ADI 42 (2022): 107-130

IPRS AND BIG DATA: A PROPOSAL FOR A FAIR BALANCE BETWEEN BUSINESSES' LEGITIMATE INTERESTS AND DATA SHARING IN THE LIGHT OF THE EU DATA ACT

DERECHOS DE PROPIEDAD INTELECTUAL Y *BIG DATA:* UNA PROPUESTA PARA UN JUSTO EQUILIBRIO ENTRE LOS INTERESES LEGÍTIMOS DE LAS EMPRESAS Y EL INTERCAMBIO DE DATOS A LA LUZ DE LA LEY DE DATOS DE LA UE

FRANCESCA GIORDANELLI*

ABSTRACT

The European Commission launched in 2020 the European Strategy for Data that is aimed at boosting the sharing of data among businesses, but also among businesses and governments. This paper analyses how the Database Directive and the Trade Secrets Directive in their current state apply to Big Data and proposes a modification of the current legal framework to align with the objective of the data strategy also taking into account the Draft Proposal of the European Commission for the Data Act. An analysis of intellectual property justifications demonstrated that an exclusive property right is not justified and the that the database sui generis right should be abolished or at least not applied neither to raw data nor to derived data. However, it is necessary for businesses to maintain a de facto control over the data they produce to recoup the costs and profitably share data. This paper argues that this control can be maintained through secrecy and that is why it is submitted that the Trade Secrets Directive provides the right framework of protection, even if some aspects need clarification.

Keywords: big data, database *sui generis* right, trade secrets, Data Act, European Data Strategy.

RESUMEN

La Comisión Europea presentó en 2020 la Estrategia Europea de Datos que tiene como objetivo impulsar el intercambio de datos entre las empresas, pero también entre las empresas y los gobiernos. Este documento analiza cómo la Directiva de Bases de Datos y la Directiva de Secretos Comerciales en su estado actual se aplican a los *big data* y propone una modificación

* Doctor in Law, graduated in law at Alma Mater Studiorum University of Bologna (Italian Law Master's Degree) and King's College London (LLM in Intellectual Property and Information Law). Dirección de correo electrónico: *francesc.giordanell3@unibo.it*.

Fecha de recepción: 30 de marzo de 2022 // Fecha de aceptación: 10 de mayo de 2022.

del marco legal actual para alinearse con el objetivo de la estrategia de datos teniendo también en cuenta el Proyecto de Propuesta de la Comisión Europea para la Ley de Datos. Un análisis de las justificaciones de la propiedad intelectual demostró que un derecho de propiedad exclusivo no está justificado y que el derecho sui generis de las bases de datos debería ser abolido o al menos no aplicado ni a los datos brutos ni a los datos derivados. Sin embargo, es necesario que las empresas mantengan un control de facto sobre los datos que producen para recuperar los costes y compartir los datos de forma rentable. En este documento se argumenta que este control puede mantenerse mediante el secreto y por ello se afirma que la Directiva de Secretos Comerciales proporciona el marco de protección adecuado, aunque haya que aclarar algunos aspectos.

Palabras clave: big data, derecho *sui generis* de las bases de datos, secretos comerciales, Ley de Datos, Estrategia Europea de Datos.

«Information wants to be free [...] but information also wants to be expensive because it can be immensely valuable».
STEWARD BRAND, *The Media Lab: Inventing the future at MIT*.

I. INTRODUCTION

In February 2020 the European Commission announced the «European Strategy for Data» for the next coming years. The goal of the Commission is to increase «the use of, and demand for, data and data-enabled products and

services throughout the Single Market»[1], as it is submitted that the full potential of data sharing has not been unleashed. The ultimate aim would be to build a single internal market like the one for goods and services, but for data where all types of data (both personal and non-personal) are shared in a secure way in compliance with the EU *acquis*[2]. On February 23, 2022 the Commission issued a proposal for the Data Act[3] that deals with actual rights on the access and use of data and aims at clarifying the IPR framework with a view to further enhance data access and use.

Also, on the 25 of November 2020 the Commission published an Action Plan on Intellectual Property[4]. One of the objectives of the Action Plan is to make sure that the European Intellectual Property framework is fit for the difficult task of balancing between the need to foster data sharing and the need to be able to safeguard legitimate interests (especially those of businesses that need IP protection of their creations and innovation). The Commission aims at devoting its attention to the clarification of the scope of protection granted by the Trade Secrets Directive[5] and the modification of the Database Directive[6].

This paper will focus on whether the database *sui generis* right and the protection of trade secrets applied to Big Data and Big Data analysis outputs are compatible with the overall policy objective of reaching abundance of high-quality data and incentivizing the sharing of data among private actors (B2B sharing). This paper will then provide a proposal for the modification of the Database Directive and the clarification of the Trade Secrets Directive in line with such policy objective.

This study will only deal with the sharing of data among private businesses and will not tackle the EU legislation and policy objectives that relate to B2G sharing of data, as the latter involves public interest considerations as well as issues of accountability, transparency and compliance with ethical principles[7] that go beyond this analysis.

It will be suggested that the database *sui generis* right should be abolished or at least not applied neither to raw data nor to derived data and that instead the Trade Secrets Directive is useful in the data economy because it is aimed at banning some unfair behaviours, but it does not create an exclusive proprietary right on data (therefore granting a degree of protection of businesses' interests without locking information in the hands of just one owner and favouring the exchange of information).

---

[1] European Commission, *Communication on a European Strategy for Data,* 19 February 2020, COM(2020) 66 final, pág. 1.

[2] *Ibid*., pág. 4.

[3] European Commission, *Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)*, 23 February 2022, COM(2022) 68 final.

[4] European Commission, *Communication on Making the most of the EU's innovative potential: An intellectual property action plan to support the EU's recovery and resilience,* 25 November 2020, COM(2020) 760 final.

[5] Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure [2016] OJ L 157/1.

[6] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 77/20.

[7] European Commission, *Towards a European strategy on business-to-government data sharing for the public interest Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing*, 2020, pág. 8.

1. Big Data and the European Data Strategy

Data have been said to be similar to «oil»[8], because they led to the creation of a new «lucrative, fast-growing industry» that resembles the oil industry at the beginning of the 20th century, but also to «sunlight», because they are «everywhere like solar rays»[9] and to diamonds, because of the many facets that they have[10]. However, all the previous analogies are not completely right. Data are very valuable, but they are not like oil because they are non-rival, by nature non-excludable and durable[11], they are not like sunlight because they can be excludable (*e.g.* through encryption), and they are not like diamonds, because their worth is not intrinsic, but they need to be analysed to produce value. As a matter of fact, the data economy and the «data revolution»[12], that changed not only the way business is made but also how knowledge is transferred and produced[13], is anything but something that can be described in simple terms.

Not even coming up with a definition for «data» is simple. It is possible to go back to the etymology of the word: it derives from the verb «dare» in Latin (that means «to give»)[14]. Kitchin proposes a definition based on the etymology according to which data are «elements that can be abstracted from (given by) phenomena measured and recorded in various ways»[15]. Moreover, the relationship between «data» and «information» has been dealt with by the literature: Kitchin argues that data are «pre-analytical and pre-factual» because they need interpretation and context to become «information»[16]. Zech instead distinguished between the semantic level of data (that relates to the meaning of the information conveyed through the data) and the syntactic level (that relates to the «sequence of zero and ones», so the signs and characters through which data are expressed)[17]. However, it was suggested by part of the literature that this distinction is «rather artificial» given that ultimately data always carry information[18].

As for the concept of Big Data, it has been developed for the first time in the mid-1990s to refer to the practice of handling and analysis of massive datasets[19].

Both the concepts of Big Data and Big Data analysis are relevant for the purposes of this study; the former refers to massive data sets characterised by the 5 V[20]: volume, veracity, velocity, variety and value. The first «V», volume, relates to the fact that big data are large amounts of data, and this has been possible

---

[8] «The world's most valuable resource is no longer oil, but data» (*The Economist*, 6 May 2017) *https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data,* accessed 21 March 2022.

[9] «Are data more like oil or sunlight?» (*The Economist*, 20 February 2020) *https://www.economist.com/special-report/2020/02/20/are-data-more-like-oil-or-sunlight,* accessed 22 March 2022.

[10] Schovsbo and Kokoulina (2020), pág. 2.

[11] Floridi (2022), pág. 67.

[12] Kitchin (2014), pág. 1.

[13] *Ibid.*

[14] *Ibid.*, pág. 2.

[15] *Ibid.*, pág. 2.

[16] *Ibid.*, pág. 3.

[17] Zech (2016), pág. 462.

[18] Surblytė-Namavičien (2020) pág. 60; Aplin (2017), pág. 68.

[19] Kitchin (2014), pág. 67.

[20] Gervais (2019), pág. 4; Hurley (2019), págs. 7-9. Other authors restrict the number of «V» to 3: volume, variety and velocity see: Strowel (2020), págs. 107-108; De Mauro, Greco and Grimaldi (2016), pág. 130.

thanks to reduction in cost of storage. The second «V», veracity, relates to the accuracy and reliability of data. The third, velocity, relates to the speed of movement and processing of data (indeed Big Data are not static datasets, but they are «real time heterogenic data» that are updated constantly[21]). The fourth, variety, relates to the different types of data and sources from which data are collected. The fifth, value, as Gervais[22] points out, is a consequence of the other 4 features: because if all the other features are present, Big Data are extremely valuable.

Big Data analysis can be defined as the process of research into huge amounts of data to reveal hidden patterns and secret correlations.

As both Kitchin[23] and Hurley[24] correctly point out, data are not useful in themselves, but their value comes from the analysis through computers that are able to find hidden patterns, associations and trends that may exist in the data. As it has been pointed out by the literature, the way in which Big Data are generated and used can be separated into two phases.[25] First, data are collected from different sources in order to create Big Data. For example, data can be collected by sensors installed in objects (*e.g.* by a smart car)[26] or they could be user-generated data that are collected from a website (*e.g.* Facebook). Secondly, Big Data are analysed through an Artificial Intelligence (AI) algorithm (in particular with a process called Text and Data Mining) to discover hidden patterns, correlations, useful insights and predictions. The AI gets better the more it goes on analysing data. It is in this second phase that it is possible to see the double function of data: on the one hand, they are used to train the AI to get better to find correlations and insights, on the other hand they are used to find those correlations and produce *new* data.

Given the important functions of Big Data it is only reasonable that the European Commission with the EU Data Strategy wants to increase the access and availability of data to allow «Big Data» pattern detection or machine learning. This intervention and in general, the relevance of data regulation through law is of vital importance: data are at the core of AI technology, blockchain, the cloud and through the regulation of data, EU law also indirectly regulates those technologies and any other data-dependent technology[27].

2. EU Data Strategy: some interests of stakeholders to take into account

The EU legislators find themselves in a difficult situation because of the different and opposed interests that are at stake and of the difficult task of finding a regulatory outcome that proportionally balances them[28].

The first interest concerned is the public interest in innovation. The limited access to Big Data causes issues both in relation to the growth of tech businesses

[21] Nordberg (2020), pág. 193.
[22] Gervais (2019), pág. 4.
[23] Kitchin (2014), pág. 100.
[24] Hurley (2019) pág. 4.
[25] Gervais (2019) pág. 4.
[26] *Ibid*., pág. 4: «cars as personal vehicles are one of the main sources of (personal) data - up to 25 Gigabytes per hour of driving».
[27] Streinz (2021), pág. 903.
[28] European Commission, *Factual Summary Report of the Public consultation on the IP action plan roadmap*, 16 November 2020.

and to research: as stated by Danah Boyd and Kate Crawford[29], it causes inequality in the system because those who can pay the companies that retain Big Data for data access or those inside the companies can conduct researches with these data and come to results that those that do not have access can neither reproduce nor critically evaluate. As it emerged from the Open Public Consultation on The European Strategy for Data[30], almost 80% of the respondents to the consultation have encountered difficulties in using data from other companies connected to interoperability but also to denied access to data and very high prices.

The second interest that is concerned is the one of businesses. The data collected by businesses through their activity is extremely precious. Aside from those who made very profitable business models from their collection of data[31], there are also the companies that manufacture data collecting objects (Internet of Things) that have in their availability huge amounts of data. While these companies have an interest in protecting the data they collect, small businesses have an interest to access this enormous amount of data for the purpose of research and innovation of their products and services. In the European Data Space as envisioned by the Commission those who contribute data would get enhanced access to data of others, analytical results from the data pool, services such as predictive maintenance services, or licence fees[32]. As Bruzzone and Debackere[33] point out openness and exclusivity are two sides of a coin that the European legislators must balance very carefully. In their article they highlight that the European Strategy for Data must be «as open as possible, as closed as needed», meaning that the impact of the EU action on incentives for the production of data by economic actors must not be overlooked because the imposition of an overly extensive obligation to share data may lead to discourage the production of data and therefore ultimately hinder the objective of maximizing the availability of data in the EU Single Market.

Lastly, there are the interests of individuals that on the one hand have a right to privacy and data protection when it comes to personal data and on the other, they have an interest in having efficient governmental and commercial services that can be enhanced through the use of Big Data. As it is correctly pointed out by Streinz[34], in the Data Strategy data are presented as a resource that should be exploited for European citizens' benefit. This recognition of the value (economic value) of data can be in tension with the fundamental rights of privacy and data protection law recognized under the Charter of Fundamental Rights.

The objectives pursued by the European Data Strategy face a number of challenges: alongside those posed by the necessity to respect intellectual prop-

---

[29] Boyd and Crawford (2012), pág. 675.

[30] European Commission, *Summary Report on the open public consultation on the European strategy for data*, 2020.

[31] See: Hartmann, Zaki, Feldmann and Neely (2016) págs. 1382-1383: «Some studies estimate an increase in annually created, replicated and consumed data from around 1,200 exabytes in 2010 to 40,000 in 2020 (Gantz and Reinsel, 2012). In some industries big data has led to the creation of entirely new business models. [...] Although companies relying on data - such as insurance companies - is not a new concept, it was only recently that companies began to make use of other data sources such as social media, smartphones or sensors, and new technologies designed to exploit this data».

[32] European Commission, *Communication on a European Strategy for Data,* 19 February 2020, COM(2020) 66 final, pág. 5.

[33] Bruzzone and Debackere (2021), pág. 41.

[34] Streinz (2021), pág. 903.

erty rights and privacy and data protection rights, the Commission itself highlighted that there are some issues holding the EU back from realising its potential in the data economy[35]. Among these issues, there are those connected to the imbalances in market power and anticompetitive behaviours, data governance and cybersecurity. This paper, however, will deal only with the role of intellectual property in the pursuing of the policy objective of enhancing data access and sharing.

II. THE STATE OF THE ART: OF INTELLECTUAL PROPERTY LAW FOR BIG DATA

1. The application of the database *sui generis* right to Big Data

The question of whether or not article 7 of the Database Directive[36] that establishes the *sui generis* right applies to Big Data is debated. There are two conditions to satisfy in order to obtain the *sui generis* protection: the subject matter must be a database as defined in article 1 of the Directive and there must be a qualitatively and/or quantitatively substantial investment in either the obtaining, verification or presentation of the contents of the database.

1.1. *Are Big Data databases?*

The first issue is whether or not Big Data fit into the definition of database provided by the Directive[37].

The most controversial aspect relates to the application of the definition of databases to «raw data» detected through machines and flows of indistinctly detected data (data that are detected every day through various sources - from social media platforms for example- before they are analysed). While there is a part of the literature[38] that argues the definition of database is a very wide definition that will be fulfilled by the overwhelming majority of Big Data collections of whichever kind, this interpretation does not convince most interpreters that highlight the difficulty of considering some kinds of Big Data «arranged in a systematic or methodical way». Bernt Hugenholtz[39] correctly upholds that the definition requiring the individual elements of the database to be «arranged in a systematic or methodical way» excludes protection of collections of raw machine-generated data. So, for example, data captured through the sensors of a smart car would not be encompassed within the definition of database. This condition excludes,

[35] European Commission, *Communication on a European Strategy for Data,* 19 February 2020, COM(2020) 66 final, pág. 6.

[36] Article 7 of the Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 77/20: «Member States shall provide for a right for the maker of a database which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database. [...]».

[37] Article 1 of the Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 77/20: «For the purposes of this Directive, "database" shall mean a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means. Protection under this Directive shall not apply to computer programs used in the making or operation of databases accessible by electronic means».

[38] Leistner (2017), pág. 27.

[39] Hugenholtz (2017), pág. 88.

in particular, that mere data flows indistinctly detected, for example by a natural phenomenon, can be considered protectable[40]. Recital 21 of the Directive, however, does not require data to be physically stored in an organised manner to satisfy this condition but rather it is only necessary that the arrangement is methodical and systematic for the purposes of machine-to-machine processes.

Whether Big Data would fall in the definition of database is already a first element of uncertainty. The definitional problem arises in relation to raw data, while machine produced outputs based on analyses of Big Data are most certainly comprehended in the definition of databases.

1.2. *Is there a substantial investment in the obtaining, verification or presentation of Big Data?*

The answer to this question is divided in two parts: first there needs to be an assessment on whether in Big Data the investment is in obtaining, verification or presentation of the database; secondly there needs to be an analysis of the criteria to establish whether the investment is substantial and whether those criteria would be satisfied in most Big Data situations.

1.2.1. Investment in the obtaining of data or creation of data?

The European Court of Justice («ECJ») in *British Horseracing*[41] and *Fixtures Marketing*[42] maintained that the only kind of investment relevant for the database protection is the one aimed at seeking out existing independent materials (that's what it is meant by the Directive with the verb «to obtain») and that, importantly, the investment aimed at the creation of the content of the database is irrelevant. Indeed, the Court affirmed that the purpose of the protection by the *sui generis* right is «to promote the establishment of storage and processing systems for existing information and *not the creation of materials capable of being collected subsequently in a database*»[43] (emphasis added). From this case law part of the literature[44] rightly noticed that both investments aimed at the creation of smart products equipped with sensors capable of collecting data and investments in Big Data analysis that produce new data (insights, correlations etc.) are not relevant for the substantiality threshold.

The EU Commission in its 2018 evaluation of the Database Directive[45] embraced an even more restrictive approach called «spin-off theory»[46] according to which databases that are by-products of the main activity of the maker should not enjoy protection under article 7 of the Directive. The Commission consequently affirmed that «the *sui generis* right does not apply broadly to the data

[40] Ottolia (2017), pág. 75.

[41] Case C-203/02, *The British Horseracing Board Ltd and Others vs. William Hill Organization Ltd.* (2004), ECLI:EU:C:2004:695.

[42] Case C-444/02, *Fixtures Marketing Ltd vs. Organismos prognostikon agonon podosfairou AE (OPAP)* (2004), ECLI:EU:C:2004:697.

[43] Case C-203/02, *The British Horseracing Board Ltd and Others vs. William Hill Organization Ltd.* (2004), ECLI:EU:C:2004:695, para 31.

[44] Drexl (2017), pág. 268; Gervais (2019), pág. 10.

[45] European Commission, *Evaluation of Directive 96/9/EC on the legal protection of databases*, 25 April 2018, SWD(2018) 146.

[46] See Corrales Compagnucci (2020), pág. 33.

economy, but it only covers databases that contain data obtained from external sources (for example industries like publishers, who seek out data in order to commercialise databases)»[47].

While the Commission's said approach is too strict and does not reflect the ECJ case law, it is still more acceptable that what is sustained by part of the literature according to which the Directive applies to the creation of new data. For example, Leistner[48] taking into account a case of the ECJ[49] that considered investments into establishing an infrastructure to obtain pre-existing data on sales or geographical data to assess the substantiality under article 7 (1), affirmed that the radical conclusion according to machine-generated data will typically not be covered by the *sui generis* right is not correct. Burdese[50] put forward a controversial interpretation and he said that databases generated by data-recording or data-mining processes involve obtaining rather than creating and that the creation/obtaining dichotomy should be abandoned. This interpretation, while going against the literal interpretation of the Directive, also does not take into consideration that at the time the Directive was adopted, most of the situations of creation of data in the Big Data context were not even conceived by the European legislator and it would be a stretch to claim that the database right is actually designed to encompass Big Data.

In the proposal for the Data Act published by the Commission on February 23, 2022[51], article 35 is aimed at clarifying that the database *sui generis* right does not apply to the the raw data, i.e. the databases containing machine-generated data obtained from or generated by the use of products or related services, such as sensors. The Commission recognised that, despite the ECJ case law, there was still legal uncertainty concerning the application of the database *sui generis* right to machine-generated data (*e.g.* the Commission recalls the German *Autobahnmaut*-case, where the German Supreme Court upheld the interpretation according to which machine-generated data would be included in the *sui generis* right[52]) and affirmed that a provision like article 35 was needed given the «need to balance the policy objectives of IP protection of such databases in the context of the data economy, where the exclusivity of data as a non-rival good is in general considered an impediment to innovation»[53].

This legislative decision is therefore the expression of precise policy objective: not to make data subject to an exclusive proprietary right. As much as this should be welcomed, an even more radical solution will be put forward in this paper. Indeed, the database *sui generis* right should not apply neither to the raw data (as the Commission provided in article 35 of the Proposal), nor to the data inferred by Big Data analysis (inferred data).

---

[47] European Commission, *Executive Summary of the Evaluation of Directive 96/9/EC on the legal protection of databases*', SWD(2018) 146, pág. 2.

[48] Leistner (2017), pág. 29.

[49] Case C-490/14, *Freistaat Bayern vs. Verlag Esterbauer GmbH* (2015), ECLI:EU:C:2015:735.

[50] Burdese (2020), pág. 14.

[51] European Commission, *Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)*, 23 February 2022, COM(2022) 68 final.

[52] European Commission, *Commission Staff Working Document Impact Assessment Report Accompanying the Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act),* 23 February 2022, SWD(2022) 34 final, pág. 136.

[53] European Commission, *Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)*, 23 February 2022, COM(2022) 68 final, pág. 10.

1.2.2. Substantiality of the investment

The requirement of substantiality is usually interpreted extensively, as a *de minimis* exclusion[54]. Leistner points out that some scholars have pushed for an interpretation of the «substantial investment» oriented towards a market failure approach which would focus on whether the investment would not have been made in a specific case if there were no *sui generis* protection. However, he rightly says that this market-failure approach would be unpractical in real life and in courts. It could also be added that the balancing of the interests has already been made by the legislator and therefore it would be inappropriate to add an additional requirement: the law just requires a quantitatively or qualitatively substantial investment that respectively refers to quantifiable resources (i.e. time and money), and to intellectual effort or energy[55]. If the approach would be the one of market failure it should be specified in the letter of the law.

1.3. *Is the rationale of the Database Directive in line for Big Data to be protected under the* sui generis *right?*

As Drexl[56] affirms the Database Directive is based on a kind of databases that are very different from Big Data.

The CJEU affirmed that the objective of the Directive is to create incentives for the obtaining of the materials of databases and not for the creation of the data of the database[57]. This interpretation of the rationale of the Directive carries with it two consequences, as Drexl correctly points out[58]: firstly, the objective of the Directive is not the one to safeguard subjects' investment in the creation of IoT products with sensors that collect data, given that those costs are not to be considered for the assessment of whether the investment in the database was «substantial»; secondly, the same is true for the investment in Big Data analysis that lead to the creation of new data.

Moreover, the 15-years term of protection denotes a static view of database technology that is not fit for Big Data that are constantly updated datasets and real-time data services.

Also, exclusive rights that are attributed to the database maker [art. 7(2): extraction and re-utilization] have been interpreted broadly by the case law: the CJEU enlarged the scope of these rights also to indirect extraction and extraction for the compilation of substantially changed and value-added databases[59] and considered the activities of compiling and gathering data from different sources[60], as well as the indexing and copying of the contents of a database by an

---

[54] Leistner (2017), pág. 30; Aplin (2014), pág. 62.

[55] Burdese (2020), pág. 8.

[56] Drexl (2017), pág. 268.

[57] See: Case C-203/02, *The British Horseracing Board Ltd and Others vs. William Hill Organization Ltd.* (2004), ECLI:EU:C:2004:695, para. 31.

[58] Drexl (2017), pág. 268.

[59] Case C-304/07 *Directmedia Publishing vs. Albert-Ludwigs-Universität Freiburg* (2008), ECLI:EU:C:2008:552, paras 29 et seq; see: Leistner (2017), pág. 30.

[60] Case C-202/12, *Innoweb vs. Wegener* (2013), ECLI:EU:C:2013:850, paras 37 et seq. see Leistner (2017), pág. 31.

internet search engine[61], as infringing. As Leistner[62] suggests, the fact that only uses of *substantial* parts of the databases or systematic and repeated extraction of insubstantial parts which add up to be a substantial part of the database without the authorisation of the database maker are considered infringements, is not enough to protect the freedom of competition and to avoid leveraging potentials in typical Big Data uses because in Big Data situations it is always necessary to use substantial parts of the database. This is one more point that shows that the Database Directive is not fit for the (Big) Data economy.

It is also the case to note that the text and data mining exceptions provided for in the Copyright in the DSM Directive[63] are constructed too narrowly to actually change the anti-competitive nature of a database *sui generis* right applied to Big Data[64], especially considering that paragraph 3 of article 4 of that Directive provides that rightsholders can expressly opt-out from the text and data mining exception.

Therefore, for all the above considerations, it is possible to conclude that the rationale of the Database Directive is not in line with the protection of Big Data and of Big Data analysis outputs.

2. The application of the Trade Secrets Directive to Big Data

The definition of trade secret that is outlined in article 2(1) of the Trade Secrets Directive requires that an information to be a trade secret satisfies 3 requirements: 1) it must be secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question; 2) it must have commercial value because it is secret; 3) it must have been subject to reasonable steps, under the circumstances, by the person lawfully in control of the information, to keep it secret.

A first aspect that has been questioned in the application of trade secrets protection to Big Data is whether «data» can be considered to be «information». On the matter, it is submitted that data are information and distinguishing between the syntactic and the semantic level of data is rather artificial, as it was underlined in I.1. In addition to these definitional remarks, it can be relevant to refer to the US Defend Trade Secrets Act[65] that comprises in the realm of information that can be object of trade secrets protection «all forms and types of financial, business, scientific, technical, economic, or engineering information, including [...] *programs, or codes,* whether tangible or intangible, and whether or how stored, compiled, or memorialized physically, electronically, graphically, photographically, or in writing». In contrast, the EU Trade Secrets Directive does not specify what kind of information can be object of trade secrets protection but using a comparative approach can help in saying that Big Data are information protectable as trade

[61] Case C-762/19 *SIA «CV-Online Latvia» vs. SIA «Melons»* (2021), ECLI:EU:C:2021:434, para 47.

[62] Leistner (2017), pág. 31.

[63] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC [2019] OJ L 130/92, art 3-4.

[64] Meys (2020), pág. 457. Musso (2020), pág. 411.

[65] Defend Trade Secrets Act of 2016 Pub. L. No. 114-153 130 Stat. 376.

secrets. However, recital 2 explains that enterprises use trade secrets protection in relation to «a diverse range of information that extends beyond technological knowledge to commercial data such as information on customers and suppliers, business plans, and market research and strategies», while recital 14 excludes that «trivial information» could be protected as trade secrets. It is submitted that these two recitals should be interpreted in the sense that both the raw data and the data inferred by data analysis are object of trade secrets protection, because the triviality of an information is a relative and contextual concept and, in the instance of Big Data, correlations of single data that are «trivial» on their own can actually result in new valuable outcomes, so that «triviality» falls short[66].

As for the three requirements for trade secrets protection, according to Aplin the «fairly low»[67] threshold of commercial value is easily satisfied by data given that there are well-developed markets for data and even in those fields where there are not yet, there is a potential market[68], while the fulfilment of the requirements of secrecy and reasonable steps of course depends on the circumstances of the case but are «flexible enough to embrace datasets and data analysis techniques»[69].

Drexl[70] is more critical and underlines that there are some shortcomings for the application of the 3 requirements of the definition to Big Data situations, especially in the instances where data are produced by smart products (IoT). He argues that the requirement of secrecy is easier to be fulfilled in the case of data collected by sensors within a factory, but the situation becomes more dubious for example if data are collected by smart cars on freely accessible roads because potentially the same information could be collected by the cars of many manufacturers[71]. Nevertheless, it must be noted that this is not a problem at all because the secrecy of art. 2(1) refers to the fact that the information is not known «as a body or in the precise configuration and assembly of its components» and therefore the fact that other manufacturers could capture the same raw data is of no problem because the precise assembly[72] that one enterprise realises trough AI both of the raw data and especially of the inferred data stays secret. However, the literature underlined that the phrase «precise configuration and assembly» denotes a static view of information that is ill suited to Big Data because Big Data are often made of real-time data continuously updated taken by different sources[73]. Moreover, this characteristic of Big Data can cause problems also in relation to the precise identification of the trade secret that is necessary to assess whether there has been a violation[74]. However, it is submit-

---

[66] Shur and Wiebe, (2019), pág. 817; Drexl (2017), pág. 269.

[67] Dessemontet (2008), pág. 282.

[68] Recital 14 of Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure [2016] OJ L 157/1. «[...] Such know-how or information should be considered to have a commercial value, for example, where its unlawful acquisition, use or disclosure is likely to harm the interests of the person lawfully controlling it, in that it undermines that person's scientific and technical potential, business or financial interests, strategic positions or ability to compete [...]».

[69] Aplin (2017), pág. 67.

[70] Drexl (2017), pág. 269.

[71] Drexl (2017), pág. 269.

[72] Shur and Wiebe (2019), pág. 816.

[73] Nordberg (2020), pág. 203.

[74] Aplin (2017), pág. 67; especially in the United Kingdom the case *CMI Centers for Medical Innovation vs. Phytopharm* [1999] FSR 235 has established that one of the conditions for protection is the precise identification of the information that is object of the trade secret.

ted that it is possible to reconcile the nature of Big Data with this requirement of precise configuration: it can be said that the trade secret is referring to some specific kind of information that are categorised «under defined criteria (such as customer mentions in social media, sales, and so on)»[75].

There is also another definitional issue with regards to the secrecy of data that deserves clarification on the side of the Commission: indeed, it is uncertain what «generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question» means. This clarification, that will be dealt with in paragraph IV.2, is even more important given that the Commission aims at enhancing the sharing of data: businesses need to be sure what are the disclosures that lead to the loss of the status of secrecy.

As for the requirement of «commercial value», Drexl argues that the most controversial issue is whether it is possible to demonstrate that data are valuable *because* of their secrecy[76]. An author described demonstrating the casual link between secrecy and commercial value «a procedural shot in the dark» for Big Data[77]. However, Aplin correctly points out that this requirement is easily satisfied by Big Data; moreover, a good interpretation of this requirement is given by the Italian literature and case law that argue that the commercial value can be found not only in a competitive advantage derived from the knowledge of information that a competitor does not have but also in a saving of time and / or money that the availability of the information in question allows[78]. Therefore, a *de minimis* interpretation of this requirement should be applied, given that it is just a threshold, but nothing more[79], and that the fact that the owner wants to enforce his right already signals that there is a commercial value[80].

As for the requirement of reasonable steps, the Directive does not specify what they would imply. It is sure that a contextual and case-by-case interpretation is needed and in the case of Big Data great part of the reasonable steps would imply technical physical measures of protection (*e.g.* encryption).

Another element that needs to be clarified is how to clearly identify a trade secret «owner» in the data economy and in a context where sharing will be encouraged.

While most of the aspects highlighted above (especially the one that relates to the application of the definition of «trade secret» to Big Data), can be solved through a correct interpretation of the text of the law, there are two aspects that should be clarified by the EU legislator. In particular, it should be clarified: 1) who is the «trade secrets holder» and if they should differ from the trade secret «owner»; 2) whether «broadly shared data» can still be considered secret and what it is meant with the phrase «persons within the circles that normally deal with the kind of information in question». These aspects will be dealt with in paragraph IV.2.

---

[75] Nordberg (2020), pág. 204.
[76] Drexl (2017), pág. 269.
[77] Nordberg (2020), pág. 207.
[78] Banterle and Blei (2017), pág. 202.
[79] Dessemontet (2008), pág. 250.
[80] Aplin, Bently, Johnson and Malynicz (2012), pág. 803.

III. POLICY CONSIDERATIONS AND IP JUSTIFICATIONS: TRADE SECRETS ARE BETTER FIT FOR THE PROTECTION OF BIG DATA AND BIG DATA ANALYSIS OUTPUTS

Intellectual property is an artificial creation of the legislation, so there must be a policy reason behind the protection of a given subject matter[81]. Preliminarily it is important to note that there is a big difference among an exclusive absolute right (like the database *sui generis* right) and trade secret protection because the latter is closer to an unfair competition model only forbidding some unfair practices, while allowing independent discovery and reverse engineering, and not creating an exclusive proprietary right on information[82].

In the next paragraphs it will follow an analysis of the justifications that have been elaborated for IP exclusive absolute rights, in particular the utilitarian justification and the market-failure theory (that should provide a justification for the application of the database *sui generis* right to Big Data and Big Data analysis outputs) and it will be shown that those theories applied to Big Data do not lead to the conclusion that there should be a protection as strong as an exclusive proprietary right to incentivise subjects to generate, collect and analyse and share data. Then, the justification for granting trade secrets protection will be analysed and it will be submitted that applying trade secrets protection to Big Data and Big Data analysis is, on the one hand, in line with the EU Data Strategy objective to enhance access and sharing of data, and on the other, it ensures the protection against «third-party interference in undertakings' sphere of confidentiality»[83].

1. Theories for the justification of Intellectual Property absolute rights

There are a series of theories that have been developed by scholars and philosophers to justify the existence of intellectual property protection.

Among these justifications, there are the deontological justifications based on natural law: according to the labour theory people should be granted property rights on their intellectual creation/invention because of the labour they put into it; according to personality theory, intellectual property is granted to protect the individual creator/inventor because of the personal relationship between the work/invention and the creator/inventor; according to reward theory, intellectual property is granted to reward people that enriched society with their creation/invention[84].

As it has been pointed out by the literature[85], when it comes to Big Data and the outputs of Big Data analysis, the establishment of exclusive IP rights based

---

[81] Hilty, Hoffmann and Scheuerer (2021), pág. 52.

[82] Recital 16 of Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure [2016] OJ L 157/1: «In the interest of innovation and to foster competition, the provisions of this Directive should not create any exclusive right to know-how or information protected as trade secrets. [...]».

[83] Drexl, Hilty, Desaunettes, Greiner, Kim, Richter, Surblytė and Wiedemann (2016), pág. 8.

[84] Hilty, Hoffmann and Scheuerer (2021), pág. 52.

[85] *Ibid.,* pág. 56.

on the abovementioned deontological justifications, that are intrinsically connected with the human impact on the creation/invention, is problematic. Indeed, it has been pointed out that the protection could be justified under those theories only if it is assumed that the human impact behind the development of the machines and the tools to analyse the data «lives on in further derivative generations by this tool»[86] or that the human «deserves a reward for making these possible»[87]. However, it is submitted that applying these theories to Big Data and Big Data analysis would be «an overstretch» of the theories themselves[88] that are indeed based on the relevance of human direct input on the work/invention and should not extend to derivative generations like Big Data and Big Data analysis that are fully machine-generated.

Another theory that has been developed is the utilitarian justification, according to which the role of IP is to incentivize innovation and creation in order to benefit the public. According to this theory, intellectual property is either supposed to provide incentives for innovation itself, in particular for socially desirable innovations[89], or for the commercialisation of the subject-matter of protection.

To understand if the utilitarian theory would justify the provision for exclusive rights on Big Data or Big Data analysis, the questions that need to be answered are those that have been formulated by Gervais: «do entities that collect, process and use Big Data *need IP incentives* or *deserve additional rewards* to do what they do? Does the creation of incentives help generate *new or better* data corpora, analyses, and thus produce welfare increases, taking account of welfare losses that rights *in* Big Data might cause, such as increased transaction and licensing costs?»[90].

The issue on whether there is a need for an incentive provided by intellectual property for the generation and collection of data has been issued by the literature within the framework of the debate on the need of a data ownership right. Stakeholders, scholars and the Commission itself noticed that there is no empirical evidence that there is a market failure in relation to the incentives for producing data[91]. Kerber explains that the reason behind this is that the contexts where data are produced are those where the benefits for the data producers are more than the costs[92]. The costs are less than the benefits not necessarily because the cost of producing data are close to zero, but rather because the data producer can defray the expense through the use of the data in-house or through data sharing, data trading and data licensing[93].

But, as it has been pointed out[94], there is only one way to guarantee revenues: the data holder must be able to exclude others from using the data unless they pay for it, and this can easily be done through providing a regime of secrecy.

[86] *Ibid.*, pág. 56.
[87] *Ibid.*, pág. 56.
[88] *Ibid.*, pág. 56.
[89] Mermer (2018), pág. 5.
[90] Gervais (2019), pág. 5.
[91] Kerber (2017), pág. 116.
[92] *Ibid.*, pág. 117.
[93] *Ibid.*, pág. 117.
[94] *Ibid.*, pág. 117.

However, the European Data Strategy, whose aim is to also to boost the B2B sharing of data could take into consideration the role of IP in incentivising the commercialisation of data. The circumstances where it is necessary to have an IP right are those where the creator/innovator cannot fully exploit her creation and therefore licences it in order to commercialise it (*e.g.* copyright)[95]. In the instance of data trading, therefore the question to ask is whether an IP right would result in a more efficient allocation and utilization of data in order not to create market failures[96]. However, to reach these goals there is no need to have a property right but trade secrets protection can easily solve this problem given that also trade secrets can be licenced[97].

2. Trade Secrets Protection justifications and coherence with the EU Data Strategy

As for the justifications for trade secrets protection, it has been argued that the utilitarian theory applies also to commercial secrets and confidential information. Even if it may seem paradoxical, trade secrecy like other kinds of IP actually favours the spread of information. This is because if there was no right to protection of trade secrets, companies would over-invest in maintaining secrets and impose onerous physical and contractual restrictions in an attempt to prevent a competitor from acquiring their information; on the other hand, thanks to the legal protection of commercial secrets, confidential information can circulate more easily even if protected by confidentiality clauses[98].

At the same time, differently from an IP exclusive absolute right, trade secrets holders can only benefit from relative protection against the forbidden acts listed in article 4 of the Trade Secrets Directive; the European Commission indeed in the Explanatory Memorandum for the Proposal of the Directive clearly stated that the holder of a trade secret does not have exclusive proprietary rights (therefore there is no such thing as «ownership» of data) over the information covered by the trade secret, however restrictions to the use of the trade secret are justified in cases where the relevant know-how or information has been obtained from the trade secret holder against its will by a third party through dishonest means[99].

As there is not a need for an exclusive (intellectual) property right in relation to Big Data, like in the case of the database *sui generis* right where the owner captures almost all the commercial benefits of the use of that data[100], it is reasonable to grant protection against misappropriation as a consequence of unfair practices through trade secrets law to ensure that the legitimate interests of businesses are preserved even in a context where access to information and sharing are encouraged. This is because businesses will give access to Big Data and Big Data analysis outputs, knowing that there is a legal protection by trade

[95] Drexl (2017), pág. 274.
[96] Kerber (2016), pág. 995.
[97] Hull (2009), pág. 203.
[98] Sappa (2019), pág. 414; See Lemley (2008), pág. 311.
[99] European Commission, *Proposal for a Directive of the European Parliament and of the Council on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure*, 28 November 2013, COM(2013) 813 final, pág. 3.
[100] Davison (2003), pág. 240.

secrets law, and they will share and authorise the use of the data that are in their control by contracting with other businesses availing themselves of non-disclosure clauses and trade secrets licences where the permitted uses of the shared data are agreed upon among the parties[101].

Even if the concepts of secrecy and access/sharing might seem in contrast, actually, trade secrets protection is in line with the objective of the European Commission to encourage the sharing of data given that one rationale for trade secrets protection is to allow a safe sharing of information without incurring in excessive transaction costs[102].

Moreover, considering that protection is conditioned on secrecy and on the reasonable steps that the companies will implement, it is submitted that this means that not all data that have given characteristics will fall under the protection (like it would happen with the database *sui generis* right[103]), but rather only those that are subject to those measures to keep them secret. The benefit of this is that it gives room to situations where the economic actors will be free to decide if they want to protect some data as trade secrets or if they want to share them on altruistic grounds, without subjecting them to NDAs.

Also, the Trade Secrets Directive already provides for some instruments that could allow the establishment of compulsory access rights that could be established by the EU Data Law. Indeed, article 3(2) states that the acquisition, use or disclosure shall be considered lawful to the extent that such acquisition, use or disclosure is required or allowed by Union or national law, and art. 5 (d) establishes an exception for the acquisition, use or disclosure of a trade secret for the purpose of protecting a legitimate interest recognised by Union or national law. As it has been underlined by the Study of the IPOL at the request of the European Parliament's Committee on Legal Affairs[104], access rights foreseen by EU law could be considered as relevant legitimate interests or obligations recognised/set out by Union law. However, access rights should prevail compared to the interest to protect secret information only under the strict conditions of article 102 TFEU[105].

As a final consideration, it is submitted that, a situation where there is no legal means to ensure at least some sort of excludability (like the kind of excludability ensured by trade secrets) is not to be encouraged, because while broader access and sharing of data is a an important goal, «undertakings can remain competitive only if they possess a certain degree of autonomy in business operations»[106] and to totally deprive enterprises from controlling the data that result from their business activity is not a fair nor proportionate compromise when it comes to B2B relationships (that are the relationships this paper is concerned with).

---

[101] See Hull (2009), pág. 212 for the challenges posed by trade secrets licences: «The experience of many practitioners is that this can be done, despite the "inherently perishable" nature of the subject matter and that licensing in this field, though more complex that in related areas is perfectly possible».

[102] Rowe and Sandeen (2015), pág. 12.

[103] European Commission, *Commission Staff Working Document Evaluation of Directive 96/9/EC on the legal protection of databases*, 25 April 2018, SWD(2018)146 final, pág. 79: «Many makers do not want to automatically acquire sui generis right».

[104] Policy Department for Citizens' Rights and Constitutional Affairs Directorate-General for Internal Policies, *IPR and the use of open data and data sharing initiatives by public and private actors*, May 2022, pág. 64.

[105] *Ibid*., pág. 64.

[106] Drexl, Hilty, Desaunettes, Greiner, Kim, Richter, Surblytė and Wiedemann (2016), pág. 8.

IV. *DE LEGE FERENDA* CONSIDERATIONS

In the following paragraphs, a proposal for the modification of the Database Directive and for the clarification of specific aspects of the Trade Secrets Directive will be put forward.

1. Proposal for the modification of the Database Directive

The application of the Database Directive – in particular the database *sui generis* right – to Big Data is debated as it was shown before. The European Commission's aim is to amend it in way that does not pose an obstacle to the access and use of machine-generated data and facilitate the sharing of those data[107]. Through the Proposal for the Data Act, the Commission intends to clarify that the database *sui generis* right does not apply to machine-generated data.

This decision on the part of the Commission is very welcome as it finally settles a debate that has been causing legal uncertainty, however it is submitted that, given that there is not a satisfactory justification for an exclusive proprietary right in *any kind* of data, the *sui generis* right should be abolished, or at least not applied neither to raw data nor to inferred data in Big Data contexts. The European Commission itself admitted that «the *sui generis* right continues to have no proven impact on the production of databases»[108]. The decision of the Commission is already moving towards the right direction, even if it could take a broader approach by including that inferred data are not protected by the *sui generis* right.

Considering that the Proposal of the Data Act will be discussed during the legislative process, and it could be amended, it is important to consider some submissions that are in favour of applying the database *sui generis* right to Big Data and to refute them.

Mermer[109] affirmed that intellectual property not only serves the purpose to incentivise innovation but also to *control* innovation. Therefore, it could be said that the database *sui generis* right could be used as a legislative instrument that actually makes sure that the rules for access to and use of data are fair, practical and clear. Burdese[110], as a supporter of this position, highlights two points that are worth of careful consideration.

First, the European Court of Justice in the *Rayanair* case[111] (2015) affirmed that articles 8 («Rights and obligations of lawful users») and 15 («Binding nature of certain provisions») do not apply to databases that are not protected under the *sui generis* right and the contractual limitations on its use by third parties can be more restrictive that what is permitted under the Directive. This decision led to a paradox, because it suggested that the maker of an unprotected

---

[107] European Commission, *Inception Impact Assessment on the Data Act (including the review of the Directive 96/9/EC on the legal protection of databases)*, 28 May 2021, pág. 4.

[108] European Commission, *Commission Staff Working Document Evaluation of Directive 96/9/EC on the legal protection of databases,* 25 April 2018, SWD(2018)146 final, pág. 19.

[109] Mermer (2018), pág. 14.

[110] Burdese (2020), pág. 9.

[111] Case C-30/14 *Ryanair Ltd vs. PR Aviation BV* (2015), ECLI:EU:C:2015:10.

database that has a strong bargaining power can enjoy greater protection than that of the maker of a protected database and not be burdened by the mandatory exceptions[112].

Secondly, Burdese, relying on the Ryanair case, upholds that: it would not be necessary to extend the database *sui generis* protection to machine-generated data but it could possible to just extend the application of article 8 of the Directive to all databases falling within the definition set forth in article 1(2)[113].

Burdese also proposes to solve the problem of access to data in key sectors through the modification of the Database Directive through the broadening of the scope of the *sui generis* right (so that to encompass cases in which creation and collection of data are not separable, in particular sensor-generated and mined data) and the provision of mandatory licences (with FRAND terms)[114].

This line of reasoning could lead to think that there should be some sort of legislative control over the terms of the contract and that the best way to do this is to extend some provisions of the Database Directive to all «databases», even if the requirement of substantial investment in obtaining, verifying and presenting the contents is not present, or even expand the scope of application of the Directive altogether to machine generated data and add a provision on mandatory licences and apply the exceptions to the *sui generis* right to all Big Data.

While in the latter case there would then be a provision of an exclusive proprietary right for Big Data, that, as it has been explained earlier, would be totally unjustified, the former approach would be, at best, not in line with the purpose of IP law. The extension of intellectual property provisions to subject matter that does not have the requirements provided for by the law just to ensure fair contractual terms is really a stretch of the function of IP law.

The role of intellectual property is most certainly not the one to ensure fairness in contracts or provide for access rights, but it is the one to incentivise the production of some intangible goods that would not be produced as much without IP protection because of a market failure.

Rather, there are specific bodies of laws (contract law, unfair competition, antitrust law) and specific solutions that can be adopted. For example, it is submitted that the best way to ensure fair contractual practices is through the development of non-mandatory best practices and, in situations where there is a power asymmetry between the parties, of mandatory rules on control of unfair B2B standard terms and conditions[115].

---

[112] Burdese (2020), pág. 9.

[113] *Ibid*., pág. 12.

[114] *Ibid*., pág. 13.

[115] Leistner (2017), págs. 38-39; see Chapter IV of European Commission, *Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)*, 23 February 2022, COM(2022) 68 final, pág. 16: «Chapter IV addresses unfairness of contractual terms in data sharing contracts between businesses, in situations where a contractual term is unilaterally imposed by one party on a micro, small or medium-sized enterprise. This Chapter guarantees that contractual agreements on data access and use do not take advantage of imbalances in negotiating power between the contractual parties. The instrument of an unfairness test includes a general provision defining unfairness of a data sharing-related contractual term complemented by a list of clauses that are either always unfair or presumed to be unfair. In situations of unequal bargaining power, that test protects the weaker contractual party in order to avoid unfair contracts. Such unfairness impedes the use of data by both contractual parties. With that, the provisions ensure a fairer allocation of value in the data economy. Model contractual terms recommended by the Commission may assist commercial parties in concluding contracts based on fair terms».

2. Proposal for the clarification of the Trade Secrets Directive

As much as the trade secrets protection of Big Data is a good solution that fairly balances the interests of businesses and the public interest in data availability, there are some aspects of the current Directive that deserve clarification. However, the European Commission did not deal with this aspect in the Proposal of the Data Act. It is submitted that the following elements should be tackled by the future Data Act.

A clarification of article 2(2) of the Directive would be welcomed. According to article 2(2) «trade secret holder» means any natural or legal person lawfully controlling a trade secret. This definition poses two problems. First, the notion of «control of a trade secret» is not further expanded by the Directive, therefore it is not clear whether it would comprehend licencees, subjects who reverse engineered or subjects to whom the information had been disclosed with consent[116]. In the Recitals there is not a clarification, however in the Explanatory Memorandum of the Proposal of the Directive, the Commission specified that the concept of lawfulness of control «ensures that not only the original owner of the trade secret but also licensees can defend the trade secret»[117]. Therefore, looking at the *intentio legislatoris,* it can be derived that the trade secret holder is not the same as the «trade secret owner» that is the original/first holder of the information. It is submitted that, as it has been suggested by one author[118], the definition of a trade secret owner should be added to the Directive, and they should be identified as the subject that has to implement the reasonable steps.

The second problem relates to the notion of control, especially in the data economy. The characteristic of the data economy is that economic value is increased in networks and not anymore in vertical value chains[119]. In the networked environment where data are integrated both within the single enterprise but also among more businesses (horizontal integration) to maximise value, and where businesses need to use cloud services to face the enormous storage capacity that is required, the «control» of the data is difficult to allocate[120]. It is the opinion of the author that the Commission should clarify what «control» entails by explaining that the control does not have to be on the data itself but on the inferred data that result from the data analysis[121], while the data may have been taken from different sources on which the trade secrets holder does not necessarily have control.

Another element that needs clarification, that has been highlighted by the position statement of the Max Planck institute[122], is to what extent broadly shared data through commercial transactions based on NDAs can still benefit of the trade secrets protection and be considered «secret» and more generally, what is

---

[116] Aplin (2017), pág. 69.

[117] European Commission, *Proposal for a Directive of the European Parliament and of the Council on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure*, 28 November 2013, COM(2013) 813 final, pág. 7.

[118] Surblytė-Namavičienė (2020), pág. 123.

[119] Drexl (2017), pág. 265.

[120] Shur and Wiebe (2019), pág. 815.

[121] Nordberg (2020), pág. 216.

[122] Drexl, Hilty, Desaunettes-Barbero, Globocnik, Gonzalez Otero, Hoffmann, Kim, Kulhari, Richter, Scheuerer, Slowinski and Wiedemann (2021), pág. 4.

the relationship between secrecy and access to information. According to Wiebe and Shur[123], trade secrets protection in the data economy is built on different premises in respect to traditional enterprises: in the data economy characterised by a networked environment the relativity of secrecy is even more significant because access and cooperation are a key feature of these kinds of businesses. A clarification is therefore needed in relation to the interpretation of the phrase «persons within the circles that normally deal with the kind of information in question». In general, as long as there are NDAs in place, it should still be possible to maintain the secrecy status, and this would be in line with the fact that one of the rationales behind trade secrets protection is to encourage disclosure and discourage real secrecy[124]. But an unclear aspect is whether a database licensed, under confidentiality, to the majority of businesses in a particular sector should be considered «generally known» in the relevant circles[125]. The answer to this question has to balance the tension that permeates the data economy between the need of exclusivity and the need to share data. It is the author's opinion that this phrase should be interpreted broadly, and the trade secrets protection should be maintained to the highest extent possible. The correct interpretation should be that the information is «generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question» when the competitive advantage of the original trade secrets holder is destroyed[126].

This conclusion is also linked to the fact that if trade secrets protection is the only legal protection afforded to data and the B2B sharing is to be encouraged, this interpretation is the only one that can fulfil both objectives. This interpretation is even more necessary because one of the objectives of the European Data Strategy is to create data pools where data are shared by the businesses and operators of a given sector in order to enable Big Data analytics and Machine Learning. If businesses are afraid to lose the only legal protection afforded if they join data pools, they would be discouraged to do so.

V. CONCLUDING REMARKS AND THE EC PROPOSAL FOR THE DATA ACT

The European Data Strategy launched by the European Commission in February 2020 represents an ambitious project and a new step for European Data Law. It moves from the focus to privacy and finally sees data as a resource whose abundance is to be welcomed. Data are both a resource that businesses can benefit from but also an asset that can give a competitive edge to those that have them. This paper focused in particular on the application of European intellectual property instruments to Big Data (both raw data and derived data), and it showed how there are several uncertainties that relate to the application of the database *sui generis* right (that the Data Act is aimed at solving in art. 35) and of the trade secrets protection. An analysis of intellectual property justifications demonstrated that an exclusive property right is

[123] Shur and Wiebe (2019), pág. 818.
[124] Sappa (2019), pág. 414.
[125] European Commission (2016), *Legal study on Ownership and Access to Data*.
[126] Ottolia (2017), pág. 63.

not justified and the that the database *sui generis* right should be abolished or at least not applied neither to raw data nor to derived data. On the other hand, the Trade Secrets Directive is useful in the data economy because it is aimed at banning some unfair behaviours, but it does not create an exclusive right on data. However, a clarification of the Directive could be necessary in respect to two aspects: namely, the identity of the trade secrets owner and the issue of «broadly shared» data.

It has been shown that the role of intellectual property in the data economy is shifting. On the one hand cooperation and access are the characteristic features of the new economy so a proprietary regime is not needed nor wanted by stakeholders, but at the same time businesses need to be able to make money out of data and data-enabled services so some sort of excludability of data needs to be ensured through secrecy. Indeed, «information wants to be free, but information also wants to be expensive because it is immensely valuable»[127].

The Draft Proposal for the Data Act has taken a step forward because it attempts at clarifying the application of the *sui generis* right by excluding that it covers machine generated data (i.e. raw data), however a broader approach should be attempted and a repealing of the *sui generis* right both for raw data and inferred data, alongside the strengthening and clarification of trade secrets protection could benefit even more the EU data economy.

VI. BIBLIOGRAPHY

APLIN, Tanya (2014), *Copyright Law in the Digital Society: The Challenges of Multimedia*, Hart Publishing Limited, Oxford, UK.

— (2017), «Trading Data in the Digital Economy: Trade Secrets Perspective», in LOHSSE, S.; SCHULZE, R., and STAUDENMAYER, D. (eds), *Trading Data in the Digital Economy: Legal Concepts and Tools*, Hart Publishing, Baden-Baden, Germany, págs. 59-72.

APLIN, Tanya; BENTLY, Lionel; JOHNSON, Phillip, and MALYNICZ, Simon P. (2012), *Gurry on Breach of Confidence: The Protection of Confidential Information*, Oxford University Press, Oxford.

BANTERLE, Francesco, and BLEI, Marco (2017), «Alcune novità introdotte dalla Direttiva Trade Secrets», *Riv. dir. ind.* 4-5, págs. 202-224.

BOYD, Danah, and CRAWFORD, Kate (2012), «Critical Questions for Big Data», *Information, Communication & Society* 15(5), págs. 662-679.

BRAND, Stewart (1987), *The Media Lab: inventing the future at M.I.T.,* Viking Penguin, New York, US.

BRUZZONE, Ginevra, and DEBACKERE, Koenraad (2021), «As Open As Possible, As Closed As Needed: Challenges Of The EU Strategy For Data», *les Nouvelles* March 2021, págs. 41-49.

BURDESE, Paolo (2020), «AI-generated databases. Do the creation/obtaining dichotomy and the substantial investment requirement exclude the sui generis right provided for under the EU Database Directive? Reflections and proposals», *WIPO Academy, University of Turin and ITC-ILO - Master of Laws in IP - Research Papers Collection - 2019-2020.*

CORRALES COMPAGNUCCI, Marcelo (2020), *Big Data, Databases and «Ownership» Rights in the Cloud*, Springer, Singapore.

[127] BRAND (1987), pág. 202.

DAVISON, Mark J. (2003), *The legal protection of databases*, Cambridge University Press, Cambridge, UK.

DE MAURO, Andrea; GRECO, Marco, and GRIMALDI, Michele (2016), «A Formal Definition of Big Data Based on Its Essential Features», *Library Review* 65(3), págs. 122-135.

DESSEMONTET, François (2008), «Protection of Trade Secrets and Confidential Information», in CORREA, C. M., and YUSUF, A. A. (eds), *Intellectual Property and International Trade: The TRIPs Agreement*, Kluwer Law Intl, London, págs. 271-292.

DREXL, Josef (2017), «Designing Competitive Markets for Industrial Data», *JIPITEC* 8(4), págs. 257-292.

DREXL, Josef; HILTY, Reto; DESAUNETTES, Luc; GREINER, Franziska; KIM, Daria; RICHTER, Heiko; SURBLYTĖ, Gintarė; WIEDEMANN, Klaus (2016), «Data ownership and access to data Position statement of the Max Planck Institute for innovation and competition of 16 august 2016 on the current European debate», Max Planck Institute for Innovation and Competition Research Paper NO. 16-10.

DREXL, Josef; HILTY, Reto; DESAUNETTES-BARBERO, Luc; GLOBOCNIK, Jure; GONZÁLEZ OTERO, Begoña; HOFFMANN, Jörg; KIM, Daria; KULHARI, Shraddha; RICHTER, Heiko; SCHEUERER, Stefan; SLOWINSKI, Peter, and WIEDEMANN, Klaus (2021), «Artificial Intelligence and Intellectual Property Law Position Statement of the Max Planck Institute for Innovation and Competition of 9 April 2021 on the Current Debate», Max Planck Institute for Innovation and Competition Research Paper 21-10.

FLORIDI, Luciano (2022), *Etica dell'intelligenza artificiale*, Raffaello Cortina Editore, Milano, Italy.

GERVAIS, Daniel (2019), «Exploring the Interfaces Between Big Data and Intellectual Property Law», *JIPITEC* 10(3), págs. 22-38.

HARTMANN, Philipp Max; ZAKI, Mohamed; FELDMANN, Niels, and NEELY, Andy (2016), «Capturing value from big data - a taxonomy of data-driven business models used by start-up firms», *International Journal of Operations & Production Management* 36(10), págs. 1382-1406.

HILTY, Reto M.; HOFFMANN, Jörg, and SCHEUERER, Stefan (2021), «Intellectual Property Justification for Artificial Intelligence», in LEE, J.; HILTY, R. M., and LIU, K. (eds.), *Artificial Intelligence and Intellectual Property*, Oxford University Press, Oxford, UK, págs. 50-72.

HUGENHOLTZ, P. Bernt (2017), «Data Property in the System of Intellectual Property Law: Welcome Guest or Misfit?», in LOHSSE, S.; SCHULZE, R., and STAUDENMAYER, D. (eds), *Trading Data in the Digital Economy: Legal Concepts and Tools*, Hart Publishing, Baden-Baden, Germany, pág. 75-99.

HULL, John (2009), «Trade secret licensing: the art of the possible», *JIPLP* 4(3), págs. 203-212.

HURLEY, Richard (2019), *Big Data: A Guide to Big Data Trends, Artificial Intelligence, Machine Learning, Predictive Analytics, Internet of Things, Data Science, Data Analytics, Business Intelligence, and Data Mining*, Amazon Italia Logistica, Milano, Italy.

KERBER, Wolfgang (2016), «A new (intellectual) property right for non-personal data? An economic analysis», *GRUR Int*. 11/2016, págs. 989-999.

— (2017), «Rights on Data: The EU Communication "Building a European Data Economy" from an Economic Perspective», in LOHSSE, S.; SCHULZE, R., and STAUDENMAYER, D. (eds), *Trading Data in the Digital Economy: Legal Concepts and Tools*, Hart Publishing, Baden-Baden, Germany, págs. 109-133.

KITCHIN, Rob (2014), *The Data Revolution*, SAGE Publications, London, UK.

LEISTNER, Matthias (2017), «Big Data and the EU Database Directive 96/9/EC: Current Law and Potential for Reform», in LOHSSE, S.; SCHULZE, R., and STAUDENMAYER, D. (eds), *Trading Data in the Digital Economy: Legal Concepts and Tools*, Hart Publishing, Baden-Baden, Germany, págs. 27-57.

LEMLEY, Mark (2008), «The Surprising Virtues of Treating Trade Secrets as IP Rights», *Stanford Law Review* 61(2), págs. 311-353.

LUNDQVIST, Björn (2021), «The Proposed Digital Markets Act and Access to Data: A Revolution, or Not?», *IIC* 52, págs. 239-241.

MERMER, Janset Ece (2018), «The Impact of Intellectual Property in Fostering Innovation», *Institutions & Transition Economics: Microeconomic Issues eJournal SSRN*.

MEYS, Romain (2020), «Data Mining Under the Directive on Copyright and Related Rights in the Digital Single Market: Are European Database Protection Rules Still Threatening the Development of Artificial Intelligence?», *GRUR Int*. 69(5), págs. 457-473.

MUSSO, Alberto (2020), «Eccezioni e limitazioni ai diritti d'autore nella Direttiva UE n. 790/2019», *Il Diritto dell'Informazione e dell'Informatica* XXXV(4), págs. 411-464.

NORDBERG, Ana (2020), «Trade secrets, big data and artificial intelligence innovation: a legal oxymoron?», in SCHOVSBO, J.; MINSSEN, T., and RIIS, T. (eds), *The harmonization and protection of trade secrets in the EU*, Edward Elgar Publishing, Cheltenham, UK, págs. 192-218.

OTTOLIA, Andrea (2017), *Big Data e innovazione computazionale*, Giappichelli Editore, Torino, Italy.

ROWE, Elizabeth, and SANDEEN, Sharon (2015), *Trade Secrecy and International Transactions Law and Practice*, Edward Elgar Publishing, Cheltenham, UK.

SAPPA, Cristiana (2019), «How data protection fits with the algorithmic society via two intellectual property rights - a comparative analysis», *JIPLP* 14(5), págs. 407-418.

SCHOVSBO, Jens Hemmingsen, and KOKOULINA, Olga (2020), «Cutting Into Diamonds: Competition Law, IPR, Trade Secrets and the Case of 'Big Data», *University of Copenhagen Faculty of Law Research Paper No. 2020-94*.

SHUR, Nico, and WIEBE, Andreas (2019), «Protection of trade secrets in a data-driven, networked environment - Is the update already out-dated?» (2019), *JIPLP* 14(10), págs. 814-821.

STREINZ, Thomas (2021), «The Evolution of European Data Law», in CRAIG, P., and DE BÚRCA, G. (eds), *The Evolution of EU Law,* Oxford University Press, Oxford, UK, págs. 902-936.

STROWEL Alain (2020), «Big data and data appropriation in the EU», in APLIN, T. (ed), *Research Handbook on Intellectual Property and Digital Technologies (*Research Handbook Edward Elgar, Cheltenham, UK, págs. 107-135.

SURBLYTĖ-NAMAVIČIENĖ, Gintare (2020), *Competition and Regulation in the Data Economy*, Elgar Studies in Law and Regulation, Cheltenham, UK.

VEZZOSO, Simonetta (2021), «The Dawn of Pro-Competition Data Regulation for Gatekeepers in the EU», *European Competition Journal* 17(2), págs. 391-406.

ZECH, Herbert (2016), «A legal framework for a data economy in the European Digital Single Market: rights to use data», *JIPLP* 11(6), págs. 460-470.