishdiscourse.org/lessig.html> accessed 29 January 2016.
[196] PM Schwartz, 'Property, Privacy, and Personal Data' 117 Harvard Law Review 7, p. 2055.
[197] *Supra* 190.

generally embraces.[198] In particular, he sees property features in the newly-established right to data portability and the right to be forgotten.

8.1.2 (Big) data contracting

A data set will typically embed complex assignments of different rights across different data stakeholders, e.g. personal data and IPRs. This will grant stakeholders "the ability to access, create, modify, package, derive benefit from, sell or remove data, but also the right to assign these access privileges to others".[199] As we could see from the discussion on property in data above, in many cases, no single data stakeholder will have exclusive rights over data. Consequently, it will not always be easy to determine to whom certain rights apply, or, to use vague language, who owns the data.[200] This will become even more challenging when the relationship between data creators and users will include complex contractual agreements.

While contractual duty only applies between the parties of a contract (*intra partes, in personam),* property law imposes obligations on all the parties regardless of their relations *(erga omnes, in rem)*[201]. This will often mean a stronger protection, since a data owner will be able to claim a legal entitlement against third parties and not only the party of the contract. Nonetheless, owing to the uncertainty of its ownership under applicable IP law, the ownership in relation to data can be, and is most often, designated by and protected by contract.[202]

(Big) data contracting, including licensing, can be one of the greatest challenges for data reusers. Contractual arrangements in the data business are diverse and complex, as illustrated by the list of

[198] JM Victor, 'The EU General Data Protection Regulation: Toward a Property Regime for Protecting Data Privacy' (November 4, 2013) 123 Yale Law Journal 513. Similarly, "Incomplete Commodification in the Computerized World" *in The Commodification of Information*, edited by N Elkin-Korin and NW Netanel, 3-22, Information Law Series, 11 (The Hague, Netherlands: Kluwer Law International, 2002). "It makes a big difference whether privacy is thought of as a human right, attaching to persons by virtue of their personhood, or as a property right, something that can be owned and controlled by persons. . . . Human rights are presumptively market inalienable, whereas property rights are presumptively market-alienable."

[199] D Loshin (2002) 'Knowledge Integrity: Data Ownership'. <http://www.datawarehouse.com/article/?articleid=3052> accessed 29 January 2016.

[200] There are plenty of entities that can be possibly involved in the ownership of one data set: *creator* – the party that creates or generates data, *consumer* – the party that uses the data, *compiler* – the party that selects and compiles information from different information sources, *enterprise* – the legal entity that enters or creates the data within their own enterprise, *funder* – the user that commissions the data creation , *decoder* – in environments where information is 'locked' inside particular encoded formats, the party that can unlock the information , *packager* – the party that collects information for a particular use and adds value through formatting the information for a particular market or set of consumers, *reader as owner* – the value of any data that can be read is subsumed by the reader and, therefore, the reader gains value through adding that information to an information repository, *subject as owner* – the subject of the data claims ownership of that data, mostly in reaction to another party, *purchaser/licenser as owner* – the individual or organisation that buys or licenses data. *Supra* 185.

[201] *Jura in personam* are rights primarily available against specific persons. *Jura in rem* are rights only available against the world at large. Black's Law Dictionary (1910).

[202] D Glazer, H Lebowitz and J Greenberg, *Data as IP and Data License Agreements* (Practical Law Publishing Limited and Practical Law Company, Inc., 2013)

various players in the data value chain.[203] First and foremost, the dataset may consist of data that has been generated by the licensor itself, collected from users and other third parties, licensed from third parties, scraped from the Internet and/or obtained via various social media tools. Therefore, the key issues include a compromise on the ownership of the rights derived from the data, the scope of data use and reuse, the warranties of compliance with laws and regulations, duration of the relationship and risk allocation.[204]

Given the heterogeneity and the complexity of big data, it is of utmost importance that the licence agreement addresses all (or at least the majority of) possibly disputable issues to avoid lack of protection, unexpected threats and failure to secure appropriate revenue. This is preferably done before reusers enter in a written contract with their data suppliers. The type of the agreement will, however, depend on the fact how this specific data fits into their business plans. If they want to acquire ownership of the data, they use a data supply contract. If they want to determine the scope of the right to resell the source's product using the data broker's brand, they enter in a data reseller agreement.[205] Also, the agreement will depend on the negotiation power of the parties. European SMEs have been complaining by the unfavourable conditions Twitter imposes on those who want to reuse their data by using Google Analytics. For instance, terms and conditions prohibit storing any data from APIs, which, *de facto*, disables an effective data reuse.[206]

Proper evaluation, protection and ownership identification of data will be also essential in crisis and insolvency situations; evaluating and assessing ownership of data is an essential step for determining the value of a company or of its assets in case of bankruptcy.[207] To whom the data is assigned will be critical, since the client, unless the ownership of data had been clearly defined in the contract, may risk losing the data stored in the cloud. The problem amplifies when personal data is involved.

Table 17 Compliance checklist in relation to data contracting

Compliance checklist
- Always address the vital aspects of a relationship in the licence agreement: a. acknowledge vendor's/costumer's rights in the data received from the licensee b. provide an appropriately tailored definition of the licensed data set c. acknowledge if the licensor has expended significant resources gathering, assembling and compiling the data

[203] Also see: N Kranjec, P Merc and B Koritnik, 'Startup pred pravnimi izzivi (Startups facing legal challenges)', Pravna praksa, 2015, issue 42-43, p. 27.
[204] JR Kalyvas and MR Overly, *Big data: A business and legal guide*. *An Auerbach Book* (Boca Raton, Fla.: CRC Press, 2014), p. 92.
[205] *Ibidem*, p. 16.
[206] A Arrigo, 'Data reuse: can you really do it?' (*LinkedIn Pulse,* June 2015) <https://www.linkedin.com/pulse/data-reuse-can-you-really-do-alessandro-arrigo?trk=prof-post> accessed 23 January 2016.
[207] *Supra* 186.

d. acknowledge that the data is original and IPR protected, or the data is a trade secret
e. define the permitted use
f. address the issues of exclusivity, sublicensing, transfers of data

- Choose the form in which the rights in the derived data are retained: the first option is an agreement that a party retains ownership of derived data and the second is a licence through which the party obtains the rights from the initial right owner. It should be borne in mind that a licence gives the reuser some advantages over third parties, which is not the case in a contractual relation.

8.1.3 Big data contracting and related socio-economic and technological propositions

Table 18 Socio-economic and technological propositions related to data contracting

Technological propositions		***Legal response - contracts***
Data collection and delivery		
Stability of interfaces	Delivering or receiving data between partners requires high stability of their interfaces. E.g. big players such as Google or Amazon are constantly changing their APIs without the need to discuss those changes.	Agreement on interfaces represents an important aspect in the relation between data creators and reusers. Unfavourable contractual terms, standard terms and conditions drafted with a bias limit data reusers and hinder their innovation.
Legacy systems	Collecting data is a case-specific exercise as data can have many diverse origins. It is also a time-consuming process.	Not only is data collection a time-consuming process, it can also have serious legal consequences. In order to avoid future dispute, the origins of data will need to be clarified in a contract between the data owner (data creator, data reseller) and data reuser.

Technological responsibility		
e.g. Data loss	Most cloud storage providers therefore use data replication techniques to avoid this.	Data reusers can be contractually liable for data loss. Acting in an irresponsible manner, e.g. non providing sufficient technical support, can have some serious legal consequences depending on the contractual provisions.
Performance unpredictability	The performance of the infrastructure is particularly important if applications are time critical.	The contract can help data reusers avoid the non-performance or at least to be fairly compensated.
Security and compliance	Security and regulatory compliance are key issues if technological responsibility is transferred to third parties. Cloud computing!	Ideally, data protection and cybersecurity requirements should be adequately addressed in the contract. The contractual agreement cannot rule out the statutory requirements for security and privacy.
Availability		Typically, this issue will be addressed in a service level agreement (SLA) (e.g. agreement between a provider of a cloud service and their costumers).
Socio-economic propositions		***Legal response - contracts***
Licensing needs to be simplified		A licence can be simplified to a certain degree. However, it must still be comprehensive enough to allow that all the critical issues are properly addressed, e.g. the rights in relation to derived data. Model licences such as Creative commons (CC) can be a useful tool. For the use and reuse of musical works, the EU has simplified territorial licensing by introducing a multi-territorial licence, which can be invoked across the EU. The directive that has introduced this option will get in force in April 2016.

Data users need a possibility to assess the relevance of data and they need to be able to determine whether the data were trustworthy (e.g. through certificates, open data documentation).	Apart from a limited number of industries where the law itself requires a stricter level of data quality, international data standards and certificates, e.g. ISO, can be used to ensure data trustworthiness. Adherence to the standards can give leverage in the big data negotiations.

8.1.4 CAS and contractual agreements

Big data contracting reveals a high number of complex relations and sometimes conflicting interests of the stakeholders that are involved in the data reuse activities. When parties negotiate an agreement they need to keep in mind the technological limits and economic aspects of data reuse. For instance, if a contract authorizes a reuser to access the data, but it does not specify an interface or even allows for changing interfaces, then the contract has no real teeth. Namely, the data owner will simply change the interface and, unless the reuser is able to implement the new system, the access to data will be closed.

9 Public sector information/freedom to information requirements and their role in the model

9.1 Public sector information/freedom to information requirements – overview of the legal framework

Table 19 Overview of the legal framework for the EU PSI regulations

EU PSI LAW	
Primary EU law	European Convention on Human Rights (Article 10) Charter of the fundamental rights of the EU (Article 42) Treaty on the functioning of the EU (Art 15 (1))
Secondary EU law	Regulation (EC) No 1049/2001 of the European Parliament and of the Council of 30 May 2001 regarding public access to European Parliament, Council and Commission documents Directive on PSI Directive 2003/4/EC of the European Parliament and

Open data, as we understand it within the EuDEco consortium, refers to open governmental records, which may be reused to create value for citizens as well as for businesses. Several other terms are associated with the notion of open data e.g., open government, freedom to information and right to access public sector information (PSI).

International courts see access rights as part of, or closely related to, the right to freedom of expression. However, access rights are also recognized in case law of the European Court of Human Rights in the context of the right to private life. By contrast, access rights may be conceived of as stand-alone constitutional rights.[208]

In the EU, the Directive 2003/98/EC on the reuse of PSI intended to harmonise the Member States' legislation in order to open the data to public access and encourage its reuse. The directive requires Member States to make as much information available for reuse as possible. It addresses material held by public sector bodies in the Member States, at national, regional and local levels, such as ministries, state agencies, municipalities, as well as organisations funded for the most part by or under the control of public authorities e.g. meteorological institutes. In 2013, the PSI Directive was revised and its scope was extended to cultural institutions such as libraries (including university libraries), museums and archives (PSI Directive, Article 3). In 2014 the EC issued a Notice, which provides a non-binding guidance on the best practices in relation to PSI's licencing, charging and formats.[209]

Before the directive was amended in 2013, the EC launched LAPSI, an international research initiative, which brought together EU legal experts to identify the remaining legal obstacles to access and reuse of PSI on the European content market, and to propose measures and tools to stimulate the progress of the European market towards open data. The project identified a number of important legal barriers that impact PSI reuse, as well as made some needed recommendations. The highlighted topics were the protection of IPRs when embedded in PSI, competition law restrictions in case of owning the data that may be used to compete on the market, and personal data protection.[210] Furthermore, the researchers pointed out that non-harmonized national laws were blocking data reuse.[211] Some of the findings of the

[208] FJ Zuiderveen Borgesius, M van Eechoud and J Gray, 'Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework' (November 24, 2015) Berkeley Technology Law Journal, Forthcoming, p.17.
[209] European Commission, INFORMATION FROM EUROPEAN UNION INSTITUTIONS, BODIES, OFFICES AND AGENCIES, EUROPEAN COMMISSION, COMMISSION NOTICE, Guidelines on recommended standard licences, datasets and charging for the reuse of documents (2014/C 240/01).
[210] The project's website can be accessed via: http://www.lapsi-project.eu/.
[211] instance, in Slovenia, the PSI directive was implemented in 2003 and since then the public bodies have given open access to their administrative decisions, however, due to improperly written national guidelines or the absence of the European ones, the decisions can only be available in PDF format, which makes them rewritable and hard to be reused. O Salamanca and M van

research project have been incorporated in the amended directive e.g. the cap on cost charged for the access and reuse of information (PSI Directive, Article 1 (6)).[212]

The open data policy is also part of the EU digital agenda.[213] In April 2015 the EC adopted the Digital Single Market Strategy and proposed two action points that directly relate to open data. First, it committed to launching the pan-European open data portal and second, it announced the review of the EC's decision on reuse. Moreover, in the Big data communication[214], the access to public data was described as one of the funding blocks of the European data economy. For the time being, no additional legislative measures are expected on the EU level. The EC only plans to perform a 'taking the pulse' exercise rather than starting with a preparation for a new policy initiative.

Due to many positive side effects such as transparency and trust, open data initiatives have been slowly moving towards the private sector.[215] Admittedly, open data as a concept is unlikely to be seen attractive for the private businesses.[216] Still, opening up at least a minimum amount of private data could result in great social benefits and by giving the business sufficient incentives it should be indeed possible to achieve greater openness in the private sector as well.

Table 20 Compliance checklist for the EU PSI regulation

Compliance checklist
Business: • Check under which conditions the public data is accessible (time for the public body to answer, financial compensation that they can ask for) and whether it is reusable • Be aware of the measures the law grants data reuse if the request for data is unduly rejected (e.g. appeal to a higher instance)

Eechoud, (2014) *Open Legal Data for Europe: LAPSI/Openlaws Workshop*. <http://www.openlaws.eu/wp-content/uploads/2014-09-02_LAPSI-Openlaws_workshop-Amsterdam.pdf> accessed 23 January 2016.

[212] Now the maximum charge is set on the level of marginal cost that incur due to a specific request (cost of reproduction and dissemination but no more cost of collection and return on investment). For a detailed analysis see LAPSI position paper N. 1 Principles governing charging for reuse of PSI. LAPSI. (n.d.a). *Policy Recommendation N. 1: The Competition Law Issues of the Re-Use of Public Sector Information (PSI).*

[213] *Supra* 188.

[214] Commission (EC), 'Towards a Thriving Data Driven Economy' (Communication) COM (2014) 442 final, 2 July 2014.

[215] <http://blogs.worldbank.org/voices/next-frontier-open-data-open-private-sector> accessed 23 January 2016. Also, EU data forum key note speech .

[216] L Reggi (2011). *Open Data to the next level: WHY and HOW to involve the private sector*. Retrieved from <http://luigireggi.eu/2011/09/19/open-data-to-the-next-level-why-and-how-to-involve-the-private-sector/> accessed 23 January 2016.

Public authorities:

- Be aware of the limitations of data sharing and reusing (e.g. intellectual property rights, data protection rights)
- Be able to respond in time and do not inflict excessive charges

9.2 PSI regulation from the perspective of the Eudeco reuse model

9.2.1 Assessment of the open data related socio-economic and technological propositions

Table 21 Socio-economic and technological propositions related to the EU PSI regulation

Technological propositions	***Legal response - PSI rules***
Scalability and data management	
Many data types, low throughput, large number of formats, difficult processing, difficult representation	Pursuant to Article 5 of the PSI directive, sector bodies shall make their documents available in any pre-existing format or language, and, where possible and appropriate, in open and machine-readable format together with their metadata. The EC provides a more detailed guidance on the recommended formats in Section 3 (Guidelines on datasets) of its Notice. Given that strong policy nudge, public sector bodies as well as the users of PSI should be able to overcome the challenges related to data scalability and transmission.
Security and privacy aspects	
	Security and privacy regulation is fully applicable to the data held by public bodies.

Socio-economic propositions	***Legal response - PSI rules***
Data sharing/management policy pressure is needed to foster the opening of data	The EU puts pressure on the Member States by requiring them to adopt national implementation laws. Those who miss the deadline become parties of the infringement procedure which may lead to high monetary fines.[217]
Compensations for efforts need to go beyond reputation; data reuse has to be rewarded, IP has to be ensured for the data producers	The typical way to reward the data producers is through licensing. Article 8 of the PSI directive provides a possibility to licence the data when appropriate: "Public sector bodies may allow re-use without conditions or may impose conditions, where appropriate through a licence. These conditions shall not unnecessarily restrict possibilities for re-use and shall not be used to restrict competition." The EC provides a more detailed guidance on the licensing in section 2 (Guidelines on recommended standard licences) of the Notice.

[217] The formal infringement proceedings against 17 EU Member States which have not yet implemented the amended directive on PSI began in September. <https://ec.europa.eu/digital-agenda/en/news/open-data-commission-launches-infringement-cases-due-late-transposition-revised-psi-directive> accessed 23 January 2016.

10 Cybersecurity and its role in the model

Table 22 Overview of the legal framework related to the EU cybersecurity law

EU CYBERSECURITY LAW	
Primary EU law	Charter of the fundamental rights of the EU (Article 7 and 8) Treaty on the functioning of the EU (Art 16) International agreements
Secondary EU law	Network and Information Security Directive (EC proposal)
Self-regulatory measures	ISO standards e.g. ISO/IEC 27032

10.1 Cybersecurity – overview of the legal framework

As Hanover Research established based on their US experience, cybersecurity law has been a growing field of legal practice and a priority for legal practitioners.[218] In the EU, the regulation of the field has been fragmented, ranging from economic internal market elements, fundamental rights and citizens' freedoms to criminal cooperation and defence policy.[219]

In 2013 the EC set the EU cybersecurity roadmap with the Communication on Cybersecurity strategy for the EU. Primarily, the EC plans to realize the strategy with a Network and Information Security (NIS) Directive, proposed in early 2013. [220] In December 2015 the EU institutions reached agreement on the final text, which still needs to be formally approved by the European Parliament and the Council.[221] Once approved, the Member States will then have 21 months to implement the directive into national law.[222]

The proposed directive aims to improve the security of the Internet and the private networks and information systems underpinning the functioning of our societies and economies. This will be achieved by requiring the Member States to increase their preparedness and improve their cooperation with each

[218] Hanover Research, The Emergence of Cybersecurity Law, Indiana University Maurer School of Law, February 2015. <http://info.law.indiana.edu/faculty-publications/The-Emergence-of-Cybersecurity-Law.pdf> accessed 23 January 2016.
[219] NK Tsagourias and R Buchan, *Research handbook on international law and cyberspace*, 1983-[2015] p. 424.
[220] Commission (EC), "Cybersecurity Strategy of the European Union: An Open, Safe and Secure Cyberspace" (Communication) JOIN(2013) 1 final , 7 February 2013.
[221] <http://www.europarl.europa.eu/news/en/news-room/content/20151207IPR06449/html/MEPs-close-deal-with-Council-on-first-ever-EU-rules-on-cybersecurity> accessed 23 January 2016.
[222] <http://www.lexology.com/library/detail.aspx?g=ea02c5cf-6138-4664-aec0-490405de2f91> accessed 23 January 2016.

other, and by requiring operators of critical infrastructures, such as energy, transport, and key providers of information society services (e-commerce platforms, social networks, etc.), as well as public administrations to adopt appropriate steps to manage security risks and report serious incidents to the national competent authorities.[223]

In order to understand the scope of the NIS directive and its relevance for data reuse, it is first necessary to define a network and information system, which represents the directive's subject matter. According to Article 3, the term covers:

- an electronic communications network within the meaning of the Framework Directive 2002/21/EC,
- any device or group of inter-connected or related devices, one or more of which, pursuant to a program, perform automatic processing of computer data, as well as
- computer data stored, processed, retrieved or transmitted by elements covered under the two bullet points above for the purposes of their operation, use, protection and maintenance.

It is seen from the above, in particular from the second bullet point, that the scope of the directive is wide and therefore applicable to a wide range of data reusers such as eHealth apps providers or Internet of Things systems designers.

Article 14 is one of the core requirements, as it requires Member States to ensure that public administrations and market operators take appropriate technical and organisational measures to manage the risks posed to the security of the networks and information systems (Article 14(1)). In addition to this, the directive imposes a duty to notify to the competent authority about incidents having a significant impact on the security of the core services they provide (Article 14(2)) and grants the possibility for the competent authority to inform the public, or require the public administrations and market operators to do so, when in the public interest (Article 14(4)(5)). The obligation to notify now only exists for the electronic communications sector, which makes the provision in Article 14 a significant improvement.

Furthermore, the proposed NIS Directive contains an additional requirement for market operators to prevent and minimize the impact of security incidents on their core services, and thus ensure their continuity. In other words, not only must the operator providing, for instance, medical devices, take preventive and defensive measures, it must also be able to continue functioning when incidents do occur. This will also apply to data reusers as long as the secondary use of data is their core service.

[223] European Commission, Explanatory memorandum, Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL concerning measures to ensure a high common level of network and information security across the Union, COM(2013) 48 final, 7.2.2012.

The NIS Directive will look at security measures not from the viewpoint of the (processing of) data, but from the viewpoint of the relevant networks and information systems. Hence, the party addressed is the provider of the service, not the data controller. So, for instance, if a pharmaceutical company establishes an e-commerce platform, where it gathers and analyses patient data, it will not only be the data controller, but also the market operator in the sense of article 14(1) of the NIS Directive. As such, it will be responsible for the security of the networks and information systems it controls and uses, and will be obliged to report any significant adverse incidents.

The NIS Directive refers to the security of private networks,[224] whilst the security of public electronic communication networks and services is regulated by a separate body of laws, the so-called Telecoms Package from 2009: Framework directive 2002/21/EC, Universal services directive and e-Privacy directive. The E-Privacy Directive regulates security, privacy and data retention for providers of electronic communications services. The subject of the e-Privacy Directive is the "right to privacy in the electronic communication sector" and free movement of data, communication equipment and services. The E-privacy directive also contains rules on breach notification, which are further explicated in Regulation 611/2013.[225]

The proposed NIS Directive may apply where the e-Privacy Directive may not, especially when these transmissions take place over non-public communications networks. An example is the use of dedicated modems and closed networks, which do not fall under the scope of the e-Privacy Directive. The differences in scope and reach of the new NIS Directive and the GDPR will likely cover more security incidents, but may also have overlaps.

It should also be borne in mind that the DPD already contains the rules on security standards for controllers (including reusers) of personal data. However, there have been no rules relating to those that control (or reuse) non-personal data. For example, a network and information security breach affecting the provision of a service without compromising personal data (e.g., an ICT outage at a power company resulting in a blackout) would not have to be notified.[226]

The NIS Directive provides some measures to boost the overall cybersecurity across the EU, but it will take a while before the directive has been implemented in all Member States. As the law seems to lag behind the rapid technological development, those companies that are eager to exhibit better compliance and trustworthiness, often decide to adopt international standards of information and cyber security. The International Organization for Standardization offers several types of certification that fit the needs of global organizations, which are likely to be under cybersecurity threat. For example, ISO/IEC 27032 addresses "Cybersecurity" or "Cyberspace security", defined as the "preservation of

[224] *Ibidem*.
[225] More broadly, the area of cybersecurity in the EU is also regulated by some International cybersecurity and cybercrime covenants such as the UN Cybercrime convention, the Internet governance forum agreements and human rights conventions. Due to our limited scope we will not explore them in more detail.
[226] *Supra* 219.

confidentiality, integrity and availability of information in the Cyberspace". The ISO standards are by no means binding, however, by giving a guarantee of a certain degree of information security and a widely trustable business environment, they increase commercial attractiveness and raise the level of compliance within an organization.

Table 23 Compliance checklist for the EU cybersecurity law

Compliance checklist for data reusers
• check if you fall under the definition of "a network and information system" as set forth in the NIS Directive • if you do, check whether your security measures are adequate and whether your business processes allow for implementing a notification procedure

10.2 Cybersecurity law from the perspective of the EuDEco model

10.2.1 Assessing technological and socio-economic propositions

Table 24 Technological and socio-economic propositions related to the EU cybersecurity law

Technological propositions	***Legal response – cybersecurity law***
Security and audit functionalities will become crucial topics.	Cybersecurity law answers the need for secure and private systems. The regulatory pressure can bring attention to security and audit functionalities. Also, it helps improving the privacy of the systems.
Socio-economic propositions	***Legal response – cybersecurity law***
Data users need a possibility to assess the relevance of data and they need to be able to determine whether the data were trustworthy (e.g. through certificates, open data documentation).	Trustworthiness is strongly related to security of data. To exhibit solid security and trustworthiness, international data standards and certificates, e.g. ISO, are widely used among the world's leading companies (Amazon, Google, etc.)

10.2.2 CAS

Cybersecurity provisions have a strong relation with the technology. In today's business environment, disruptive technologies such as cloud computing, social computing, and next-generation mobile computing are fundamentally changing how organizations utilize information technology for sharing

information and conducting commerce online.[227] Hacking strategies are becoming more and more sophisticated and require a rougher approach. It is clear that in these changing circumstances, the law has to constantly adapt and develop.

Economic tensions indicate an increasing concern for security. The worldwide cybersecurity market is defined by market sizing estimates that range from $77 billion in 2015 to $170 billion by 2020 according to Forbes.[228]

Also from a more societal and economic perspective, information security is essential for the growth and prosperity. The EU believes a strengthened cybersecurity would result in an increase in trust and willingness for further digitalisation.[229]

[227] Unisys Corporation, "Unisys Descriptive Technology & Trends Points of White Paper Series-Cyber Security" USA, 2011.
[228] http://www.forbes.com/sites/stevemorgan/2015/10/16/the-business-of-cybersecurity-2015-market-size-cyber-crime-employment-and-industry-statistics/ accessed 27 January 2016.
[229] <http://www.euractiv.com/sections/digital/eu-lawmakers-countries-agree-cybersecurity-law-320212 > accessed 27 January 2016.

11 Conclusions

The objective of this deliverable is to provide an overview of the legal requirements for further refinement of the heuristic model of the European data economy. In order to do so, three steps were taken. First, the legal requirements from D1.2 were concretized and transformed into compliance checklists. These checklists were established for data protection law, human rights law, intellectual property law, public sector information law, consumer law and cybersecurity law. The compliance checklists can be found in the respective sections.

Second, in this deliverable (as well as in D2.3 and D2.4) a further integration of the different research perspectives (i.e., law, technology and socio-economics) was made. This was done by examining the technological and socio-economic issues from a legal perspective (and vice versa in D2.3 and D2.4). This exercise shows that many of the technological and socio-economic issues have in fact several legal aspects. For instance, there are legal requirements for technological issues like data quality and information security.

The integration of the different research perspectives included mapping the European data economy, particularly the reuse of data, as a CAS. By checking the key characteristics of CAS (such as aggregation, adaptivity, nonlinearity, feedback loops, non-normal distributions and homogeneous versus heterogeneous expectations), it was shown that the legal framework for data reuse in the European data economy can be regarded as a CAS.

Third, based on these results (compliance checklists, integration of the research perspectives and the CAS approach), some suggestions can be derived for future legislation of the laws and regulations for data reuse. When comparing the practical situation of data reuse in the European data economy with the legal framework governing it, it becomes obvious that practices are rapidly changing, but the legal framework is not. For instance, a new legal framework for data protection law was accepted in the EU in 2016, but the previous data protection law, EU Directive 95/46/EC dates from 1995. Similar examples of legislation lagging behind on practices can be observed in intellectual property law, public sector information law and cybersecurity law.

Furthermore, in practice, the legal requirements governing data reuse are not always observed or often ignored. The legal framework governing data reuse is complex. As a result practitioners may not always be fully informed of the legal requirements that should be observed. In some cases, practitioners may even willingly ignore legal requirements as they may be considered as a hindrance and enforcement is not very strong (although this may change, for instance, with the general data protection regulation).

Altogether, it may be recommended to adapt the legal framework in such a way that it is *not too complex* to understand for practitioners, *flexible* so that new technological developments can be addressed more rapidly and *practical* so that practitioners recognise and understand the legal

requirements when applying them to their business. How to achieve these things more concretely, is subject of D4.2.