Rigorous empirical research on intellectual property

Evaluating the EC Private Data Sharing Principles: Setting a Mantra for Artificial Intelligence Nirvana?

Begoña Gonzalez Otero
IP Researcher and Consultant

December 2018

Rigorous empirical research on intellectual property

This paper was initially drafted during a research stage at the Institute for Information Law (IViR), in Amsterdam. I am grateful to Bernt Hugenholtz, Niko van Ejik, Kristina Irion, Joost Poort, Steff van Gompel, Raquel Xalabarder and Claudia Tapia for all their comments and feedback.

4iP Council is a European research council dedicated to developing high quality academic insight and empirical evidence on topics related to intellectual property and innovation. Our research is multi-industry, cross sector and technology focused. We work with academia, policy makers and regulators to facilitate a deeper understanding of the invention process and of technology investment decision-making.

www.4ipcouncil.com

Evaluating the EC Private Data Sharing Principles: Setting a Mantra for Artificial Intelligence Nirvana?

Begoña Gonzalez Otero
IP Researcher and Consultant
b.otero@euruni.edu

1. Introduction

On April 25, 2018, the European Commission (EC) published a series of communications related to data trading and artificial intelligence. One of them called "*Towards a Common European Data Space*"[1], came with a working document: "*Guidance on Sharing Private Sector Data in the European Data Economy*"[2]. Both the Communication and the guidance introduce two different sets of general principles addressing data sharing contractual best practices for business-to-business (B2B) and for business-to-government (B2G) environments. On the same day, the EC also published a legislative proposal to review the Public Sector (PSI) Directive[3]. These two simultaneous actions are part of a major package of measures aiming to facilitate the creation of a common data space in EU and foster European artificial intelligence technologies' development.

This article focuses on the first action, the "*Guidance on Sharing Private Sector Data in the European Economy*". First, because it is one of its kind. So far, the discussion on data sharing in Europe has been less intense than for data transfer. Maybe because the legal basis for a transfer can be a sale, lease, rental, while data sharing legal basis is more intricate, as we are looking at network structures and co-operation. Second, because, although these principles do not qualify as soft law (lacking binding force but having legal effects) the Commission's communications set action plans for future legislation. Third, because the ultimate goal of these principles is to boost European artificial intelligence (AI) development. However, do these principles set a viable legal framework for data sharing or this public policy tool is merely a naïve expectation? Moreover, would these principles set a successful path toward a thriving European AI advancement? In this contribution, I try to sketch some answers to these and related questions.

It is crucial to mention that EC private data sharing principles evaluation has clear connections to the data ownership debate[4].This paper will neither re-examine this aspect nor the introduction of other possible doctrines[5] nor review any other

[1] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions "Towards a Common European Data Space", COM (2018) 232 final.
[2] Commission Staff Working Document "Guidance on Sharing Private Sector Data in the European Data Economy", SWD (2018) 125 final.
[3] See https://ec.europa.eu/digital-single-market/en/proposal-revision-public-sector-information-psi-directive (accessed on October 15, 2018).
[4] For an overview on the data "ownership" debate see: T. Hoeren, "A New Approach to Data Property?", AMI 2018/2, p. 58-60, available at: https://www.ami-online.nl/art/3618/a-new-approach-to-data-property (accessed on October 15, 2018); B. Hugenholtz, "Data property: Unwelcome guest in the Houes of IP", available at: https://www.ivir.nl/publicaties/download/Data_property_Muenster.pdf (accessed on October 15, 2018); J. Drexl, "Designing Competitive Markets for Industrial Data - Between Propertisation and Access" 8 (2017) JIPITEC p. 257; H. Zech, "A Legal Framework for a Data Economy in the European Digital Single Market: Rights to Use Data", 11 Journal of Intellectual Property Law & Practice (2016), p. 460-470.
[5] For an overview see: M. Dorner, "Big Data und "*Dateneingentum*"", (2014), Computer und Recht, p. 617-628; Osborne Clarke LLP, *Legal Study on Ownership and Access to Data*, Study prepared for the European Commission DG Communications Networks, Content & Technology (2016), available at:

ramifications, such as the right to information privacy and personal data protection[6]. Finally, the assessment of these principles will also stay away from specific consumer law issues related to the use of personal data, including "counter performance" as proposed in the Digital Content Directive[7].

This contribution is structured as follows: The first part will present the problems at stake: what is the current state of AI development in Europe, the availability of data for AI and the Internet of Things (IoT) research and development, and the current legal framework of data trading. The second part will evaluate the principles from an overall perspective focusing on their underlying goals. The evaluation will be addressed separately: first, the principles for business-to-business (B2B) and next, the principles for business-to-government (B2G) data trading will be considered. Last, the paper will conclude by answering the question of whether this public policy tool is merely an unrealistic expectation or whether it sets a favorable regulatory approach for a successful development of AI enabled technologies in the single market.

2. The Problems at Stake:

2.1. The Status Quo of AI Development in Europe.

Investment in artificial intelligence (AI) has rapidly increased in the last five years at international level. According to a study presented early 2018 which used basic research and market capitalization to track where AI is done, China leads the former statistic, with the U.S. behind and long followed by the UK and modestly by Germany, France and Italy[8]. When looking at market capitalization, the first four largest public companies with AI exposure are Apple, closed followed by Alphabet, Microsoft and Amazon[9], all of them headquartered outside Europe but all of them running business in the single market. Then, why is it Europe behind the US and China in capturing the opportunities of artificial intelligence[10]?

First, for AI innovation to happen, R&D is a must. In the sector of AI this translates into "for AI technologies to evolve, machine learning (ML) needs to happen". Machine

---

https://publications.europa.eu/en/publication-detail/-/publication/d0bec895-b603-11e6-9e3c-01aa75ed71a1/language-en (accessed on October 15, 2018).

[6] See N. Purtova, "Do property rights in personal data make sense after the Big Data turn? Individual control and transparency", 10(2) Journal of Law and Economic Regulation November (2017); Tilburg Law School Research Paper No. 2017/21. Available at SSRN: https://ssrn.com/abstract=3070228 (accessed on October 15, 2018).

[7] Proposal for a Directive of the European Parliament and of the Council on Certain Aspects Concerning Contracts for the Supply of Digital Content, COM (2015) 634 final; see A. Metzger, "Data as Counter-Performance – What Rights and Duties do Parties Have?" 8(2017) JIPITEC 2 p. 2; A. Metzger, Z. Efroni, L. Mischau, J. Metzger, "Data-Related Aspects of the Digital Content Directive" 9(2018) JIPITEC 90 p. 1

[8] A. Goldfarb, D. Trefler, "AI and International Trade" (2018) National Bureau of Economic Research, Working Paper 24254, available at: http://www.nber.or/papers/w24254 (accessed on October 15, 2018), p. 2.

[9] Ibid. p. 3.

[10] See J. Manyika, "10 imperatives for Europe in the age of AI and automation", Report McKinsey Globarl Institute, October 2017, available at: https://www.mckinsey.com/featured-insights/europe/ten-imperatives-for-europe-in-the-age-of-ai-and-automation (accessed on October 15, 2018).

learning is a subset of AI that allows computer systems to learn by analyzing huge amounts of data and drawing insights from it rather than following pre-programmed rules[11]. It requires lots of data to create, test, and "train" the algorithms underlying the AI. Examples can be found in several fields, for instance in drug discovery, Sanofi has signed a deal to use UK start-up Exscientia's artificial-intelligence (AI) platform to hunt for metabolic-disease therapies, and Roche subsidiary Genentech is using an AI system from GNS Healthcare in Cambridge, Massachusetts, to help drive the multinational company's search for cancer treatments[12]. Another example from a complete different sector is Alexa, Amazon's powered Echo cylinder. The household artificial intelligence device helper that can turn off the lights, tell jokes, or let us read the news hands-free. It also collects reams of data about its users, which is used to improve Alexa and add to its uses. How does this happen? 99% of the processing of Alexa's takes place in Amazon's Cloud. As the technology is based on voice recognition, the device needs to be always "alert" listening, but not recording. The moment the machine recognizes to the word Alexa or other similar wake word, it activates, starts recording and the snippet is sent to Amazon's cloud, and use for further training of the artificial intelligence device[13]. However, it is important to note that not all AI systems have the same type of data requirements, some are more "data-hungry" than others. Thus, as AI-enabled technologies are becoming more important to the economy, so too are large quality datasets. Large datasets, meaning structured (not raw) data, are critical input for companies that want to create and develop AI systems. Even the best AI algorithms would be useless without an underlying large-scale dataset, because datasets are needed for the initial training and fine-tuning of these algorithms. Therefore, we are talking about collections of separate sets of information that the computer, the algorithm, will treat as a single unit. It includes raw and processed data, information, and so on. To produce large datasets a considerable investment is necessary, and not all firms involved or who want to enter the AI technology market can afford these costs. But a business that lacks access to good datasets faces a substantial barrier to entering a market involving AI technologies.

Second, most data used for research and development of AI technologies come from the Internet of Things (IoT). Although the definition on what IoT is fuzzy[14] , expressions such as "smart cars", "smart phones", "smart homes" are common nowadays. We normally relate such expression to sensors embedded into devices of all kind, connected to the Internet, transferring data over a network. But in fact, all IoT-related

[11] The Royal Society, *Machine Learning: The Power and Promise of Computers that Learn by Example*, (2017), p. 49 available at: https://royalsociety.org/~/media/policy/projects/machine-learning/publications/machine-learning-report.pdf (accessed on October 15, 2018).
[12] See N. Fleming, "How artificial intelligence is changing drug discovery", Nature, 30 May 2018, available at: https://www.nature.com/articles/d41586-018-05267-x (accessed on October 15, 2018).
[13] For further details see: Amazon's website section on machine learning at: https://aws.amazon.com/machine-learning/?nc1=h_ls (accessed on October 15, 2018); S. Levy, "Inside Amazon's Artificial Intelligence Flywheel", Wired, (2018), available at: https://www.wired.com/story/amazon-artificial-intelligence-flywheel/ (accessed on October 15, 2018).
[14] See R. Minerva, A. Biru, D. Rotondi, "Towards a Definition of the Internet of Things (IoT)" IEEE (2015), available at: https://iot.ieee.org/images/files/pdf/IEEE_IoT_Towards_Definition_Internet_of_Things_Revision1_27MAY15.pdf (accessed on October 15, 2018).

devices, no matter how different they may be, do much more than that. IoT related devices always follow five basic steps: they sense (the environment); they transmit (data); they store (data); they analyze (datasets) and then, act on (datasets). For any IoT application to be worth buying (or making), it must demonstrate value in the last step of that chain, the "act on."[15] Artificial Intelligence and the Internet of Things are intrinsically connected and in need of each other to unleash their potential. The true value of any IoT product and byproduct is determined by AI, or more precisely, by the machine learning process. Reason is that machine learning allows the creation of smart actions that make IoT products and byproducts valuable to consumers. The key: finding insights in datasets.

Third, although the volume of data increases fast it is not really available between economic operators. Recent predictions are that by 2020, the number of IoT connections in Europe will reach 6 billion[16]. According to a 2017 research report by the Centre for the Promotion of Import from developing countries (CBI), Europe has an almost 40% share of the global Internet of Things market, projected to reach a value of around €1.2 trillion in 2020[17]. However, the existence of major issues regarding access and transmission of the data generated by IoT devices has been well recognized by the January 2017 European Commission's Communication "Building a European Data Economy". Much of those data are generated, retained and later on analyzed in "silos" by the "owners" of the technology[18]. This makes it very difficult for (European) businesses and organizations to access and use datasets. If companies face high barriers to accessing such datasets, then they may opt not to enter a market that requires large datasets as inputs, leading to less competition. Companies may forgo entry because of this difficulty, and so competition would decline in both new and established markets. Consequently, a lack of shared data access would harm consumers, sometimes via higher prices, sometimes via a reduction in the number of improved features or other innovations.

Altogether, Europe is running behind in the artificial intelligence global race and in need of a strategy that promotes the democratization of data to overcome these challenges. If this current situation would be due to a market failure, a regulatory intervention would be justified. Yet, would the EC's proposed contractual principles suit?

---

[15] "To act on" can mean an infinite number of things, ranging from a profound physical action (e.g. deploying an ambulance to the site of an auto accident) to merely providing basic information to a relevant consumer (e.g. sending a text message to alert a driver that their car needs an oil change). But no matter what the ultimate step of "act on" actually is, it's value is entirely dependent on the penultimate analysis.

[16] EC Final report - Study "Definition of a Research and Innovation Policy Leveraging Cloud Computing and IoT Combination", March 31, 2016, p.10; SMART number 2013/0037; available at: https://publications.europa.eu/en/publication-detail/-/publication/35f3eccd-f7ce-11e5-b1f9-01aa75ed71a1/language-en (accessed on October 15, 2018).

[17] See: "The Internet of Things in Europe" (2017) CBI-Ministry of Foreign Affairs; available at: https://www.cbi.eu/market-information/outsourcing-itobpo/internet-things/ (accessed on October 15, 2018).

[18] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions "Building a European data economy", COM (2017) 09 final.

2.2. Availability of Data for AI and IoT Research and Development

A pre-condition of data sharing and data transfer is data access. As mentioned, access to privately held and controlled data is considered by the EC as key to the development of AI and IoT technologies in Europe, and only accessed data can be re-used.

Datasets access and use are directed by both contractual and technical factors. At the contractual level, there is a range of permissions, policies, legal considerations, personal and organizational preferences, and other factors that impact the data access rights. Rights, in this context, may cover permissions to view, use, reuse, repurpose, or distribute data. Metadata attributes, such as "rights management," can be assigned to data manually or automatically. When applied, rights management indicates data access status and use conditions. These conventions are primarily contractual and inform technical aspects of system design. To understand the complexities of data access, both contractual and technical, it is helpful to first review the status of data access, specifically, what is meant by open and closed data.

The term open data is very specific and covers two different aspects of openness. First, the data is legally open, which in practice generally means that the data is published under an open license and that the conditions for re-use are limited to attribution. Second, the data is technically open, which means that the file is machine readable and non-proprietary where possible. In practice, this means that the data is free to access for everybody, and the file format and its content are not restricted to a particular non-open source software tool[19]. The absence of restriction surrounding open data extends to any endeavor, including commercialization. There are a range of licenses that data producers or data hosts append to data, indicating open access[20].

Following Open Data Institute's definition, closed data refers to data that can only be accessed by its subject, owner or holder[21]. Closed data often contain private or sensitive information. Closed data extend across a wide range of entities, topics, and environment. Examples of closed data include personal, institutional, or industry data identifying financial resources (e.g., sums, transactions, account numbers), personal information relating to health and well-being, or status (e.g., married, single, divorced). Data may also be designated as closed, or regulated by controlled access, due to legal restrictions or organizational policies protecting current or predicted value[22]. More specifically, data access is often restricted because of a known or perceived competitive advantage, and the associated risks with making it public, including

[19] See European Data Portal, General Definition of Open Data, available at: https://www.europeandataportal.eu/en/providing-data/goldbook/open-data-nutshell (accessed on October 15, 2018).
[20] See Creative Commons Licenses at https://creativecommons.org/ (accessed on October 15, 2018)
[21] Definition by the Open Data Institute, available at: https://www.theodi.org (accessed on October 15, 2018).
[22] See T. Aplin, "Trading Data in the Digital Economy: Trade Secrets Perspective" in S. Lohsse, R. Schulze, D. Staudenmayer (eds.), *Trading Data in the Digital Economy: Legal Concepts and Tools,* (2017) Baden Baden, Nomos, p. 59.

misuse, if the data fall into the wrong hands. Closed data are accessible to individuals or organizations who have the appropriate permissions.

Currently, most AI-centered innovation is based on a business model where most training datasets are considered closed data. Such datasets as noted before, are in private silos, not necessary in machine readable and non-proprietary formats. Data storing is already well established as a defensive strategy among AI-centric companies. Google, Microsoft and others have open-sourced lots of software, and even hardware designs, but are less free with the kind data that makes such tools useful[23]. Many startups and SMEs have no bargain power when negotiating a license to get access and use of training datasets, neither can afford the costs.

A second challenge when looking at datasets licensing is that data can be protected by an overlapping patchwork of different intellectual property rights[24] and contractual restrictions on the purposes for which the data can be used. For example, one common misconception is that any freely available online data can be re-used for any purpose. This often isn't the case; website terms and conditions along with copyright and other IP rights, such as the database right, can prevent the data from being used to train a machine learning system. From the practical point of view, many SME's are faced with the problem (and associated costs) of drafting B2B licensing contracts with a necessary degree of legal certainty in respect of the conditions for and scope of the uses allowed by third parties, and Europe lacks any sort of standard contracts or best practices on this regard.

As previously mentioned, access to closed data is considered by the European Commission as key to the data economy and to the development of AI technologies since only accessed data can be re-used. As the Commission acknowledged in their Communication "Building a European data economy"[25] when evaluating the question of "ownership" of data in the industrial context, "*voluntary data sharing might emerge, but negotiating such contracts could entail substantial transaction costs for the weaker parties, when there is an unequal negotiation position or because of the significant costs of hiring legal expertise*".

Finally, if access to data is denied, the question of compulsory licensing becomes relevant[26], as well as competition law intervention. But in the case of access to datasets, as it will be explained in a subsequent section, relying on competition law as the only regulatory tool might not be to the smartest move.

[23] T. Simonite, "AI and Enormous Data Could Make Tech Giants Harder to Topple", Wired, July, 2017, available at: https://www.wired.com/story/ai-and-enormous-data-could-make-tech-giants-harder-to-topple/ (accessed on October 15, 2018).
[24] For a detailed explanation of the current intellectual property rights framework of data in the EU, see B. Hugenholtz, supra n 4.
[25] See supra n 18.
[26] For a detailed study on compulsory license in data trading see: R. H. Weber, "Improvement of Data Economy Through Compulsory Licenses?" in S. Lohsse, supra n 22, p. 137.

Availability of training datasets for AI and IoT R&D is still a hurdle, that, if not reduced, could stifle SMEs innovation, reduce the overall size of the AI market and the benefits that AI could bring to the society.

3. Legal Framework of Data Sharing in Europe

If we look at the data trading (and sharing) relationships within the European single market, three are the existing datasets streams: Public sector information to companies (i.e. government to business or G2B); companies to public bodies (i.e. business to government or B2G); and company to company (i.e. business to business or B2B). Until now, only one these flows has been partly regulated. And this is the G2B.

The public sector is one of the most data-intense sectors within the European Union. Public Sector Information (PSI) is the wide range of information that public-sector bodies collect, produce, reproduce, and disseminate in many areas of activity while accomplishing their institutional tasks. In other words, public sector information means information public bodies produce as part of their public task. That is, as part of their core roles and functions, as defined in legislation or established through custom and practice.

Access and re-use of these data have been regulated via the Public-Sector Information Directive (PSI Directive)[27]. The PSI Directive, provides a common legal framework for a European market for government-held data. The Directive was subject to a review in 2013 and is currently again under review. The aim of the current revision is to strengthen the position of SMEs by dismantling market barriers to reusing public sector information for commercial purposes. This is because re-use of open data by private companies could contribute to the development of artificial intelligence and IoT markets.
According to the impact assessments[28], three are the main barriers:

- data generated by utilities, transport and publicly funded research have tremendous reuse potential, but are not covered by the current rules, even though much of this research is fully or partly funded by public money;
- real-time access to public sector information is rare. This prevents the development of products and services using real-time information, such as meteorological and transport apps, and;
- the re-use of PSI data can be very expensive, depending on the public institution offering them.

We need to wait and see the outcomes of the discussions between the European Parliament and the Council, before any further evaluations.

[27] Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information, OJ L 175, 27.06.2013, p. 1-8 (PSI Directive).
[28] Available at: https://ec.europa.eu/info/law/better-regulation/initiatives/ares-2017-4540429_en (accessed on October 15, 2018).

Sharing of datasets both in B2B or B2G relationships falls under contract law and the principle of freedom of contract.

As contract law is part of Member States' national law, the rules around private and public organizations entering into a contract for data sharing, access, use and re-use are essentially the subject matter of national law.

Same applies to regulations on contract terms, left for the Member States to decide under national law. Besides, B2B contract terms have long been supported by freedom of contract and distinguished from business-to-consumer (B2C) which are heavily regulated. For instance, B2B unfairness control of standard terms and conditions is an unfamiliar concept for the majority of Member States where such regime does not exist and in others, and where it does exist, like in Germany, it has been criticized[29] However, in the last years and in certain sectors, studies and consultations commissioned and launched by the EC have showed important concerns regarding specific types of B2B trading practices. They have also stem from the view that B2B relationships are not to be completely left for the parties to determine but that the weaker party, often a small and medium sized company (SME), should be given certain legal protection in a way that cannot be displaced or agreed otherwise between the parties. Example of this is the Directive (EU) 2015/2366 on payment services (PSD2 Directive)[30], which was implemented at national level in January 2018, and gives Member States discretion to treat small and medium sized enterprises as consumers in applying the conduct of business rules when a payment service is provided to them[31]. The Food Supply Chain Proposal Directive is another example into the same direction[32]. And a third example is the Proposal for a Regulation on Online Platforms[33], published in April 2018, which provides same protections for both SMEs and non-SMEs businesses using the online intermediation services.

In the current normative framework, only competition law provides a very wide one to prevent abuses in both B2B or B2G. In the case of data sharing this would be between a data holder and a party (another firm or a public body) who wants to have access and/or use to the particular data.

[29] See: M. Lehman, J. Ungerer, "Save the Mittelstand: How German Courts Protect Small and Medium-Sized Enterprises from Unfair Terms" (2017) European Review of Private Law, 25(2), pp.313, recommending not to emulate the German B2B control of standard terms model on the European level.
[30] Directive (EU) 2015/2366 of the European Parliament and of the Council of 25 November 2015 on Payment Services in the Internal Market, amending Directives 2002/65/EC, 2009/110/EC and 2013/36/EU and Regulation (EU) No 1093/2010, and repealing Directive 2007/64/EC, OJ L 337, 23.12.2015, p. 35-127 (PSD2 Directive).
[31] Article 38 PSD2 Directive.
[32] Proposal for a Directive of the European Parliament and of the Council on Unfair Trading Practices in Business-To-Business Relationships in the Food Supply Chain Com/2018/0173 Final - 2018/082 (Cod).
[33] Proposal for a Regulation of the European Parliament and of the Council on Promoting Fairness and Transparency for Business Users of Online Intermediation Services COM 2018/0112 Final - 2018/328.

Some scholars have proposed the need of regulatory intervention by crafting default contract rules[34]. This would provide a general legal standard on what a balanced distribution of rights and obligations is in a contractual relationship between the data holder and the other party requesting data access and/or use. Some stakeholders have showed their disconformity with this regulatory approach[35] and consider no legal intervention is necessary.

Additionally, as explained in the previous section, contractual relationships between parties trading in data imply the use of licenses. Model licenses or non-mandatory rules on the use and content of licenses might not be enough to democratize access and use of closed data and boost artificial intelligence in Europe. Particularly in the case of B2G supply of private data under conditions for re-use, one should wonder whether and to what extent mandatory licenses would be necessary or whether public organizations and private companies should be left on their own under the principle of freedom of contract[36].

When looking at this complex scenario, the (non-mandatory) contractual principles published by the European Commission might seem a toddler step, but we should not forget that their Communications are a public policy tool which set action plans for future legislation.

Having said the above, another fact that is worth bringing in this context: On April 23, 2018, two days before the EC's Communication and its guidance on contractual principles were published, a coalition of associations from the EU agri-food chain presented a joint "EU Code of Conduct on Agricultural Data Sharing[37]". This self-regulation instrument promotes the benefits of sharing data and enables agri-business models to swiftly move into a digital data enhanced farming. The eleven pages of the Code shed greater light on contractual relations and provide guidance on access and use of data topics. It is important to recall that both agriculture and automotive sectors have been at the heart of the debate around "data ownership" and "data access", thus the relevance of a sectorial code of conduct which focuses on data access and re-use, rather than in ownership regimes.

This can be also a symptom that self-regulation could be followed by other sectors, such as mobility, health, automotive, energy or aerospace, where industries are rather

[34] F. Graf von Westphalen, "Contracts with Big Data -The End of the Traditional Contract Concept?" in S. Lohsse, supra n 22, p. 249; Twigg-Flesner, "Disruptive Technology -Disrupted Law? How the Digital Revolution Affects (Contract) Law" in De Franceschi (ed.) *European Contract Law and the Digital Single Market*, Intersentia, 2016, p. 21.

[35] See individual responses to EC Consultation Building an European Data Economy by Bayer AG; Industry Coalition on Data Protection (ICDP); Community of European Railway and Infrastructure Companies (CER); Ibec; available at: https://ec.europa.eu/digital-single-market/en/news/public-consultation-building-european-data-economy (accessed on October 15, 2018).

[36] On the need of compulsory licenses in data sharing and transfer see: R. Weber, "Improvement of Data Economy through Compulsory Licenses?" in S. Lohsse, supra n 22, p. 137; M. Grützmacher, "Data Interfaces and Data Formats as Obstacles to the Exchange and Portability of Data: Is there a Need for (Statutory) Compulsory Licensing" in S. Lohsse, supra n 22, p. 189.

[37] Available at: http://www.cema-agri.org/publication/new-brochure-eu-code-conduct-agricultural-data-sharing (accessed on October 15, 2018).

reluctant about the establishment of data access claims[38]; maybe because they are aware that there is no one-way system and that today's plaintiff could be on the other side tomorrow, being forced to provide access to competitors.

All in all, for both, boosting Europe's artificial intelligence technology and harvesting the full benefits of IoT, companies also need to understand the practicability and impact of the principles proposed by the Commission. Thus, looking closer at the principles themselves might shed some light on what kind of legal intervention, if any, the future would bring.

4. Evaluating the Principles on Private Data Sharing

The EC Communication and its accompanying working document[39] present two separate sets of principles, which are meant to guide on contractual relations where data are shared between business organizations or where data are supplied by a business organization to public sector bodies. To evaluate them and answer the question of their practical use the analysis will go as follows: First, a look into the policy reasons motivating them, as described in the introduction of the Communication and the Guidance. Second, as these principles and their underlying goals correspond to different contractual relationships, B2B and B2G, a separate analysis of each set of principles. Within this part, the B2B analysis will concentrate on their underlying objective, namely (to) "*ensure fair markets for IoT objects and for products and services relying on data created by such objects*". This connects with the debate on contract standard terms and the challenges of leaving the prevention of abuses in B2B alone to competition law. The B2G analysis will focus on the principles' primary reason, which is to "*support the supply under preferential conditions for re-use.*" This would lead to the notion of public interest in the use and re-use of private sector (closed) data.

4.1. Policy Behind the Principles

When reading the introduction to these principles, one cannot miss the same and truthful common message in many of the Commission communications related to European Commission's big-data strategy and European data economy: "*data-driven is a key enable of growth and jobs in Europe. The importance of data collected online and generated by the Internet of Things (IoT) objects, and the availability of big data analytics tools and artificial intelligence applications are key technical drivers*."

As some economic studies have shown[40], we should take this statement with a grain of salt due to several reasons.

---

[38] See M. McCarthy, et al., EC Final Report "Access to In-Vehicle Data and Resources" May 2017, p. 55, 194 (Access to In-Vehicle Report) and M. Barbero et al, EC Final Report "Study on emerging issues of data ownership, interoperability, (re-)usability and access to data, and liability" (2016), SMART number 2016/0030, p. 31 and ff. (Emerging Issues Report).

[39] See supra n 1 and n 2.

[40] N. Duch-Brown, B. Martens, F. Mueller-Langer, "The Economics of Ownership, Access and Trade in Digital Data" (2017), JRC Digital Economy Working Paper 2017-01, available at: https://ec.europa.eu/jrc/sites/jrcsh/files/jrc104756.pdf (accessed on October 15, 2018); W. Kerber, J.S.

In the first place, it is indeed true that data can be used multiple times without inherently diminishing its value; thus, fostering the sharing and re-use of data among companies is logical. But for those who harvest data, sharing and making datasets available for re-use in certain formats come with high costs. Therefore, although data as such is a nonrival resource, it might not always be efficient for companies who have invested in data collection to share such datasets as a matter of principle with other companies only for the sake of maximum data exploitation. On this regard, the nonrival nature of data should not be *per se* turned into a maximum efficiency argument pro data sharing alone.

Second, data have no value in themselves, only at their point of use. This is why we should be talking about "datasets" instead of "data". To deliver value, datasets need to be mixed and merged with other datasets[41]. The data holder is not always best placed to extract value from those datasets: this player could lack the skills, the culture or the incentives to deliver innovation. In other words, as Walsh and Pollock said: "the coolest thing with your data(sets) will be thought of by someone else."[42] But even if in some cases the most innovative applications come from unpredictable usage of existing datasets, this should not be considered as the general rule.

Last, the same degree of caution should apply when making statements about how businesses already benefit from access to public sector information available as Open Data. For instance, one study concludes that although the focus of the PSI Directive is to encourage commercial activity in the hope that this leads to new business models and economic growth, a harmonized Digital Single Market of PSI is still far from being reality[43]. Thus, the EU institutions' ambition of the creating of a harmonized public information market across the EU both in terms of the type of underlying works and in terms of compatibility of processes, licensing and formats, is still in the works (and under review).

4.2. The Business-to-Business (B2B) Principles

Frank, "Data Governance Regimes in the Digital Economy: The Example of Connected Cars" (November 3, 2017); available at https://ssrn.com/abstract=3064794 (accessed on October 15, 2018); W. Kerber "Rights on Data: The EU Communication "Building a European Data Economy" from an Economic Perspective" (September 1, 2017). Forthcoming, S. Lohsse, R. Schulze, D. Staudenmayer (eds.), *Trading Data in the Digital Economy: Legal Concepts and Tools,* (2017) Baden Baden, Nomos; available at SSRN: https://ssrn.com/abstract=3033002 (accessed on October 15, 2018).

[41] On the question of whether these datasets could be protected under the sui generis database right, the answer is probably not. As Hugenholtz's explains, it seems that for the European Court of Justice "*investment in 'creating' data does not count towards investment* (criterion for protection)*, even if such epistemological distinction between 'creating' and 'obtaining' data is not self-evident*". For a detailed explanation, see B. Hugenholtz, "Data property: Unwelcome guest in the House of IP" (supra n 4) p. 7-8.

[42] J. Walsh, R. Pollock, "The coolest thing to do with your data will be thought of by someone else", (2007) Open Data and Componentization, XTech2007 available at: http://assets.okfn.org/files/talks/xtech_2007/ (accessed on October 15, 2018).

[43] A. Wiebe, N. Dietrich (eds.) "Open Data Protection: Study on legal barriers to open data sharing – Data Protection and PSI" (2017) Universitätverl. Göttingen, p. 248.

There are five key principles that, if respected, would ensure fair and competitive markets: Transparency; shared value creation; respect for each other's commercial interests; (to) ensure undistorted competition; and, (to) minimized data lock-in. The Communication defines each as follows:

a) ***Transparency**: The relevant contractual agreements should identify in a transparent and understandable manner (i) the persons or entities that will have access to the data that the product or service generates, the type of such data, and which level of detail; and (ii) the purposes for using such data*
b) ***Shared value creation**: The relevant contractual agreements should recognize that, where data is generated as a by-product of using a product or service, several parties have contributed to creating the data.*
c) ***Respect for each other's commercial interests**: The relevant contractual agreements should address the need to protect both the commercial interests and secrets of data holders and data users.*
d) ***Ensure undistorted competition**: The relevant contractual agreements should address the need to ensure undistorted competition when exchanging commercially sensitive data.*
e) ***Minimized data lock-in**: Companies offering a product or service that generates data as a by-product should allow and enable data portability as much as possible[44]. They should also consider, where possible and in line with the characteristics of the market they operate on, offering the same product or service without or with only limited data transfers alongside products or services that include such data transfers.*

A. Principles' Goal: Fostering Data Sharing Environments to Ensure Fair and Competitive IoT Markets

On the B2B data sharing, the underlying goal is to "*ensure fair markets for IoT objects and for products and services relying on data created by such objects.*"

When looking at the results of the Synopsis Report Consultation on "Building a European Data Economy"[45], it is interesting noting that a considerable majority of the stakeholders were against any kind regulatory intervention because in their view, some of the data access issues set out in the Communication may result from the normal dynamic of an emerging market, rather than from a market failure[46].

The question is why the Commission proposes this set of principles under the above-mentioned goal. Even though there is no clear evidence of a market failure,

[44] E.g. data produced by robots in the context of industrial processes, relevant for provision of after-sales services (e.g. repair and maintenance), or data on the rating of service providers.
[45] See Annex to the Synopsis Report: Detailed analysis of the public online consultation results on "Building a European Data Economy", available at: https://ec.europa.eu/digital-single-market/en/news/synopsis-report-public-consultation--building-european-data-economy, p. 12-13 (accessed on October 15, 2018).
[46] See individual responses by Bayer AG; Industry Coalition on Data Protection (ICDP); Community of European Railway and Infrastructure Companies (CER); Ibec; available at: https://ec.europa.eu/digital-single-market/en/news/public-consultation-building-european-data-economy (accessed on October 15, 2018).

as recent economic studies have pointed, it is not less true that we are in an ecosystem with predominant presence of (traditional) data "silos" [47].

For IoT and AI markets to emerge and consolidate in the European Union, we need a data sharing ecosystem. It is to the setting of such ecosystems that the Commission is proposing these five guiding principles. And it needs to be clearly stated that when considering IoT (and artificial intelligence applications as an extension of IoT), we are talking about several markets, thus "*markets for IoT objects and market for products and services relying on data created by such objects."*[48]

To help at understanding this previous statement it crucial to understand what the Internet of Things ecosystem consists of:

First, IoT objects do not "create" data but rather "collect" or "collect and act on" data. These objects are a different set of elements which constitute the first building block of an IoT platform. Those devices are part of the so-called physical layer, the hardware, the "thing". These sensors, actuators and devices collect data from the environment or perform actions in the environment. They need certain computing power, electric power, cooling, memory, sometimes special footprint, multimedia support and connectivity. But they do not work alone; they are part of an ecosystem, the platform. Accordingly, the electronic utility that measures physical properties, the sensor, sends collected data to an aggregator in a cloud that transforms groups of "raw data" into "intermediate data." To get to the cloud, the sensor can be connected through a variety of methods including: cellular, satellite, WIFI, Bluetooth, low-power wide-area networks (LPWAN) or connecting directly to the internet via ethernet. Once the data gets to the cloud, software performs some kind of processing and then might decide to perform and action that goes back to the user.

Second, data management of IoT data is different from traditional data management systems. In traditional systems, data management handle the storage, retrieval, and update of elementary data items, records and files. In the context of IoT, data management systems must summarize data online while providing storage, logging, and auditing facilities for offline analysis[49]. Pattern recognition and data mining techniques can be used for the multitude of IoT applications and produce datasets, that, simply put could be useful for self-improvement of the IoT sensor itself, as well as for the development of new

[47] N. Duch-Brown, supra n 40; W. Kerber, J.S. Frank, "Data Governance Regimes in the Digital Economy: The Example of Connected Cars" (November 3, 2017); available at https://ssrn.com/abstract=3064794 (accessed on October 15, 2018); W. Kerber "Rights on Data: The EU Communication "Building a European Data Economy" from an Economic Perspective" (September 1, 2017). Forthcoming, S. Lohsse, R. Schulze, D. Staudenmayer (eds.), *Trading Data in the Digital Economy: Legal Concepts and Tools,* (2017) Baden Baden, Nomos; available at SSRN: https://ssrn.com/abstract=3033002 (accessed on October 15, 2018).

[48] See supra n 2, p. 3.

[49] M Abu-Elkheir et al., "Data Management for the Internet of Things: Design Primitives and Solution, Sensors" (2013) Nov (11) p. 15582-15612; doi:10.3390/s131115582.

products, byproducts or services that might have no correlation with the initial aim for which data was collected in the first place, as illustrated in the figure below. For instance, data generated by location sensors could be potentially used by publishers to understand and reach a precise local audience or give local context to end-users.

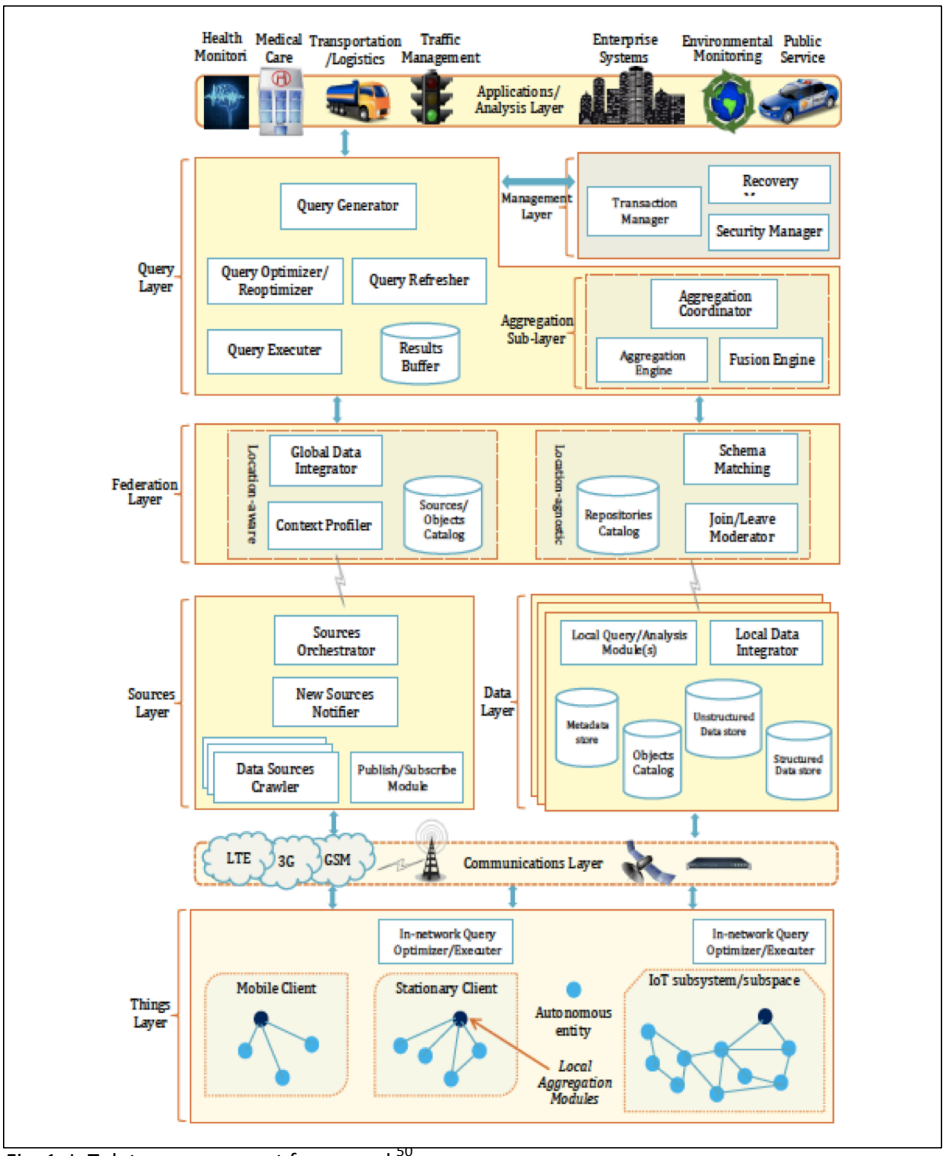

Fig. 1: IoT data management framework[50]

[50] Ibid

Next, we need to understand of what IoT platforms consist: An IoT platform is what makes IoT happen for the devices, that is, an IoT platform is an integrated service that offers the necessary tools to bring physical objects online. Trying to make it as simple as possible, and depending on the tools it provides an IoT platform can be classified as:

- End-to-end or General IoT platform, providing the hardware, software, connectivity, security and device management tools to handle millions of concurrent device connections. A well-known example is Particle;
- Connectivity management platforms, providing low power and low-cost connectivity through WIFI and cellular technologies, as in the case of Sigfox;
- Cloud platforms, mainly enterprise software vendors that are offered by cloud service providers who extend typical enterprise services to include IoT capabilities, such as Google Cloud or Amazon Web Services; and,
- Data platforms, providing data tools that allow routing device data and management and visualization of data analytics, such as Microsoft Azure[51];

Nonetheless, each of the IoT platforms listed above can provide very different byproducts, solutions and uses, completely different from a vertical perspective; from smart systems such as Salesforce that is connected to Microsoft Outlook, an Oracle Database and various sales phone systems. In this case, instead of having multiple places to sort through data, a custom designed dashboard can bring in all of this data into a single pane view. This IoT platform allows correlations discovering and process elimination of inefficiencies. Another type of IoT vertical platform is an industrial IoT, normally used by manufacturers, energy or healthcare, because it integrates Big Data, Machine-to-Machine (M2M) communication, machine learning, smart equipment or robots, and an array of sensors into optimizing processes within a system. Last but not least, if we consider Echo Amazon (popularly known as Alexa), this technology includes particular capabilities that has even made Apple's founder describing Alexa as the next big IoT platform[52]. And we could endlessly continue: There are IoT platforms of every shape and size. There are platforms for specific industries like commercial real estate and family health. Some focus on one type of device: for example, there are platforms focused on augmented-reality headsets. Some are focused on a particular function, like manufacturing[53]. There are even IoT platforms for pets[54].

Also, from a single datasets perspective, a data marketplace is a platform on which datasets can be offered and accessed[55]. Often cited examples are the Microsoft

[51] For a similar breakdown explanation see J Lee, "How to Choose the Right IoT Platform: The Ultimate Checklist" Medium, available at: https://hackernoon.com/how-to-choose-the-right-iot-platform-the-ultimate-checklist-47b5575d4e20 (accessed on October 15, 2018).
[52] See http://www.businessinsider.com/steve-wozniak-thinks-amazon-echo-is-the-next-big-platform-2016-3?international=true&r=US&IR=T (accessed on October 15, 2018).
[53] See Mckinsey Global Institute, "The Internet of Things: Mapping the Value beyond the Hype", June (2015) available at: www.mckinsey.com (accessed on October 15, 2018).
[54] See "Smart Pet Tech and The Intern et of Things" at: https://www.gomindsight.com/blog/smart-pet-tech-and-the-internet-of-things/ (accessed on October 15, 2018).
[55] F. Schomm, F. Stahl, G. Vossen "Marketplaces for data: an initial survey" (2013) ACM SIGMOD Record 42(1), p. 15-26.

Azure Marketplace, Xignite, Gnip, AggData or Cvedia. Data that are being offered may be static archives or online streams of new data. Different modes of access may be offered, for instance, whole repositories, APIs or subscriptions. These are called "data products" as well, where the estimation of the value of such datasets is a continuous challenge[56].

Finally, the latest reports on IoT platforms vendors' alone in the global market reveal that their number reached a new record in 2017, reaching 450, a 25% increase compared to the 360 of the previous year[57]. Most of the increase occurred in the industrial and manufacturing sectors with more than half of the vendors headquartered in the US; The IoT analytics' report also shows that more than 30 vendors included in 2016 have ceased to exist in 2017, they have either gone out of business or been acquired by others. Furthermore, if we search *Crunchbase*[58] for venture-funded IoT platforms, we will find well over 100 hits. This list doesn't include bigger technology players entering the market with IoT platforms like Microsoft, IBM, and SAP or several industrial companies with similar aspirations like GE, Bosch, and Siemens.

In view of this wide-ranging array of horizontal and vertical potential markets for IoT, ranging from hardware, software, connectivity and storage to humans using the information created from data analysis in order to make better decisions. In an ecosystem where IoT platforms are the essential element, collaboration by means of data sharing is more important than ever before. When business share data, it is usually for mutual benefit, determined by commercial negotiation and agreed contract terms. But as the study "Cross-Cutting Business Models for IoT" shows, in the IoT scenario, one step further than traditional cooperation like the application of an open business model, where data sharing is fundamental, will be key[59].

These principles might constitute a good first step towards enabling adequate market conditions for both IoT and AI markets and for the creation of B2B platforms.

B. Introducing Non-Mandatory Contract Terms in B2B

Overall these principles may be seen as too simplistic, but one cannot lose sight that they are framed in a Communication and that its accompanying document makes clear that "model contract terms for different types of data sharing agreements and for some sectors or types of data sharing are already being

[56] A. Muschalle, et al. "Pricing approaches for data markets". In: IEEE 15th International Workshop on Business Intelligence for the Real-Time Enterprise (2012).
[57] See https://iot-analytics.com/iot-platforms-company-list-2017-update/ (accessed on October 15, 2018).
[58] See www.crunchbase.com (accessed on October 15, 2018).
[59] PricewaterhouseCoopers, EC Final report – Study "Cross-Cutting Business Models for IoT" (2017) Study prepared for the European Commission DG Communications Networks, Content & Technology, SMART number 2016/0027.

developed."[60] The measure comes originally from the Telecommunications Sector. In particular, on page 42 of the *Annex to the Commission Implementing Decision on the adoption of the work program for 2018 and on the financing of Connecting Europe Facility (CEF)*[61]*.* We should not forget that the telecommunications sector has already faced very similar problems as to the giving access and re-using of closed data and it may be worth looking at them for useful or inspiring solutions.

The Connecting Europe Facility (CEF) in Telecom[62] is a key EU instrument to facilitate cross-border interaction between public administrations, businesses and citizens, by deploying digital service infrastructures (DSIs) and broadband networks. If recalling what IoT platforms consist of, as explaining above, it makes sense the establishment of a Core Service Platform (central hubs which enable trans-European connectivity) with a Support Centre for data sharing, to support the knowledge exchange between all actors in the data economy. The aim of this Support Centre is also to provide practical advice, best practices and methodologies for both data sharing and data analytics, and it will become operative in early 2019.

If looking at the principles in detail, the transparency one might somewhat resembles Article 5 of the Unfair Terms in Consumer Contracts Directive (UTD)[63]. Yet, it is important to recall that B2B relationships have long been underpinned by freedom of contract and distinguished from B2C relationships which are heavily regulated. For instance, the European Commission's Green Paper which looked into B2B relationships in the sector of food supply chain[64], described freedom of contract as a "cornerstone of any B2B relationship in the market economy"[65]; consequently, parties should be able to design contract that best suit their needs. Nonetheless, this well-established legal principle is increasingly questioned in recent times due to lack of bargaining position of one of the parties to negotiate the terms on which they trade datasets.

Transparency is a precondition for fairness and good faith. In that sense, it might be worth looking at what the European Court of Justice (ECJ) has ruled on Article 3(1) of the UTD and its unfairness test. Because although the Directive applies exclusively to B2C relationships, the ECJ has applied this unfairness test to some B2B transactions. The UTD defines unfairness by resorting to broadly formulated

[60] See p. 6 of EC SWD (2018) 125 final, supra n 2. (Certain increase level of clarity or better placement of this non-regulatory measure would have been welcome, as one needs literally to fish in to find it).
[61] Annex to the Commission Implementing Decision on the adoption of the work program for 2018 on the financing of Connecting Europe Facility (CEF) – Telecommunications Sector, C(2018) 568 final – Annex, February 5, 2018.
[62] See: https://ec.europa.eu/inea/en/connecting-europe-facility (accessed on October 15, 2018).
[63] Council Directive 93/13/EEC of 5 April 1993 on unfair terms in consumer contracts, OJ L 95, 21.04.1993, p. 29-34, Article 5: "*In the case of contracts where all or certain terms offered to the consumer are in writing, these terms must always be drafted in plain, intelligible language. Where there is doubt about the meaning of a term, the interpretation most favorable to the consumer shall prevail. This rule on interpretation shall not apply in the context of the procedures laid down in Article 7 (2).*"
[64] Green Paper on Unfair Trading Practices in the Business-to-Business Food and Non-Food Supply Chain in Europe, COM (2013) 37 final.
[65] Ibid p 6.

standards of good faith and significant imbalance. The ECJ has stated in both *Invitel* and *VB Pénzügyi* that it is up to the national courts to adjudicate whether such "significant imbalance" exists in view of the respective contract term and all other terms, based on the applicable contract rules of the national law of the Member State[66]. Therefore, national rules must construe the benchmark for finding whether a contractual term causes a "significant imbalance" and is "contrary to good faith"[67].

At the European level[68] recent legislative proposals have agreed upon that B2B relationships are not to be completely left for the parties to determine but that the weaker party, often a small and medium sized company, should be given certain legal protection in a way that cannot be displaced or agreed otherwise between the parties. Declarations made by *Elżbieta Bieńkowska*, Commissioner for Internal Market, Industry, Entrepreneurship and SMEs, on April 24, 2018 follow this line of thinking: *"We want to prevent the fragmentation of the Single Market through a patchwork of national rules. Today, the Commission is coming forward with an approach that will give EU businesses – particularly smaller ones – the transparency and redress mechanisms that will help them embrace the digital economy. It also gives platforms legal certainty."* Moreover, as explained in previous sections, in the PSD2 Directive, there is an example where a small or medium sized company is treated as a consumer in a B2B relationship with regards to transparency of conditions and information requirements for payment services[69]. All the above builds up on the studies and consultations related to data ownership and data sharing[70].

In the Guide, the principle of **transparency** is linked to clearly express who has access to the datasets, what type of datasets are given access to and to what level of detail, and also for what purpose(s) is access and/or use license, all key to gain trust among parties. Whether this could also be a matter of **unfairness**, the truth is that to be able to identify who has been given access to datasets is essential to either determine any kind of liability for accuracy or completeness problems, damages arising from further connections or use of the dataset by machines, devices, data user or third parties. But also, for determining liability in case of

[66] Case C-472/10 Nemzeti Fogyasztóvédelmi Hatóság v Invitel Távközlési Zrt ("Invitel"), EU:C:2012:242, para 30; Case C-137/08 VB Pénzügyi Lízing Zrt. v Ferenc Schneider ("VB Pézügyi"), EU:C:2010:659 para 44.

[67] For further details see R. Manko, "Unfair contract terms in EU law" Library of the European Parliament, ref. 130624REV1, 2013, available at: http://www.europarl.europa.eu/RegData/bibliotheque/briefing/2013/130624/LDM_BRI(2013)130624_REV1_EN.pdf (accessed on October 15, 2018).

[68] See PSD2 (supra n 30); Proposal for a Directive of the European Parliament and of the Council on unfair trading practices in business-to-business relationships in the food supply chain, COM (2018) 173; EC Press Release "Online Platforms: Commission sets new standards on transparency and fairness", April 26, 2018 (IP/18/3372).

[69] See PSD2 recital 53 and article 38 (supra n 30)

[70] See Access to In-Vehicle Report and Emerging Issues Report (supra n 38); Annex to the Synopsis Report (supra n 45); N. Duch-Brown et al., "The Economics of Ownership, Access and Trade in Digital Data" (2017), JRC Digital Economy Working Paper 2017-01, available at: https://ec.europa.eu/jrc/sites/jrcsh/files/jrc104756.pdf (accessed on October 15, 2018).

unlawful disclosure of trade secrets. Tentatively, a transparency principle could potentially help at assessing a refusal to license situation as the more information provided in the contract on the datasets, the easier it could be to evaluate datasets substitutivity.

Similar reasons fall under **the shared value creation** principle and **respect for each other's interests**. The assurance of undistorted competition is limited to the exchange of commercially sensitive data. This could suggest a reassurance of the protection of trade secrets, and protecting against tampering in particular. Both were flagged in the Synopsis Report as two core fears for B2B relationships not to share information as well as why business partners in joint projects are sometimes not allowed to receive data[71]. Also, if we look at the relationship between suppliers and an end producer, a contractual principle advocating undistorted competition could fit. Let's consider the Block Exemption Regulation in the Motor Vehicle Sector for the repair and maintenance of motor vehicles and for the supply of spare parts.[72]. The treatment of data on the functioning of the vehicle between the supplier of part and the manufacturer of the vehicle is not regulated within the block exemption. Accordingly, there is the risk that the vehicle's manufacturer could implement contractual terms on data treatment concerning the parts that would place the supplier at a disadvantaged position.

More complicated at first glance is the last principle, namely, (to) minimized data lock-in by enabling data portability. Arguments supporting it are to be framed under two paradigms: on the one hand, the need to train artificial intelligence applications to boost innovation[73]; on the other hand, the need to develop open, technical standards to foster interoperability (enabling data portability)[74] Both combined would ultimately improve Europe's competitiveness in the international dimension.

An example of a data-sharing platform that illustrates the above is the joint venture of the three German car manufacturers, Daimler, BMW and Audi. They acquired Nokia's digital map HERE[75] in 2015 as an important element of their systems for autonomous driving; in 2017, Intel bought 15% of HERE, joining forces and last April 2018, Bosch did, acquiring a 5%. There are other strategic partners such as Pioneer, Esri, DJI, NVIDIA or Oracle. And it is feasible to become a partner. The data produced by HERE are shared and simultaneously used by the partners not only for systems of autonomous driving, but for other mobility sectors such as

[71] See Annex to the Synopsis Report (supra n 45) p. 15-16.
[72] Commission Regulation (EU) No. 461/2010 of 27 May 2010 on the application of Article 101(3) of the Treaty on the Functioning of the European Union to categories of vertical agreements and concerted Parties in the motor vehicle sector OJ L 129, 28.05.2010, p. 52-57.
[73] For arguments supporting that data portability would favor AI see "Data Economy Workshop Report" (2017) p. 4, available at: https://ec.europa.eu/information_society/newsroom/image/document/2017-28/data_economy_ws_report_1A1E8516-DE2A-B8C4-54C4F7CA98621166_45938.pdf (accessed on October 15, 2018).
[74] See Section 6.2., JRC Report (supra n 40) p. 42-46.
[75] See www.here.com (accessed on October 15, 2018).

transportation; logistics, publishers and advertising; improvement of cities infrastructures, secure payment services, just to name a few[76].

Other examples into a similar direction are Automotive Grade Linux (AGL) and Mobilityxlab[77]. The former is a collaborative open source project aiming at bringing together car manufacturers, suppliers and technology companies to build a Linux-based, open software platform for automotive applications that can serve as the *de facto* industry standard. Its underlying idea is that by adopting a shared platform across the industry will reduce fragmentation and allow car manufacturers and suppliers to reuse the same code base, same data-format, leading to innovation and faster time-to-market for new products. The latter, Mobilityxlab, is a coalition of leading Swedish firms that cooperate with startups to develop joint projects for solutions to the transport of the future, primarily to multiply the use of AI in the areas of electrification, connectivity and self-driving vehicles[78].

Yet, discussing about interoperability in the context of data portability or Art. 20 GDPR[79] still raises a number of controversial issues. On the one hand, the lack of obligations for interoperability in Art. 20 could have detrimental effects on users. For instance, the lack of interoperability and compatibility requirements could lead to a race to the "lowest common denominator" of standard datasets provided by data controllers. Adoption of universal requirements to interoperate with all other services would be expensive for companies with uncertain benefits for most users and such a burden would fall disproportionately on start-ups and SMEs, who would have to enter the market with systems in place to interoperate with all other systems already in the market[80]. Eventually, where competing services would need to have common features and functions, it would result in less variety and feature competition, also reducing consumer choice and finally reducing innovation[81]. Additionally, as a Joint Research Center's report indicates, many of the economic results supporting that a welfare-maximizing policy maker would prefer interoperable services in both traditional and platform markets, have been extracted from analyses that do not take data considerations explicitly. Therefore, more economic research is necessary to launch definitive conclusions[82].

All in all, there are quite many incentives for the private sector to follow or at least to not disregard these set of guiding principles. Under these conditions, and as both scholars and industry operators have tabled over the last years in their

[76] Ibid.
[77] See https://www.automotivelinux.org/ and https://www.mobilityxlab.com/en/news/artificial-intelligence-focus (accessed on October 15, 2018).
[78] Ibid.
[79] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L119, 4.5.2016, p. 1-88.
[80] See Robin Wilton's opinion, from Internet Society during the OECD Expert Workshop on Enhanced Access to Data: Reconciling Risks and Benefits of Data Re-Use, May (2018), para 95
[81] Ibid.
[82] See JRC Report (supra n 38), p. 46.

dialogues and consultations with the Commission, it seems the approach taken goes finally towards "(regulating) self-regulation" borrowing Prof. Dr. Hilty's pun[83].

C. Challenges for Competition Law: The Example of a Refusal to Grant Access to Datasets

It is not the intention of this analysis to compare a public policy tool such as the principles contained in the Commission's communication "Towards a Common European Data Space" with a regulatory tool such as competition law. Yet, some reflections are necessary here for two reasons:

First, the results of the public consultation on "Building a European Data Economy" showed that a majority of stakeholders where satisfied with the effectiveness of competition law and its enforcement in addressing potentially anticompetitive behavior of companies holding or using data[84]. Yet, several respondents pointed to the difficulties that the concept of "data sharing" could pose to competition law, as well as that stakeholders believed that competition law should evolve in order to adapt to the digital economy and duly account for the reality of data-driven markets.

Also, some scholars have pointed out that access to data is a disputed topic under general competition law[85]. As this contribution looks at data sharing, the paper circumscribes to the example of refusal to license access to datasets. It is article 102 TFEU, which bans the misuse of a dominant position by one or more undertakings. The CJEU has ruled that this provision may be used for the granting of compulsory licenses (even) to information protected by intellectual property rights[86].

Compulsory licensing for data access is a topic that has also been discussed in reference to sector specific regulations such as the PSI Directive[87], the eCall Regulation[88] and in the field of financial services[89], or in reference to e-platforms[90].

---

[83] See R. Hilty, "Big Data: Ownership and Use in the Digital Age" in Global Perspectives and Challenges for the Intellectual Property System, No 5, June 2018, p. 87-94. In the same line, see also M. Leistner, "Big Data and the EU Databases Directive 96/9/EC" in S. Lohsse, supra n 22, p 38.

[84] See Annex to the Synopsis Report (supra n 45), p. 13.

[85] B Lundqvist "Big Data, Open Data, Privacy Regulations, Intellectual Property and Competition Law in an Internet of Things World – The Issue of Access" (2016) Stockholm Faculty of Law Research Papers, p. 3, available at SSRN: https://ssrn.com/abstract=2891484; J Drexl "Designing Competitive Markets for Industrial Data - Between Propertisation and Access" 8 (2017) JIPITEC 257 para 1.

[86] RTE and ITV v Commission ("Magill"), C-241/91 P and C-242/91 P, ECLI:EU:C:1995:98, [1995] ECR I-743; IMS Health GmbH & Co. OHG v NDC Health GmbH & Co. KG., C-218/01, ECLI:EU:C:2004:257 [2004] ECR I-5039.

[87] See PSI Directive (supra n 27)

[88] Regulation (EU) 2015/758 of the European Parliament and of the Council of 29 April 2015 concerning type-approval requirements for the deployment of the eCall in-vehicle system based on the 112 service and amending Directive 2007/46/EC (E-call) OJ L 123, 19.5.2015, p. 77-89.

[89] See PSD2 Directive (supra n 30).

What all these *ex ante* sectorial regulations and proposals have in common, is that they imply an obligation either to share the data or to grant open access to the data collecting device.

For a unilateral refusal to license access to dataset to be found in violation article 102, the following considerations are to be considered:

For starters, the definition of the relevant market plays a central role in all three areas competition law regulates. To determine abuse of a dominant position, it has first to be determined whether a company has a dominant position in the first place. And to that end, it needs to establish on which market it occupies that dominant position. In 1997, the European Commission published a notice on the definition of relevant markets for the purposes of EU competition law[91]. Accordingly, the market definition is composed of the relevant product market and the relevant geographic market. Ever since, the Commission has continuously "commissioned" reports or launched consultations on market definition in different sectors such as the media (1997), pharmaceutical (2009), telecoms (2002), etc.[92] However, the application of competition law in general, and the definition of the relevant market in particular, are inherently case-specific. For example, while assessing merger control involves a prospective analysis, application of Art. 102 (and 101) TFEU look into past behavior.

Second, when looking at the current practice on refusals to deal and to license as a guide[93], there is one difficult obstacle to overcome when considering datasets. Data is a non-rivalrous resource; if datasets could be substitutable, meaning the same individual data could be found in various datasets, this would count against the requirement of dominance. Thus, a refusal to deal or to license would not prosper.

Finally, if we consider datasets negotiations for analytics involving techniques of data mining by searching datasets for correlations, necessary to improve algorithms of artificial intelligence applications, contractual agreements on access to datasets may simply fail because of asymmetries regarding the value of the

[90] See W Maxwell and T Pénard "Regulating digital platforms in Europe – a White Paper" (2015). available at: www.digitaleurope.org against the French National Digital Council's (CNN) report recommending legislation targeting digital platforms, (accessed on October 15, 2018).
[91] Commission Notice on the definition of relevant market for the purposes of Community competition law (97/C 372/03) (1997) OJ C 372/5.
[92] The media sector is the more prolific, all the studies can be found at http://ec.europa.eu/competition/sectors/media/documents/index.html; in the case of pharmaceutical industries: http://ec.europa.eu/competition/sectors/pharmaceuticals/inquiry/index.html; for telecommunications industries: http://ec.europa.eu/competition/sectors/telecommunications/overview_en.html. For studies on different sectors can be found at http://ec.europa.eu/competition/sectors/ (accessed on October 15, 2018).
[93] For a detailed explanation see Drexl (supra n 4) p. 281-282.

datasets, not because of anticompetitive conduct[94]. And this could be the case with IoT platforms.

Therefore, Article 102 may not be readily applicable to provide access to datasets per se, except when those datasets are indispensable to access an industry, or a relevant market and parties are not able to agree on price[95].

All in all, in such an emerging market sector as the IoT platforms, with so many players and different niches, abuse of dominant position and refusals to grant access to data might be very problematic to articulate.

Thus, relying on competition law as the only regulatory tool, might not be the smartest move. On the other hand, following the results of the consultation launched in 2017, the idea of setting the ground via recommending standard contract terms was generally preferred to the proposal of legislating laying down non-mandatory rules for B2B contracts[96]. So, the idea proposed by the Commission to test *ex-ante* measures in the field of contractual relations may be beneficial towards supporting fair markets for IoT products, byproducts and services.

4.3. Business-to-Government (B2G) Principles

The primary reason to put forward a set of contractual principles regarding the supply of private data to public sector bodies for public interest purposes is to "*support the supply (…) under preferential conditions for re-use.*" This goal could be rephrased as the wish to turn closed data into open data for a public interest reason (AI innovation).

The Commission proposes the six following principles as guidance: Proportionality in the use of private sector data; purpose limitation; "do no harm;" conditions for data re-use; mitigate limitations of private sector data; and, transparency and societal participation.

They read as follows[97]:

a) ***Proportionality in the use of private sector data:*** *Requests for supply of private sector data under preferential conditions for re-use should be justified by clear and demonstrable public interest. The request for private sector data should be adequate and relevant to the intended public interest purpose and be proportionate in terms of details, relevance and data protection. The cost and effort required for the supply and*

[94] This is known as the "information paradox" framed by Arrow in the context of patent law. See Kenneth J Arrow, "Economic welfare and the Allocation of Resources for Invention" in: National Bureau of Economic Research (ed.), The Rate and Direction of Inventive Activity (1962) p. 609.
[95] Huawei Technologies Co. Ltd v ZTE Corp. and ZTE Deutschland GmbH, C-170/13, ECLI:EU:C:2015:477 [2015]. For a commentary on the case see C. Tapia, S. Makris, "Negotiating Licenses For FRAND-accessible Standard Essential Patents In Europe After Huawei v ZTE: Guidance from National Courts" Managing Intellectual Property, May 2018, available at: http://www.managingip.com/Article/3804014/Negotiating-SEP-licences-in-Europe-after-Huawei-v-ZTE-guidance-from-national-courts.html (accessed on October 15, 2018).
[96] See Annex to the Synopsis Report (supra n 45) p. 20-21.
[97] See EC COM (2018) 232 final, p. 13.

re-use of private sector data should be reasonable compared with the expected public benefits.

b) ***Purpose limitation**: The use of private sector data should be clearly limited for one or several purposes to be specified as clearly as possible in the contractual provisions that establish the business-to-government collaboration. These may include a limitation of duration for the use of these data. The private sector company should receive specific assurances that the data obtained will not be used for unrelated administrative or judicial procedures; the strict legal and ethical provisions governing statistical confidentiality in the European Statistical System could serve as a model in this regard.*

c) ***'Do no harm':** Business-to-government data collaboration must ensure that legitimate interests, notably the protection of trade secrets and other commercially sensitive information, are respected. Business-to-government data collaboration should allow companies to continue being able to monetize the insights derived from the data in question with respect to other interested parties.*

d) ***Conditions for data re-use:** business-to-government data collaboration agreements should seek to be mutually beneficial while acknowledging the public interest goal by giving the public sector body preferential treatment over other customers. This should be reflected in particular in the level of compensation agreed, the level of which could be linked to the public interest purpose pursued. Business-to-government data collaboration agreements that involve the same public authorities performing the same functions should be treated in a non-discriminatory way. Business-to-government data collaboration agreements should reduce the need for other types of data collection such as surveys. This should reduce the overall burden on citizens and companies.*

e) ***Mitigate limitations of private sector data:** To address the potential limitations of private sector data, including potential inherent bias, companies supplying the data should offer reasonable and proportionate support to help assess the quality of the data for the stated purposes, including through the possibility to audit or otherwise verify the data wherever appropriate. Companies should not be required to improve the quality of the data in question. Public bodies, in turn, should ensure that data coming from different sources is processed in such a way to avoid possible 'selection bias'.*

f) ***Transparency and societal participation:** business-to-government collaboration should be transparent about the parties to the agreement and their objectives. Public bodies' insights and best practices of business-to-government collaboration should be made publicly available as long as they do not compromise the confidentiality of the data.*

A. Principles' Goal: Incentivizing B2G Data Sharing to Foster AI Innovation

From a business-to-government perspective, the question would be how to find a way that private companies would share and open their private datasets to public bodies to support AI development not only for matters of public interest but for innovation[98]. Add on top of that such openness would need to be in a way that privacy of individuals is respected and guaranteed. And if this would be possible, how to set the conditions for collaborating without harming business legitimate interests, while also mitigating potential limitations of private sector data.

[98] The Commission also adds in their communication the goal of "the economization of public resources". Yet, the only example explaining it is: "this can also lower the burden on companies and citizens by avoiding survey questionnaires." It would be very helpful if this concept is explained in further communications.

Three of the principles proposed by the Commission, namely "**do no harm**", **conditions for data re-use** and **mitigation of limitation of private sector data**, show that there is a clear understanding that pursuing a public good is not a sufficient driver to incentive data sharing for innovation. Businesses are profit driven. They share data typically by selling integrated analytics services, and they can provide different levels of access under different business models. From this perspective, these principles aim at creating incentives for the private sector by either securing monetization, compensation or by lowering costs:

- *"Business-to-government **data collaboration should allow companies to continue** being able to **monetize** the insights derived from the data in question with respect to other interested parties."*
- "*Business-to-government data collaboration agreements should **seek to be mutually beneficial** while acknowledging the public interest goal (…) reflected in **particular** in the **level of compensatio**n agreed".*
- *"Business-to-government data collaboration agreements should **reduce the need for other types of data collection** such as surveys. This should **reduce the overall burden** on citizens and **companies**."*
- *"Companies supplying the data should offer reasonable and proportionate support to help assess the quality of the data for the stated purposes, (but), **should not be required to improve the quality of the data**"*

If these principles would turn into a legislative proposal, it would be critical not to lose sight on how to develop incentives mechanisms. This would comprise an assessment on the legal, economic and technical obstacles preventing B2G data sharing, and advise on concrete actions to promote B2G data sharing for public interest purposes.

Beyond that, there are many questions left open in the air such as whether private data shared with public bodies could become open data, and if so, which and to what extent; or whether it could be re-used for official statistics. The good news is that the Directive on the Re-Use of Public Sector Information is currently under review, and some of its objectives are aligned with these proposed guiding principles. In particular, addressing the risk of excessive first-mover advantage by requiring a more transparent process for the establishment of public-private arrangements by:

a) Allowing any company to learn about the data being available, and,
b) Increasing the chance of a wider range of re-users actually exploiting the data in question[99].

The bad news is that we do not know how the PSI Directive would move forward, nor whether these principles would have any impact at all. In the meantime, besides giving these B2G principles an overall weak evaluation, we would need to see whether the Commission moves relatively quickly on developing this strategy.

[99] COM (2018) 125 final, p. 5 and footnote (19). For details on the current review of PSI2, see Proposal for a Directive of the European Parliament and of the Council of the re-use of public sector information (recast), COM (2018)/234 final – 2018/0111 (COD).

B. Re-Use of Closed Data for Public Interest: A Win-Win Situation?

The famously Walsh and Pollock's quote "the coolest thing with your data will be done by someone else" comes in handy here. Government agencies or researchers make use of private company data to address societal issues. As the Communication points out, statistical offices in some EU Member States use data from mobile telecom operators as an alternative source for official statistics, for instance on mobility or demography[100]. Nonetheless, a private telecom company as Vodafone offers packaged services to public bodies based on the mobility data gathered by their antennas. In developing countries, they offer their data services as an alternative to poor-quality official statistics, and their main incentive lies in corporate image and the potential indirect business benefits[101]. These exact same datasets have proved an invaluable source for controlling outbreaks, surveilling and modeling of infectious diseases[102].

Symmetrically, as explained previously, the re-use of (certain) public sector information by private companies is regulated by the Public-Sector Information, and in force since December 2003[103]. The evolving approach of this Directive is to overcome the resistance among public bodies in Member States to make public data more accessible to the private sector, obviously safeguarding the fundamental right of privacy and personal data protection of individual citizens.

There are other examples in the *acquis* where access to information is promoted by specific legislative means based on the nature of the information. For instance, scientific information is often controlled by academic publishers who tend to seek exclusive licenses so to digitally management of such information (publications). While public institutions tend to promote open-access systems. The Commission Recommendation of 17 July 2012 on access to and preservation of scientific information[104] provides a set of tools as to ensure incentives so that business benefit as well as society and ultimately promote the use of open-access systems.

Yet, when considering public interest, some comments are deemed necessary.

First, the Commission's proportionality principle reiterates that the public interest reason for requesting data should be clearly and demonstrably justified. It shows a clear intention of an enhanced public interest reason, i.e. to give an extra assurance to private companies when handing over their private data. There are examples in the European *acquis* such as the processing of data for archiving, scientific or historical research or statistical purposes, and safeguarded by the

[100] EC Com (2018) 125 final, p. 12.
[101] D2.2 First Report on Policy Conclusions – Update of the European Data Market Study (SMART 2016/0063), p. 31.
[102] See S. Bansal et al., "Big Data for Infectious Disease Surveillance and Modeling", J Infect Dis. (2016) Dec 1; 214 (Suppl. 4), p. 375– 379. Published online 2016 Nov 14. doi: 10.1093/infdis/jiw400,
[103] See PSI (supra n 27).
[104] Commission Recommendation of 17 July 2012 on access to and preservation of scientific information, C(2012) 4890 final.

General Data Protection Regulation[105]. In the field of patent law, for instance the EU Regulation on compulsory licensing of patents for the manufacture of pharmaceutical products for export to countries with public health problems outside the EU, where access to the patent information shall be given to others against a fee[106]. Or in the case of law enforcement and national security[107].

The question in the case of these principles comes with their legal status. If they are a non-binding instrument, how a request to supply private data based on (enhanced or not) public interest can be enforced? It looks good on paper, but there are no instruments that allow this principle to actually operate.

Second, can the fundamental right of privacy be overridden by public interest? And if so, how would this affect a supplying of private data by a company to a public body in the context of these principles?

These questions arise after a ruling by the Court of Justice of the EU in 2017, related to the Universal Services Directive and telephone guides data, *Tele2 (Netherlands) and Others*[108]. European Directory Assistance (EDA) is a Belgian company offering directory enquiry services and directories accessible from Belgian territory. EDA requested the companies which assign telephone numbers to subscribers in the Netherlands (namely, Tele2, Ziggo and Vodafone Libertel) to make available to EDA data relating to their subscribers, relying on an obligation

[105] See Art. 89 of the General Data Protection Regulation (supra n 79).

[106] See Regulation (EC) no 816/2006 of the European Parliament and of the Council of 17 May 2006 on compulsory licensing of patents relating to the manufacture of pharmaceutical products for export to countries with public health problems, 9.6.2006, OJ L 157/1.

[107] A good example is the Mutual Legal Assistance Treaties (MLATs) which are in effect between and among countries around the world and can provide governments the ability to access data in one jurisdiction but needed for lawful investigative purposes in another. For example, Germany signed a Mutual Legal Assistance Treaty in Criminal Matters with the United States in 2003 and a Supplementary Treaty to the Mutual Legal Assistance Treaty in Criminal Matters in 2006. Both treaties entered into force on October 18, 2009 and allow authorities in each country to request and receive information located in the other's jurisdiction (including information stored in third-party facilities clouds). For further information see: W. Maxwell, "A Global Reality: Governmental Access to Data in the Cloud", Hogan Lovells White Paper, May 2012. At international level, the EU-U.S. and Swiss-U.S. Privacy Shield Frameworks. These were designed by the U.S. Department of Commerce, the EC and the Swiss Administration to provide companies on both sides of the Atlantic with a mechanism to comply with data protection requirements when transferring personal data from the European Union and Switzerland to the United States in support of transatlantic commerce. More information at: https://www.privacyshield.gov/welcome (accessed on October 15, 2018). For further information see also: J. V. J. van Hoboken, A. Arnbak, N.A.N.M. van Eijk, N.A.N.M., "Obscured by Clouds or How to Address Governmental Access to Cloud Data from Abroad" (June 9, 2013). Privacy Law Scholars Conference 2013. Available at SSRN: https://ssrn.com/abstract=2276103 (accessed on October 15, 2018); T. Christakis, "Lost in the Cloud? Law Enforcement Cross-Border Access to Data After the "Clarifying Lawful Overseas Use of Data" (Cloud) Act And E-Evidence", FIC Observatory, June 28, 2018, available at: https://observatoire-fic.com/en/lost-in-the-cloud-law-enforcement-cross-border-access-to-data-after-the-clarifying-lawful-overseas-use-of-data-cloud-act-and-e-evidence/ (accessed on October 15, 2018).

[108] Case C-536/15 Tele2 (Netherlands) BV, Ziggo BV and Vodafone Libertel BV v Autoriteit Consument en Markt (ACM), ECLI:EU:C:2017:214 [2017].

provided for under Dutch law, which is itself the transposition of Article 25(2) of the European Universal Service Directive[109].

The Court was asked whether an undertaking is required to make data relating to its subscribers available to a provider of directory enquiry services and directories established in another Member State; and whether it is necessary to leave the subscribers with the choice to give or not their consent depending on the country in which the undertaking requesting that data provides its services. To the first question, the CJEU declared that the Universal Service Directive covers all requests made by an undertaking established in a Member State other than that in which the undertakings which assign telephone numbers to subscribers are established. To the second question, the Court confirmed that the passing of the same data to another undertaking intending to publish a public directory did not required subscriber's "renewed consent".

It is undeniable that data held by private companies can be invaluable for addressing social issues. They are not a low hanging fruit: they require substantial investment and a degree of direct involvement for the supplier of the datasets. Thus, a mandatory data sharing measure without contemplating returns on investment could put in jeopardy the emerging data driven economy as well as the development of artificial intelligence. Each ecosystem is building its own set of business models and organizational arrangements to fit their particular system of incentives, thus for a B2G data sharing relationship to maximize, this should be the way too. And last but not least, as regards to the information contained in private data or better said, private datasets, a distinction between which are in the public interest and which are only of commercial interest is very difficult to make. To overcome this highly challenging task, the principles proposed by the Commission try to set a framework where the supply of private datasets should be mutually beneficial and proportionately compensated to the supplier. The use of words and expressions such as "proportionality", "purpose limitation", "clear and demonstrable public interest", "do no harm", "mitigate limitations of private data", clearly suggest the Commission's goal is to build on trust while creating business incentives to foster this kind of data flow. To take into account the investment in data collection or adaptation that would be necessary before any private dataset could be supplied and used by public bodies (conversion into relevant formats, anonymization of personal data or confidential business information) while allowing companies to keep on monetizing the insights derived from the datasets provided to public bodies with respect to third parties.

In this scenario there is no "silver bullet" to ensure a boost of Europe's technology and the democratization of artificial intelligence technology. It is a matter of

[109] Art. 25: "Operator assistance and directory enquiry services. (2). Member States shall ensure that all undertakings which assign telephone numbers to subscribers meet all reasonable requests to make available, for the purposes of the provision of publicly available directory enquiry services and directories, the relevant information in an agreed format on terms which are fair, objective, cost oriented and non-discriminatory."

setting the right policy mix of raising awareness among the market players and providing information and guidance about options, modalities and building trust to remove fears. In this sense, the set of principles as such, without any further enforcement measures and the articulation of real incentive mechanisms, would amount to a quite naïve proposition.

5. Conclusions

In this digital era of sharing supply chain data, companies on the move need to develop business growth strategies with AI playing a central role to gain insights, knowledge and ultimately innovate and be competitive. Data held by private companies can be invaluable for addressing societal issues, or for generating new products and services. Nevertheless, it is still unclear if all data or only certain datasets or even whether such datasets, since they are not real time data and have been analyzed and processed according to certain interests, are already bias. Therefore, one needs to be very careful and maybe before jumping into sharing data as a matter of principle, further research on what kind of data are we in need of sharing to address the above objectives.

The EU has been struggling for some time over the need for legal protection of data "ownership" in terms of property, even considering the creation of a new intellectual property right. These two sets of principles on private data sharing, despite of their simplicity, put on the table an important question for reflection: *Should Europe move away from discussing about a regulatory approach to data property and access to data, and rather focus on elaborating on the problem of how to foster data sharing and data collaboration to find better solutions?*

Creating economic incentive is necessary to evolve from a "one-company philanthropy" model for data sharing to an open data sharing community including competing firms. It is also critical to clarify the responsibilities and roles by governments and by private sector actors on issues such as data access, data sharing and data quality. New legislation will just take too long to address these questions, while the amount of power data give to companies cannot be left without regulatory intervention, and just in the hands of stakeholders to be sorted out by the market. However, instead of looking towards a vertical approach, the Commission should look horizontally, as Europe do counts with considerable established rules in different fields such as competition law or intellectual property that could be applied or adapted to the new "data driven" reality. At a sectorial level, it would not hurt looking closer to the telecommunications sector, already experienced at establishing formal and "quasi-formal" standards for the industry, in particular the standardization processes, standard setting and developing organizations, the use of FRAND commitments, etc. Same goes to the Open Source movement, a prototype for open innovation, as it allows independent companies to innovate in a collaborative process, where sharing is the key.

Moving toward a data sharing mantra is urgent to encourage not only further quality datasets training contributions, but to boost the development of AI-enabled technologies, and these basic principles are an approach very worth considering, but we need more. Also, the development of instruments within the context of freedom of contract aiming at protecting the weaker party (or a third party) from unfair exploitation, need to be taken into account. Therefore, the approach needs to include more than recommendations and models for how the parties can design their own contractual arrangements. We need a normative approach with strong regulators, in order to protect both parties' freedom of contract. But at least for now, similar to Buddhism, these principles set the right mantra for a potential artificial intelligence nirvana.