The data-driven society from European Union legal perspective

Introduction

We are living in a data-driven society. A lot has been written about the vast quantities of data our society currently generates. In a single day, we generate more than 2.5 Exabyte of data and 90% of existing data was created in the last two years. It is estimated that an average person will generate 1.5GB per day by 2020. Computing & data storage capacities have lately augmented exponentially and digital data collection has become an inseparable part of our lives.

As we go about our daily lives, we leave a continuous trace of data and information that is being collected by numerous private companies and public institutions. Often we are not even aware of it. On the other hand, companies store vast amounts of data related to their businesses and manufacturing processes. In some cases, they are legally obliged to do so, in other cases, they collect them for further analysis and optimization of production and sales. Or simply it is just cheaper to store the data instead of disposing of them. Lately, with the surge of the Internet of Things (IoT) we are facing a new phenomenon of data collection. It is expected that 20.4 billion of devices will be connected to the internet by 2020[1]. Each and single one of these devices collects data of some kind and all of the data is being stored by someone. Only autonomous cars will generate about 0.75GB per second[2]. That is some big data.

Kitchin defined big data as "vast quantities of dynamic, varied digital data that are easily conjoined, shared and distributed across ICT networks, and analysed by a new generation of

[1] http://www.gartner.com/newsroom/id/3598917
[2] http://mashable.com/2016/08/17/intel-autonomous-car-data/#jy_cIttTMqq9

data analytics designed to cope with data abundance as opposed to data scarcity."[3]

There are many questions arising in relation to the unprecedented collection of data. In the first place technical ones: How to effectively manage data and data flows to bring value to it? Big data is just the ability to harness and store huge amounts of data which only becomes information by its interpretation. Whether it is geographical information, statistics, weather data, research data, transport data, energy consumption data, or health data if not properly analysed to gather intelligence and foster innovation and growth, then data becomes just a Big Junk. It is the normalization of datasets and the use the algorithms that can order and analyse raw data in order to extract information and add value to it.

With the rise of new analysis techniques, data has acquired high commercial and social value. As Mattioli[4] explains: "Big data represents a secondary, and largely speculative, public value that resides far downstream from the commercial exchanges that take place between the data producers and their customers." But data do not only have a present value, as Mayer-Schönberger[5] points out: "Data's value shifts from its primary use to its potential future uses."

If data are so valuable, then it is crucial to determine the ownership of data and who should have (lawful) access to it. Is there a need for a new legal concept of data ownership or is the current legislation sufficient?

Gartner, a consultancy, estimated that 30% of the companies in 2016 would treat their information assets as if they were a currency and thus would be dealing with them on a regular basis. It might be just true. There are increasingly more and more legal transactions in which "proprietary rights" on data are being claimed. It allows one party to set strong contractual restrictions on the use of the data being provided under the agreement (usually as

[3] R. Kitchin, *The Data Revolution*, Sage, 2014, p. 15

[4] M. Mattioli has written Disclosing Big Data (February 20, 2014). Available at SSRN: https://ssrn.com/abstract=2358985 or http://dx.doi.org/10.2139/ssrn.2358985

[5] Viktor Mayer-Schönberger, Kenneth Cukier: *Big Data: A Revolution that Will Transform how We Live, Work, and Think,* Houghton Mifflin Harcourt, 2013

a license). The reason for such appropriation is typically the high commercial value (either real or potential) granted to the data. Restrictions are usually based on two different presumptions, that the data is owned by its possessor (in the sense of *in rem* ownership of tangible objects) and that the data is kept confidential and thus constitute a business secret.

Both of the foregoing presumptions are potentially problematic and bring many questions. Should we assume that all big data fall into the scope of copyright? Can data be a secret if collected from individuals or from third-party devices? Is data susceptible of appropriation? And if so, what is the correct legal institution?

There is a range of rules applicable on data: copyright, database protection, competition laws, trade secrets and contractual law. In many occasions it is hard to fit them in any of these rules. Maybe it is time to redesign the rules and find a specific, unified rules for data.

Data can have different features and modalities. It can comprise copyright protected works, personal data, publicly accessible information, open data or proprietary information. Different types of data can be included in separate databases or mixed all together. Data can be gathered from various sources and the same data can be collected by several entities at the same time. These facts make the question of the "ownership" of data very difficult to answer.

To give an example, new cars count with numerous sensors and can include third-party software, such as Apple CardPlay or Android Auto. Information about cars and driving conditions, geo-localization, surrounding area, the driver and much more is being collected or is easy to collect. But who "owns" the data? The car manufacturer, the manufacturer of the sensors, the software developer, the owner or the driver of the car or even the car itself? All of them might feel entitled to be the "owner" of such data, even if not necessarily collected by them.

Data is an intangible asset and as such, it cannot be possessed in the same fashion as tangible assets. Intangible assets allow simultaneous use by persons or entities and they are not-

perishable by nature, contrary to tangible assets. As it was said (Macmillan) "because of the non-rivalrous and non-wasteable nature of things in intellectual space they are all incapable by their nature of being exclusively owned or appropriated."[6] (...)

Notwithstanding the theoretical incapability of being owned, data is starting to be traded as if it was exclusively owned[7]. On the other hand, due to factors such as the different nature of each data set, the incapability of market players to set any pricing rules on data exchange, transaction costs and the non-rivalrous nature of data, most tech firms prefer to keep their data for themselves and when it comes to potential competitors, to buy them up and get hold of their data.

European Union legal framework

With the increasing importance of new technologies and the exponentially increasing amounts of data being collected and generated by connected devices, the EU is pushing forward new rules on the management and protection not only of personal data but also raw machine-generated data in order to establish a clear playing field for all market players. In this sense, Commissioner G. Oettinger, in several of his speeches about the Digital Single Market (DSM) has spoken about the need for a unified "civil code" for data that would address all the issues of ownership, access and interoperability.

So far, several laws which affect data are in the pipeline or have been adopted already. In 2016, the EU adopted the General Data Protection Regulation (GDPR)[8], which will unify

[6] F. Macmillan, *Altering the contours of the public domain, in Intellectual Property – The many faces of the public domain*, ob.cit., p. 100

[7] Beatrice Covassi from DG Connect gave in a presentation in November 2015 two real world examples. The first example refers to social media platforms that offer access to their data (ergo user data) by third parties but the data stays with the platform. As a second example, she looks onto an open API built by a Spanish bank that offers third parties to build value on data collected on the usage of credit cards by its users. In this second scenario, data initially "owned" by individuals were anonymised and aggregated. The "ownership" of such aggregated and anonymised data passed onto the bank, which offered its use to third parties in form of the API.

[8] 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

personal data protection across the EU, when it comes into force in May 2018. In the same year, it also adopted the Trade Secret Directive[9] and started working on the Draft Directive on Copyright[10], which impacts data mainly through the new exception on text and data mining (TDM).

In January 2017, new documents within the free flow of data in the Digital Single Market initiative[11] were published, including a Communication on Building a European data economy[12] and a proposal for a Regulation on Privacy and Electronic Communications[13] which would substitute the e-Privacy Directive (2002/58/EC) and extend its application to the over-the-top service providers.

Personal data protection

On 25 May 2018, the GDPR will substitute the previous Directive[14] which has been in force since 1995. It strengthens rights of data subjects even further, a necessary move in the midst of massive collection of and non-transparent trading with personal data. Just to give an

---

of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) http://ec.europa.eu/justice/data-protection/reform/files/regulation_oj_en.pdf

[9] http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:32016L0943

[10] Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market (2016/0280 (COD)) https://ec.europa.eu/digital-single-market/en/news/proposal-directive-european-parliament-and-council-copyright-digital-single-market

[11] The EU Commission is looking into the issues arising around big data in the context of free flow of data initiative within the DSM, which "(...) will address the emerging issues of ownership, interoperability, usability and access to data in situations such as business-to-business, business to consumer, machine generated and machine-to-machine data. It will encourage access to public data to help drive innovation", taken from Inception impact assessment of European free flow of data initiative within the Digital Single Market; http://ec.europa.eu/smart-regulation/roadmaps/docs/2016_cnect_001_free_flow_data_en.pdf

[12] 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” https://ec.europa.eu/digital-single-market/en/news/communication-building-european-data-economy

[13] Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications) https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-privacy-and-electronic-communications

[14] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:31995L0046

example, there are at least 50 data broker companies only in the EU and highly detailed profiles of our digital selves are being created and sold as cheap as 0,07€[15].

Although the GDPR avoids the concept of "ownership", some of the provisions are directed in that direction, especially regarding the right to be forgotten and data portability. The first one gives the data subject right to demand the erasure of personal data under certain conditions. The second one gives him the right to receive the data in machine readable format and to transfer it to another data controller.

There are concerns about greater risks of leakage of personal data and of a *de facto* ownership of such data if it is traded with under the data portability provision. Nevertheless, it brings novel and more protective approach to personal data, including penalties of up to 4% of annual worldwide turnover, strengthened consent, right to access or confirmation of the privacy by design concept and it should be seen as an important step forward.

Trade secrets

The new Trade Secrets Directive which was adopted in June 2016 seeks to harmonise the level of protection of trade secrets across the EU[16].

The Directive in its art. 2(1) defines trade secrets as any information that meets all three requirements set forth in the same article: (i) it has to be secret, as not generally known or readily accessible to persons within the circles that normally deal with such kind of information, (ii) it has commercial value and (iii) reasonable steps have been taken to keep it secret.

The Directive covers both undisclosed essentially technical information (know-how) and

[15] http://tecnologia.elpais.com/tecnologia/2017/05/03/actualidad/1493835469_309268.html

[16] The need for harmonisation is explained extensively in the recitals of the Directive" "*(N)ot all Member States have adopted national definitions of a trade secret or the unlawful acquisition, use or disclosure of a trade secret, therefore knowledge on the scope of protection is not readily accessible and that scope differs across the Member States*." (recital 6) and "*It is important to establish a homogenous definition of a trade secret without restricting the subject matter to be protected against misappropriation*." (recital 14)

commercial information (trade secrets). On the other hand, it does not cover trivial information (with no or little commercial value) nor experience and skills honestly acquired by employees.

Could then big data and databases be protected by trade secrets? Art. 2(1)(a) of the Directive refers to a secret "as *a body or in the precise configuration and assembly of its components*". It thus seems that if the contents of a database containing large amounts of data are not generally known, have at least a potential commercial value and steps are being taken to protect it, one might readily answer - yes. But looked at it separately, most of the data might be seen as trivial information and as such should not be protected. Potentially, due to the current mosaic legal framework applicable to data, market players may try to protect as trade secrets data and information that shouldn't be protected as such.

Upcoming regulations

As it was said before, service providers and manufacturers in many occasions act as *de facto* owners of data they possess and "data producers" are not able to transfer the data to another provider nor other market players can access such data. The EU Commission is trying to adopt an innovative approach to such situation and open up access to certain types of data. In the future, data possessors could be obliged to offer data under FRAND terms or to develop Application Programming Interfaces (APIs) or even to give free access to data when it is in public interest or for scientific research; possibly in similar terms as the text and data mining exception included in the copyright reform which is currently being discussed in the European Parliament. Data producers might be also granted a right to handle data produced by their devices, for example to authorise third parties to use such data.

Another possible option would be to regulate the portability of non-personal data in similar way as the portability of personal data included in the GDPR. It could include the interoperability of different platforms and creation of new technical standards. End users

would be thus able to switch between service providers without losing their data or even use new cross-platform services.

In order to successfully implement such regulations across the EU, the European data market would have to be opened up, establishing a cross-border free flow of data and abolishing the current restrictions on place of storage of data that many Member States have in place.

Text and data mining (TDM) of copyrighted works

TDM is a tool to analyse large amounts of data using machine processing, in particular to identify correlations, trends and patterns. It is also involved in information visualization, question answering or deep learning and artificial intelligence activities. It is something like reading in a machine way.

So far, most research groups and commercial entities have been mining under licensing agreements with the content holders and publishers[17].

The new Draft Directive on copyright in the Digital Single Market has introduced a TDM exception to copyright[18]. So far, the latest version issued by the Committee on Legal Affairs of the European Parliament[19] would give right to mine to any person with lawful access for the sole purpose of TDM. It also includes an obligation to give access to datasets for the purposes of TDM to research organisations even when they don't have lawful access to the works.

The proposal has been criticized by academics and practitioners from all around the EU[20].

[17] Typically TDM is licensed by the publishers or aggregators such as Copyright Clearance Center either together with or separately from access rights.
[18] Limited TDM exceptions were introduced in June 2014 in the United Kingdom and in France in October 2016. Germany and Switzerland have similar draft bills. Outside of Europe, Japan was the first to introduce a TDM exception already in 2009. The exceptions typically cover only non-commercial uses and require previous lawful access to the data. The Japanese exception does not extend to databases.
[19] Committee on Legal Affairs, DRAFT REPORT on the proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)) http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+COMPARL+PE-601.094+01+DOC+PDF+V0//EN&language=EN
[20] Most notably, the European Copyright Society, an association of renowned scholars from across the EU, issued on 24 January General Opinion on the EU Copyright Reform Package criticizing the limited

Some of the concerns raised are over the oversimplification of TDM, which covers many technical solutions and that copyrighted works are just a small part of all data being used. The exception for some critics means an implicit recognition that all TDM means in principle a copyright infringement.

In that respect, copyright was never intended to give the right to control secondary, value-added uses of the works. It was meant to protect human creativity and intents of fitting TDM into copyright seem as stretching copyright far outside of its original purpose.

What comes next

In our data society, the most valued companies are the ones which possess vast quantities of data and quantitatively they are few: Google, Facebook, Apple, Amazon and Microsoft, interestingly the five most valuable listed firms in the world. They and many others have a commercial interest in owning the data they gather to monetize them to a maximum extent. On the other hand, there is a clear public interest in accessing such data for research, innovation, statistical purposes or market control. These questions do not belong anymore to the realm of intellectual property rights. Antitrust authorities will have to consider new innovative approaches and regulators will have to step in and redefine the rules of the game. Public policy will have to change, favouring more open data streams and possibly declaring some types of data a public utility.

In effect, we are facing a new kind of data dominance which will have to be addressed. On that account, when in 2014 the merger between Facebook and WhatsApp was allowed by the EU Commission, no definition of market for data was given. The competition authorities failed to understand the importance of data and its market power. At least, this is slowly starting to change. In 2016, the Bundeskartellamt (German Federal Cartel Office) initiated investigation

extend of the proposed limitation. Another academic institutions, such as The Max Planck Institute for Innovation and Competition, have positioned themselves against the Commission proposed text of TDM exception, proposing that the exception should cover data mining performed by anyone for both commercial and non-commercial purposes and without the need of previous legal access to the works.

into possible Facebook's abusive practices and the EU Commission alleged that Facebook provided misleading information about data matching between Facebook's and WhatsApp's databases. Most recently, on 27 June 2017, the EU Commission fined Google 2.42 billion euros for unlawful abuse of dominance with its shopping search tool.

All in all, this new oil of our society is still an unknown territory not only for the legislation but for the society as a whole. We will continue seeing rapid changes and new applications for data and legislators will have to learn how to keep up. Maybe, soon enough we will see artificial intelligence helping to craft new laws for data with the help of algorithms and data.