

AIDA

ANNALI ITALIANI DEL DIRITTO D'AUTORE
DELLA CULTURA E DELLO SPETTACOLO

Anno XXVII

2018

Isbn 9788828809807

(Estratto)

GUIDO NOTO LA DIEGA

Artificial Intelligence and databases in the age of big
machine data



GIUFFRÈ FRANCIS LEFEBVRE

Artificial Intelligence and databases in the age of big machine data (*) (**)

« Hoc qui existimat fieri potuisse, non intellego cur non idem putet, si innumerabiles unius et viginti formae litterarum vel aureae vel qualeslibet aliquo coiciantur, posse ex iis in terram excussis annales Ennii ut deinceps legi possint effici; quod nescio an ne in uno quidem versu possit tantum valere fortuna »

Cicero, *De natura deorum*, 2.37

SUMMARY. 1. Introduction. – 2. Artificial intelligence, machine learning, deep learning: definitions and developments. – 3. Context and definitions in the Database Directive. – 4. Database copyright in *cerca d'autore*. – 5. A sui generis right for a sui generis database? – 6. Infringement, exceptions, and data mining in the Digital Single Market. – 7. The AI database's owner armoury beyond the Database Directive: technical protection measures, contracts, and unfair competition. – 8. Not the end of the road.

1. This paper deals with the databases made by means of Artificial Intelligence (1) technologies (hereinafter 'AI databases'), which can intervene in the stage of obtaining the contents of a databases, as well as in the verification and presentation (2). The overarching research question is whether AI databases can be protected under the copyright and sui generis regimes provided by the Database Directive (3). The topic is interwoven with some of the most pressing issues in copyright law, i.e. the alleged unfitness of the sui generis right for machine data and big data (4), the originality of dehumanised works (5), the text and data mining exception (6) in the

(*) Questo scritto ha ricevuto un giudizio positivo di un referee.

(**) The author is indebted to Estelle Derclaye, Marco Ricolfi, Rossana Ducato, James C. Bell, and an anonymous reviewer for their helpful suggestions. Thanks to Valentina Borgese and Alessia Palladino for the precious research assistance. Opinions and errors are solely the author's.

(1) AI, umbrella term for autonomous and 'intelligent' technologies, will be analysed in the next section.

(2) Conversely, this paper will leave out those databases whose makers use AI as an aid, but where the human element remains prevalent.

(3) 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 (hereinafter the Database Directive or the Directive).

(4) JIIP and Technopolis Group, Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases – Final report, European Commission, Brussels, 2018, v.

(5) See, e.g., IHALAINEN, *Computer creativity: artificial intelligence and copyright*, in *II-PLP* 2018, IX, 724; RAMALHO, *Will Robots Rule the (Artistic) World? A Proposed Model for the Legal Status of Creations by Artificial Intelligence Systems (13 June 2017)*, available at <https://ssrn.com/abstract=2987757>; BRIDY, *Coding Creativity: Copyright and the Artificially Intelligent Author*, in *Stan Tech L Rev* 2012, V, 1; GUADAMUZ, *Do androids dream of electric copyright? Comparative analysis of originality in artificial intelligence generated works*, in *IPQ* 2017, II, 169.

context of the proposed EU reform of copyright (7), and data ownership (e.g. the infamous data producer's right) (8).

In 2018, the European Commission evaluated the impact of the Database Directive, with a focus on the sui generis right and machine data (9). When the directive was drafted, it was the time of CDs (10), and the now forgotten song 'End of the Road' was dominating the Billboard charts. Since then, technologies have been developing dramatically, with AI providing unprecedented tools to extract value from Big Machine Data (11), and databases coming back into style thanks to the blockchain (12) and other distributed ledger technologies (13). The argument has been put forward that the database protection regime is no longer adequate, also in light of the importance of the data economy, whose overall value reached € 300 billion in 2016, and it is set to increase to € 739 billion with an overall impact of 4% on the EU GDP by 2020 (14). Conversely, the database market in EU is utterly stagnant (15). Said inadequacy led the European Parliament to call on the Com-

(6) Hereinafter, text and data mining will be referred to as simply data mining for readability purposes.

(7) Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market (COM/2016/0593 final) (Draft Copyright Directive in the Digital Single Market).

(8) See Commission Staff Working Document on the Free Flow of Data And Emerging Issues of the European Data Economy - accompanying the document Communication Building a European data economy, COM(2017) 9 final (10 January 2017).

(9) Commission Staff Working Document, *Evaluation of Directive 96/9/EC on the legal protection of databases* {SWD(2018) 147 final}.

(10) The fact that the draft Database Directive had the CD-ROM market as reference is confirmed by the Explanatory Memorandum to the Directive. For a history of the introduction of the Database Directive, see DAVISON, *The legal protection of databases*, Cambridge University Press, Cambridge, 2003, 51 ff..

(11) By 'Big Machine Data', we mean the big quantity of data produced by machines, in particular in the context of the Internet of Things, AI, machine-to-machine (M2M) communication, Industry 4.0, cloud computing, and robotics.

(12) The blockchain in the technology used by cryptocurrencies such as BitCoin and Ether, but goes beyond that, being "a type of database that takes a number of records and puts them in a block [...] each block is then 'chained' to the next block, using a cryptographic signature" (UK GOVERNMENT CHIEF SCIENTIFIC ADVISER, *Distributed Ledger Technology: Beyond block chain*, 2016, Government Office for Science, London, 17). Its main features are a persistent, tamper-evident record of the relevant transactions (e.g. smart contracts) and an infrastructure to authenticate the parties of the transaction. BACON ET AL., *Blockchain Demystified*, Queen Mary University of London, School of Law Legal Studies Research Paper No. 268/2017.

(13) Distributed ledgers are a type of database that is spread across multiple sites, "[r]ecords are stored one after the other in a continuous ledger, rather than sorted into blocks, but they can only be added when the participants reach a quorum" (UK GOVERNMENT CHIEF SCIENTIFIC ADVISER, *op. cit.*, 17-18).

(14) IDC and OPEN EVIDENCE, *European data market. Final report*, European Commission, Brussels, 2017, 126.

(15) JIIP and TECHNOLIS GROUP, *Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases – Annex 2: Economic analysis*, European Commission, Brussels, 2018.

mission to abolish the sui generis right (16) and the Commission to decide to keep it (17) while suggesting the introduction of a data producer's right as a new property that would have done what the sui generis right had been unable to and, therefore, incentivise investments in data production (18).

It is this paper's contention that, contrary to popular belief, the sui generis right is fit for AI databases and that a different solution would lead to an overprotection of said subject matter by contractual means. Moreover, if the sui generis right were revitalised, there would be a strong argument to reject any proposal of data property. This contention has crucial consequences, because the sui generis right may be the best, if not the only, way to protect AI 'authorial' works. Indeed, even if AI works currently fall outside the scope of copyright law for lack of originality, they could nonetheless be protected if part of a database. Thus, thanks to AI, the sui generis right may become more important than it ever was.

In terms of methods, this work focuses on statutes and case law on databases from an EU perspective. National implementations will be considered only marginally and the UK will be the main reference for a threefold reason (19). First, the UK has been the first Member State to regulate computer-generated works (20). Second, whilst in principle AI works can hardly be considered original under the EU standard of originality – the author's own intellectual creation (21) – they might be under the British one, i.e. skill, labour, or judgement (22). Third, the UK is the most productive database maker in the EU (23). The paper looks at the law as it

(16) European Parliament resolution of 19 January 2016 on Towards a Digital Single Market Act (2015/2147(INI)) (2018/C 011/06), para 108.

(17) Commission, *Evaluation*, cit..

(18) Commission, *Free flow of data*, cit..

(19) Copyright and Rights in Databases Regulations 1997, S.I. 1997 n. 3032, amending the Copyright, Designs and Patents Act 1998 (Part II), and introducing the 'database right' (Part III). Not all Member States were timely in the transposition. It is noteworthy that the Court of Justice's first rulings on the Database Directive found the Member States liable for non-transposition of the Directive. See Court of Justice 13 April 2000, *Commission v Luxembourg*, case C-348/99, in ECR, 2000, I, 2917 and Court of Justice 11 January 2001, *Commission v Ireland*, case C-370/99, *ivi*, 2001, I, 297.

(20) Copyright, Designs, and Patents Act 1988, Section 9(3), 12(7), 79(2)(c), 81(2), 178, 214, 263.

(21) Court of Justice 16 July 2009, *Infopaq*, case C-5/08, in *European Court Reports* 2009, I, 6569, para. 36. See VALENTI's comment in this *Journal*, 2009, 428 ff., DERCLAYE, *Wonderful or Worrisome? The Impact of the ECJ ruling in Infopaq on UK Copyright Law*, in *EIPR* 2010, V, 247 ff., and TRABUCO, *Com onze palavrinhas apenas (...): a reprodução temporária de obras e a actividade de press clipping*, in *Cadernos de Direito Privado* 2009, 38 ff..

(22) In theory, after *Infopaq*, the UK adopted the EU standard of originality, though this is said to have no practical consequences in *The Newspaper Licensing Agency and others v Meltwater Holding BV and others* [2011] EWCA Civ 890. However, with the UK leaving the EU, the country may return to the previous, arguably lower and certainly more AI-friendly, standard of originality.

(23) JIIP and TECHNOLIS GROUP, *Final report*, cit., 6. This is used as an argument to assert that it is unlikely that the EU-stemming database regime will be repealed by the UK after Brexit. This has been argued by RAMALHO and GOMEZ GARCIA, *Copyright after Brexit*, in *JIPLP* 2017, VIII, 669, 670, who refer to the data presented in DG Internal Market and Services Working Paper, *First evaluation of Directive 96/9/EC on the legal protection of data-*

currently stands, and it assesses whether and how it can be applied in the selected technological environment. De lege ferenda considerations will be kept to a minimum. The focus of this paper is on AI making databases, but AI is relevant from a database perspective also because the relevant technologies can be used to infringe database rights, prevent their infringement, and enforce them (e.g. filters and automated takedown procedures). These aspects will be dealt with only in so far as they can contribute to answer the main research question.

The structure of this paper is as follows. The first part will introduce AI technologies and their relevance from a database perspective. Second, the database copyright will be analysed more in – depth to resolve the problems of non – human authorship, originality, and ownership of AI databases. Third, a study of the aspects of the sui generis right which are more relevant and problematic follows. Fourth, infringement and exceptions will be critically analysed, with a focus on the legality of AI-powered data mining. Before concluding, an analysis of the database protection beyond the Database Directive will be presented, with a focus on technological protection measures (TPMs), contracts, and unfair competition.

2. In observing the AI debate, it has been noted (24) that there is a polarisation between ‘Singularitarians’ (25) and ‘AItheists’ (26). The former are sure that true superintelligence is around the corner and it will disrupt everything we know, thus leading to an apocalyptic scenario where human labour will become useless and human being will become the machines’ slaves. In turn, the latter argue that even imagining an intelligent machine is preposterous and in any case no real disruption will come, since we will be able to keep machines under our control. While there is no agreement on the timescales and the degree to which AI will change our world, it is accepted that this set of technologies is already having a palpable impact on several aspects of our life (27), with wide-ranging consequences, including the necessity to rethink some intellectual property principles and rules (28).

bases (Brussels, 12 December 2005).

(24) L. Floridi, *Should we be afraid of AI?*, in *Aeon* (9 May 2016), available at <https://aeon.co/essays/true-ai-is-both-logically-possible-and-utterly-implausible>.

(25) For those who are not familiar with this kind of literature, the reference is to K. KURZWEIL, *The singularity is near: When Humans Transcend Biology* (New York: Viking, 2006). An eminent intellectual belonging to this class was Stephen Hawking, who said that ‘the development of full artificial intelligence could spell the end of the human race.’ (R. CELLAN-JONES, *Stephen Hawking warns artificial intelligence could end mankind* (2 December 2014), available at <http://www.bbc.co.uk/news/technology-30290540>).

(26) See, eg. J.R. SEARLE, ‘What your computer can’t know’ *The New York Review of Books* (2014).

(27) APPENZELLER, *The AI revolution in science*, in *Science*, 7 August 2017 and HARARI, *Reboot for the AI revolution*, in *Nature*, 2017, VMMDCCLXXVI, 324.

(28) Cf., to name some of the relevant recent literature. LUPU, *Artificial Intelligence and Intellectual Property*, in *World Patent Information*, 2018, LIII, A1; DICKENSON, MORGAN, and CLARK, *Creative machines: ownership of copyright in content created by artificial intelligence applications*, in *EIPR* 2017, VIII, 457; IHALAINEN, *op. cit.*, 724; RAMALHO, *Will Robots Rule*, *cit.*; SCHAFER, *Editorial: The future of IP law in an age of artificial intelligence*, in *Scripted* 2016, III, 283; BRIDY, *op. cit.*, 1; GINSBURG, *People Not Machines: Authorship and What It Means in the Berne Convention*, in *IIC* 2018, II, 131; GUADAMUZ, *op. cit.*, 169; KA-

Even though there is not generally accepted definition of artificial intelligence, it is useful to briefly refer to the most common attempts to better define the scope of this paper.

The scholar who coined the phrase (29) ‘Artificial Intelligence’ (30) defined AI as “the science and engineering of making intelligent machines, especially intelligent computer programs” (31). According to another commonly cited (32) definition, AI is the “simulation of human intelligence on a machine, so as to make the machine efficient to identify and use the right piece of ‘Knowledge’ at a given step of solving a problem” (33). Both definitions shift the problem to the understanding of what is ‘intelligence’, problem that had one of its most fortunate, albeit open to criticism (34), solutions in the Turing test, according to which a machine is intelligent if, playing the ‘imitation game’ a human could not distinguish between responses from a machine and a human (35).

RAGANIS and URBAN, *The Rise of the Robo Notice*, in *Communications of the ACM* 2015, IX, 28; LEVENDOWSKI, *How copyright law can fix artificial intelligence’s implicit bias problem*, in *Washington Law Review*, 2018, II, 579; PEREL and ELKIN-KOREN, *Accountability in Algorithmic Copyright Enforcement*, in *Stan. Tech. L. Rev.* 2016, 473; PEREL and ELKIN-KOREN, *Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement*, in *Fla. L. Rev.* 2017, 181; YANISKY-RAVID and MOORHEAD, *Generating Rembrandt: Artificial Intelligence, Accountability and Copyright - The Human-Like Workers Are Already Here - A New Model*, in *Michigan State Law Review (forthcoming)*, and NOTO LA DIEGA, *Machine Rules. Of Drones, Robots, and the Info-Capitalist Society*, in *Italian LJ.* 2016, II, 367 ff.. The literature on AI and intellectual property dates back to the Eighties, with one of the first works being the seminal BUTLER, *Can a computer be an author-copyright aspects of artificial intelligence*, in *Comm/Ent LS.* 1981, IV, 707 ff., and SAMUELSON, *Allocating Ownership Rights in Computer-Generated Works*, in *University of Pittsburgh Law Review* 1985, 1185 ff..

(29) Unlike the phrase, the concept and the relevant studies date back at least to the early Forties, and, in particular, to MCCULLOCH and PITTS, *A logical calculus of the ideas immanent in nervous activity*, in *Bulletin of Mathematical Biophysics* 1943, IV, 115, as noted inter alia by RUSSELL and NORVIG, *Artificial Intelligence: A Modern Approach*, 3rd ed, Prentice Hall, Upper Saddle River, 2009, 16.

(30) John McCarthy coined the term in 1955 when he set up a research group and then organised the Dartmouth Summer Research Project on Artificial Intelligence. See MCCARTHY, MINSKY, ROCHESTER, and SHANNON, *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence (31 August 1955)*, available at [http:// raysolomon-off.com/dartmouth/boxa/dart564props.pdf](http://raysolomon-off.com/dartmouth/boxa/dart564props.pdf). The role of McCarthy and the conference is universally recognised, but see the critical remarks of KLINE, *Cybernetics, Automata Studies*, and the Dartmouth Conference on Artificial Intelligence, in *IEEE Annals of the History of Computing* 2010, IV, 5 ff..

(31) MCCARTHY, *What is artificial intelligence?* (12 November 2007), available at <http://www-formal.stanford.edu/jmc/whatisai>.

(32) See, e.g., SHUKLA, TIWARI, and KALA, *Real Life Applications of Soft Computing*, CRC, Boca Raton, 2010, 9.

(33) KONAR, *Artificial Intelligence and Soft Computing. Behavioral and Cognitive Modeling of the Human Brain*, CRC, Boca Raton, 1999, 1.2.

(34) It would seem, however, that the Turing test is still the main test to assess machine intelligence. See LACURTS, *Criticisms of the Turing Test and why you should ignore (most of) them*, in *Official Blog of MIT’s Course: Philosophy and Theoretical Computer Science* (2011).

(35) TURING, *Computing Machinery and Intelligence*, in *Mind* 1950, CCXXXVI, 433, 441.

One of the main distinctions in the field is between a general, strong, or full AI on the one hand, and an applied, narrow, or weak AI on the other hand. Whereas Artificial General Intelligence (AGI) may finally replace humans because it seeks to “engineer human-level general intelligence-based theoretical models” (36), narrow AI “develops software to solve limited practical problems” (37), hence it is intrinsically aimed not to replace humans, but to improve their life, for instance in the fields of predictive analytics, driverless cars, care robots, speech recognition, and data mining. While to some AGI “starts looking like an attainable goal” (38), this paper will consider chiefly narrow AI for a twofold reason. First, because of the methodological option to focus on current scenarios, as opposed to future ones. Second, because of the importance of data mining, a typical example of narrow AI (39), in the context of the creation of databases and of the infringement of the relevant rights, as it will be explained below. For the purposes of this paper, AI is an umbrella term encompassing a number of technologies that make machines (hardware and software) increasingly autonomous (40) from the human beings (developers and users), the main of which are machine learning and deep learning. It may be useful to briefly examine said technologies.

Machine learning is a subset of AI that, existing at the intersection of statistics, AI, and computer science, aims at extracting knowledge from data sets (41). It enables automated learning, by having computers learning from input available to them, i.e. converting experience into expertise or knowledge (42). A key element machine learning is the expectation that “the accuracy of the computer algorithm will improve over time [...] as a result of feedback concerning previous accuracy” (43). As an example, Facebook uses machine learning algorithms inter alia to rank feeds, ads, and search results (44).

(36) GOERTZEL, *The path to more general artificial intelligence*, in *Journal of Experimental & Theoretical Artificial Intelligence* 2014, III, 343 ff..

(37) Ibidem.

(38) BARONI ET AL., *CommAI: Evaluating the first steps towards a useful general AI*, in *ICLR 2017 : 5th International Conference on Learning Representations - Workshop Track*, Toulon, 24-26 April 2017.

(39) Ibidem.

(40) On the relation between AI and autonomy see, for instance, MORRIS, SCHLENOFF, and SRINIVASAN, *A Remarkable Resurgence of Artificial Intelligence and Its Impact on Automation and Autonomy*, in *IEEE Transactions on Automation Science and Engineering*, 2017, II, 407 ff., that underline how “[t]he transition from automation to autonomy is one of the striking features of the rise of AI and [machine learning]” (ivi, 408).

(41) DAVIS, HOFFERT, and VANLANDINGHAM, *A taxonomy of artificial intelligence approaches for adaptive distributed real-time embedded systems*, in *2016 IEEE International Conference on Electro Information Technology*, 2016, 233.

(42) SHALEV-SHWARTZ and BEN-DAVID, *Understanding Machine Learning: From Theory to Algorithms*, Cambridge University Press, Cambridge, 2014, 1.

(43) *Machine learning*, in UPTON and COOK (eds), *A Dictionary of Statistics*, 3 ed., Oxford University Press, Oxford, ad vocem.

(44) On the hardware and software infrastructure that supports Facebook’s machine learning algorithms, see HAZELWOOD ET AL., *Applied Machine Learning at Facebook: A Data-center Infrastructure Perspective* (2017) <<https://research.fb.com/wp-content/uploads/2017/12/hpca-2018-facebook.pdf>> accessed 29 August 2018.

Deep learning algorithms, in turn, are “inspired by the structure and function of the brain called artificial neural networks” (45).

The computer that uses deep learning develops “complicated concepts by building them out of simpler ones” (46); the graph representing how the concepts are built is ‘deep’, “with many layers” (47). Machine learning, be it deep or not, is usually ‘supervised’, where the learning happens as a result of the training of the algorithm with labelled datasets (48).

For instance, Facebook image recognition (49) is supervised (50), which means that there is a human operator labelling a picture, say, of a cat as ‘cat’.

The importance of the involvement of humans in this type of machine learning made some emphatically, albeit not incorrectly, say that machine learning is a myth (51).

On the opposite end of the spectrum there is unsupervised learning, where the algorithm is trained with unlabelled data, therefore it learns in a way which is similar to the human one, i.e. by experiencing the world, rather than by being told the name of every object (52).

In recent years, finally, semi-supervised learning has emerged and gained popularity (53).

Less labour-intensive than the supervised learning and more accurate (54) than the unsupervised one, an example of it is Alexa, Amazon’s AI-powered virtual assistant, which learns how to decipher its users’ voice both with operators listening and

(45) DAVIS, HOFFERT, and VANLANDINGHAM, *op. cit.*, 233. However, as noted by MUELLER, REINHARDT, and STRICKLAND, *Neural Networks: An Introduction*, Springer Science & Business Media, Berlin-Heidelberg, 1995, 13, neural networks models can be deemed to be derived from research into the nature of the brain only in “a loose sense”. It has been noted that neural networks “while they indeed have something to do with brains, their study also makes contact with other branches of science, engineering and mathematics” (GURNEY, *An Introduction to Neural Networks*, CRC Press, Boca Raton, 2014, 1).

(46) GOODFELLOW, BENGIO, and COURVILLE, *Deep learning*, MIT Press, Cambridge (Ma), 2016, 2.

(47) *Ibidem*.

(48) LECUN, BENGIO, HINTON, Deep learning, in *Nature* 2015, VMMDLIII, 436 ff.

(49) It may be useful to keep in mind that the true challenge of AI is “solving the tasks that are easy for people to perform but hard for people to describe formally – problems that we solve intuitively, that feel automatic, like recognizing spoken words or faces in images” (GOODFELLOW, BENGIO, and COURVILLE, *op. cit.*, 1).

(50) See MAHAJAN ET AL., *Exploring the Limits of Weakly Supervised Pretraining*, in *ArXiv*, 2018, available at https://research.fb.com/wp-content/uploads/2018/05/exploring_the_limits_of_weakly_supervised_pretraining.pdf?

(51) BRADSHAW, *Self-driving cars prove to be labour-intensive for humans*, in *Financial Times*, 9 July 2017, who denounces the poor conditions of the relevant workers.

(52) LECUN, BENGIO, HINTON, *op. loc. ult. cit.*

(53) See, for instance, RODRIGUEZ, *Google Expander and the Emergence of Semi-Supervised Learning*, in *Medium*, 20 November 2016, available at <https://medium.com/@jrodthoughts/google-expander-and-the-emergence-of-semi-supervised-learning-1919592bfc49>, accessed 29 August 2018.

(54) Cf. LIANG and KLEIN, *Analyzing the Errors of Unsupervised Learning*, in *Proceedings of ACL-08*, Columbus, Cwmbrian, 2008, 879.

labelling, as well as autonomously interpreting unlabelled data (55).

The difference between the different forms of learning is relevant from our perspective because in most scenarios it is easier to assess whether there is infringement of a database if data are extracted or otherwise used using a supervised learning model; indeed, this will require the involvement of human operators to label the data. Conversely, it may prove more complicated to assess whether there has been infringement in instances where there are no humans involved. More on this later.

In the Eighties (56), it was predicted that the integration of AI and databases would have become critical “for the next generation of computing” (57). However, only recently (58) AI databases are becoming popular, thanks to the fact that AI promises to “simultaneously ingest, explore, analyze, and visualize fast-moving, complex data within milliseconds” (59). The ongoing relevance in the academic debate is confirmed by an annual international conference on AI and databases (60), and by the International Journal of Intelligent Information and Database Systems.

Four scenarios can illustrate how AI and databases are interwoven, from a lawyer’s perspective. The first two regard databases created by human beings using AI-produced data, either by creatively selecting and arranging the data, or by investing

(55) ANDERS, “*Alexa, Understand Me*”, in *MIT Technology Review*, 9 August 2017, available at <https://www.technologyreview.com/s/608571/alex-understand-me/>, accessed 29 August 2018.

(56) In the Eighties, AI scholars started looking into the relationship between AI and databases. See e.g. SCHOEN, SMITH, and BUCHANAN, *Design of knowledge-based systems with a knowledge-based assistant*, in *IEEE Transactions on Software Engineering*, 1988, XII, 1771; MYLOPOULOS and BRODIE (eds), *Readings in Artificial Intelligence and Databases*, Kaufmann, San Mateo, 1989; the Working Conference on the Role of Artificial Intelligence in Databases and Information Systems, Guangzhou, PR China, 4-8 July, 1988 (proceedings published as MEERSMAN (ed), *Artificial Intelligence in Databases and Information Systems (DS-3)*, Elsevier Science, New York, 1990). From then on, there has been a constant interest for the topic, as exemplified, in the Nineties, by PARSAYE and CHIGNELL, *Intelligent database tools & applications*, Wiley, 1993, BEYNON-DAVIES, *Expert Database Systems: A Gentle Introduction*, McGraw-hill, 1991, and MARIK and LAZANSKY, *Database and Expert Systems Applications*, Springer, 1993; in the 2000s, by LAST, KANDEL, and HORST, *Data Mining In Time Series Databases*, World Scientific, 2004, BERTINO, CATANIA, and ZARRI, *Intelligent Database Systems*, Addison-Wesley, 2001, and Ma, *Intelligent Databases*, Idea Group, 2007; and, finally, in the current decade, by LI, *Intelligent Multimedia Databases and Information Retrieval*, IGI Global, 2011, Brodie, Mylopoulos, and Schmidt, *On Conceptual Modelling*, Springer Science, Boca Raton, 2012, and BARBUCHA, NGUYEN, and BATUBARA (eds), *New Trends in Intelligent Information and Database Systems*, Springer, Cham-Heidelberg-New York-Dordrecht-London, 2015.

(57) BRODIE, *Future Intelligent Information Systems: AI and Database Technologies Working Together*, in MYLOPOULOS and BRODIE (eds), *op. cit.*, 623.

(58) The integration of AI and databases is recently gaining commercial success, however, it is not a new phenomenon. See, e.g. SCHOEN, SMITH, and BUCHANAN, *op. cit.*, 1771; SINGH and HIHNS, *Automating workflows for service order processing: integrating AI and database technologies*, in *IEEE Expert* 1994, V, 19 ff.

(59) A spokesperson of data management company Kinetica interviewed by MARVIN, *AI Databases: What they are and why your business should care*, in *PCMag UK*, 26 October 2017.

(60) The Asian Conference on Intelligent Information and Database Systems has reached its 11th edition, that will take place in Indonesia, 8-11 April 2019.

significantly in the obtainment, verification, or presentation thereof. At the centre of the third scenario, is an AI selecting or arranging data from various sources, whilst the last one regards an AI that obtains, verifies, or present said data. Artificial intelligence is relevant from a database perspective for a number of reasons, but mainly because, on the one hand, the underlying technologies require large datasets to use to train the algorithms (61), which begs the question whether AI-enabled data mining, scraping, and crawling are lawful and whether the output of these processes can be protected. On the other hand, AI produces big machine data and can create databases with the information derived therefrom. AI needs big data (62), produces them through data mining and other techniques (63), and has the ability to set up and manage proper databases (64), the legal regime of which does not seem entirely clear (65).

Finally, from a legal perspective it is immaterial whether a consensus can be reached on how to define AI and its technologies. Conversely, it is crucial to keep in mind two circumstances. First, most state-of-the-art AI applications require human involvement, for instance in the forms of labelling in supervised and semi-supervised learning. Second, intelligence and autonomy are still weak; whereas applied AI is indeed becoming ubiquitous, strong AI is still quite not here yet, though it may be on the horizon. While we sail towards it with a strong wind behind us, we had better focus on the legal issues in the technologies that are already here and constitute a non-negligible challenge to lawyers and lawmakers. A technological confusion between discrete concepts such as AI, big data, and the Internet of Things characterises the 2018 evaluation of the Database Directive and may contribute to explain why the Commission seems inclined to think that the *sui generis* right is not fit for machine data.

3. In the early Nineties, supported by the TRIPs (66), at the EU (then Europe-

(61) However, there are a number of difficulties deriving from the use of large datasets. Several solutions have been developed, see e.g. BELABBAS and WOLFE, *Spectral methods in machine learning and new strategies for very large datasets*, in *PNAS* 2009, II, 369 ff.; XING ET AL., *Petuum: A New Platform for Distributed Machine Learning on Big Data*, in *IEEE Transactions on Big Data* 2015, II, 49 ff., YE ET AL., *Building feedforward neural networks with random weights for large scale datasets*, in *Expert Systems With Applications* 2018, 233 ff.

(62) See, e.g., ASSEFI ET AL., *Big data machine learning using apache spark MLlib*, in 2017 *IEEE International Conference on Big Data*, Boston, 11-14 December 2017; HUANG and LIU, *Big data machine learning and graph analytics: Current state and future challenges*, in 2014 *IEEE International Conference on Big Data*, Washington DC, 27-30 October 2014.

(63) Cf. XUE and ZHANG, *Evolutionary feature manipulation in data mining/big data*, in *ACM SIGEVOlution*, 2017, I, 4 ff.

(64) See, e.g., VARDE, MANIRUZZAMAN, and SISSON, *QUENCH ML: A semantics-preserving markup language for knowledge representation in quenching*, in *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 2013, I, 65; RHEINECKER, *DB Networks Debuts AI-Based Agentless Database Activity Monitoring*, in *Wireless News*, 16 December 2016; MARVIN, *op. loc. ult. cit.*; SCHOEN, SMITH, and BUCHANAN, *op. loc. ult. cit.*, and SINGH and HIHNS, *op. loc. ult. cit.*

(65) Commission, *Evaluation*, cit.

(66) Agreement of Trade-Related Aspects of Intellectual Property Rights, Article 10(2).

an Community) level was felt that something needed to be done to bridge the gap between the flourishing US database industry and the floundering local one (67). This was the main economic justification of the creation of a much contested (68) sui generis right protecting investments of database makers, regardless of the originality of databases. Alongside, it was considered that the different national laws on databases could constitute a factor of fragmentation of the single market and, therefore the Directive harmonised, although only partly, the copyright laws applicable to the original databases (69). The third objective was the safeguard of “the balance of interests between database users and database makers” (70). The evaluations conducted by the Commission in 2005 (71) and 2018 (72) and a review of the relevant EU case law (73) confirms that the adoption and transposition of said instrument did not achieve any of the objectives above.

In terms of scope of protection, the Database Directive defines a database broadly, setting out three requirements (74): i. A collection of independent works, data or other materials; ii. A systematic or methodical arrangement; iii. Individual accessibility.

First, the materials need to be separate and not interact with each other (75). The materials must be separable from one another without the informative, literary, artistic, musical or other value of their contents being affected (76). The concept of

GHIDINI, *Rethinking Intellectual Property. Balancing Conflicts of Interest in the Constitutional Paradigm*, Edward Elgar, 2018, 241-242 points out that whereas the TRIPS justifies database copyright, there is no international foundation for the sui generis right.

(67) Database Directive, recitals 11 and 11; Commission, *Evaluation*, cit., 18.

(68) The most significant case in this regard is Judgment of the Court of Justice 15 January 2015. *Ryanair*, case C-30/14, in *Computer Law Review International* 2015, 83 with a comment by ELTESTE. *EU: Contractual Limitations for Database Use - Screen Scraping*. See also VOUSDEN, *Autonomy, comparison websites, and Ryanair*, in *IPQ* 2015, 386; CASTETS-RENNARD, *La liberté contractuelle et la réservation de l'information des bases de données non protégées devant la CJUE*, in *Droit de l'immatériel* 2015, 8; GUPTA and DEVAIAH, *Databases: The Database Directive "contracting out" bar: does it apply to unprotected databases?*, in *II-PLP* 2015, 669; ROSS, *"Not Getting into a Scrape": Dispute over "Screen Scrape" Data*, in *Computer and Telecommunications Law Review* 2015, 103; SYNODINOU, *Databases and screen scraping: lawful user's rights and contractual restrictions do not fly together*, in *EIPR* 2016, V, 312.

(69) Database Directive, recitals 2, 3, and 4.

(70) JIIP and Technopolis Group, *Final report*, cit., ii.

(71) DG Internal Market and Services, cit., 24.

(72) Commission, *Evaluation*, cit., passim.

(73) *Ryanair*, cit., 312.

(74) Database Directive, Article 1(1).

(75) PILA and TORREMANS, *European Intellectual Property Law*, Oxford University Press, Oxford, 2016, 510.

(76) Court of Justice 9 November 2004, *Fixtures Marketing v OPAP*, case C-444/02, paras. 29 and 33, in this *Journal*, 2005, 407 ff. with a comment by COGO. See also APLIN. *The ECJ elucidates the Database Right*, *IPQ* 2005, 204. For this reason, a recording or an audio-visual, cinematographic, literary or musical work as such does not fall within the scope” of the Database Directive (recital 17). This is because of the “semantic continuity” of such works, as noted by OTTOLIA, *Big data e innovazione computazionale*, in *I quaderni di AIDA*, n. 28, Giappichelli, Torino, 73. See also Court of Justice 26 October 2011, *Dufour*, case T-436/09,

independence has been interpreted broadly in *Verlag Esterbauer* (77) by considering immaterial the reduction of the autonomous value after the extraction and by pointing out that the “autonomous informative value of material which has been extracted from a collection must be assessed in the light of the value of the information not for a typical user of the collection concerned, but for each third party interested by the extracted material” (78). The reference to ‘materials’ encompasses both copyright and non-copyright works (79). This has a twofold consequence. First and foremost, the sui generis right could be used to protect AI works (e.g. a song written by an AI) even if they were not in themselves copyrightable, for instance because the originality conundrum were not untangled. The sui generis right might, therefore, become of unprecedented importance to grant some indirect protection to AI works, as a form of last recourse. Second, given that databases can include also copyright works, the AI, or its owners, will need to seek a licence from the author of the work they want to access or rely on some other legal basis (e.g. copyright exception for research purposes) (80). In turn, ‘data’ suggests that a database can include also personal data, should the relevant data protection requirements be met. One should keep in mind that the right of access as recognised by the EU General Data Protection Regulation (GDPR) (81) should not affect adversely the intellectual property of others, which, in turn, cannot be used to refuse to provide all information to the data subject (82). This has a threefold consequence when the AI includes personal data in the database: i. data subjects can access their data

in ECR. 2011, II, 7727, paras. 87, 102. On this case see LARCHÉ, *Accès aux documents*, in *Europe* 2011, XII, 14 ff.. In the most recent EU case about databases, the Court stated that “geographical information extracted from a topographic map by a third party so that that information may be used to produce and market another map retains, following its extraction, sufficient informative value to be classified as ‘independent materials’ of a ‘database’ within the meaning of that provision” (Court of Justice 29 October 2015, *Verlag Esterbauer*, case C-490/14, in *Dir. inf.*, 2016, 191, para. 30, with a comment by RESTA, *Sulla tutelabilità delle carte geografiche ai sensi della direttiva sulle banche di dati*). See also WIEBE, *Landkarten als Datenbanken: Der Informationswert von daten*, in *Gewerblicher Rechtsschutz und Urheberrecht PRAX* 2016, III, 49 ff..

(77) *Verlag Esterbauer*, cit..

(78) *Ivi*, para. 27.

(79) The Database Directive is “without prejudice to the freedom of authors to decide whether, or in what manner, they will allow their works to be included in a database” (recital 18).

(80) As stated in 5 March 2009. *Apis-Hristovich EOOD v Lakorda AD*, case C-545/07, in ECR, 2009, I, 1627, para. 71, the fact that the “materials contained in a legal information system are, by reason of their official nature, not eligible for copyright protection does not, as such, justify a collection consisting of those materials being refused classification as a ‘database’”. On this ruling see COGO, in this *Journal* 2009, 405 ff.. EICKEMEIER, *Relevanter Zeitpunkt und Umfang einer Datenentnahme*, in *Gewerblicher Rechtsschutz und Urheberrecht*, 2009, 578 and RAMBAUD, *Droit sui generis des bases de données: vers un équilibre?*, in *Droit de l’immatériel*, 2009, 6.

(81) Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation or GDPR) [2016] OJ L 119/1.

(82) GDPR, recital 63.

as long as this does not affect the rights on the database in an adverse way; ii. Even when the access adversely affects the rights on the AI database, this is no good justification for a refusal to provide all information; iii. The other data subject's rights are not conditional to the non-infringement of an intellectual property right, therefore for instance data portability requests could not be rejected.

The second requirement to fall under the definition of database is that an element internal to the database must organise the information according to methodical criteria e.g. chronological order (83). Conversely, it is not necessary for those materials "to have been physically stored in an organized manner" (84). In principle, the elements of a database are not presented "in any fixed, immutable order, but may be presented in a multitude of different combinations, using the technical and other means available" (85).

Third, the database must have a system for the retrieval of each of its constituent materials (86). This does not mean, however, that the rightholder cannot restrict the access to only part of the database (87). Therefore, for instance, Amazon's Alexa collection of independent data including the end-user's voice, organised chronologically, retrievable individually, might qualify as a database, even though the user can only access his or her data, and not the entire collection. Other users, for instance, the operators who train the algorithm and label the data, can access the whole of the database (88). The computer programs used in the making or operation of databases are outside of the scope (89). Therefore, the AI-powered software – e.g. Alexa – may be protected under the Software Directive (90) or as a computer-implemented inventions, if the relevant patentability requirements are met (91), while the relevant algorithm is more likely to be covered by a trade secret (92). It seems that the definition is broad enough to be applicable in the context of AI and related phenomena, therefore this author does not share the view of those who deem "necessary to clarify/address the problem of big data, IoT, sensor-

(83) The way in which the arrangement is achieved is immaterial, unless the arrangement is provided by an element outside the database. See Court of Justice 19 December 2013, *Innoweb*, case C-202/12, in *IIPLP* 2014, 458, with comment of BONADIO and ROVATI. *Use of dedicated meta-search engine infringes database right: the CJEU's stance in Innoweb v Wegener*. For instance, search engines do not make a database of the internet because they are external to the collection itself.

(84) Database Directive, recital 21.

(85) *Dufour*, cit., para. 107.

(86) *Fixtures Marketing v OPAP*, cit., paras. 31-32.

(87) One can infer this from Article 6(1) of the Database Directive.

(88) This database may fall outside the scope of the sui generis right, nevertheless, because it may be regarded as a "spin-off" database, issue on which more will be said later.

(89) Database Directive, Article 1(2).

(90) Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs [2009] OJ L 111/16.

(91) On the legal protection of software, see NOTO LA DIEGA, *Software patents and the Internet of Things in Europe, the United States, and India*, in *EIPR* 2017, III, 173 ff..

(92) NOTO LA DIEGA, *Against the dehumanisation of decision-making. Algorithmic decisions at the crossroads of intellectual property, data protection, and freedom of information*, in *JIPITEC* 2018, I, 3, 12 and passim.

generated data as part of the definition” (93).

4. Regardless of the *sui generis* right (94), the copyright protection covers only original databases, defined as “databases which, by reason of the selection or arrangement of their contents, constitute the author’s own intellectual creation shall be protected as such by copyright” (95). The protection regards the structure of the materials and, accordingly, the creative effort is relevant only if it regards the selection and arrangement of the materials, not their creation, as pointed in *Football Dataco v Yahoo* (96). The protection concerns only the database itself, not its contents that may or may not be covered by other intellectual property rights (97).

Originality plays a pivotal role when it comes to AI databases, compared to other copyright works, because it is the only criterion for the copyright protection of databases (98) and because originality, or the lack thereof, is arguably the main argument for the unfitness of copyright for protecting AI works. Understanding the originality of AI databases requires a preliminary understanding of originality of AI works. AI is increasingly creating works that, if made by humans, would probably qualify for copyright protection. One need only think that in August 2018 auction house Christie’s offered an AI work (99), thus signalling that buyers may consider such works as art (100). Another example of the change in perceptions when it comes to creativity is the fact that an AI art-generating algorithm has been acclaimed as “the biggest achievement of the year” (101). Many legal systems recognise that copyright works can be created with the help of software and hardware technologies. A famous example is the British regime of protection of computer-generated works (102), which are “generated in circumstances when there is no

(93) Legal annex, cit., 13, who admit, nonetheless, that this should happen in due course, because “it may be too early to legislate in this matter” (Ibidem). Assuming, for the sake of argument, that in the future the definition of databases will need a revision, it is crucial that this will happen in compliance with the principle of technological neutrality. See, inter alia, CRAIG, *Technological neutrality: Recalibrating copyright in the information age*, in *Theoretical Inquiries in Law* 2016, II, 60 ff.

(94) There can be double protection (copyright and *sui generis*) on the same right, as well only copyright or only *sui generis* right. See the Database Directive, Article 7(4).

(95) Database Directive, Article 3(1).

(96) Court of Justice 1 March 2012, *Football Dataco v Yahoo! UK*, case C-604/10, in *Diritto comunitario e degli scambi internazionali* 2012, 269, with comment by ADOBATI, *La Corte di giustizia interpreta la direttiva n. 96/9/CE sulla tutela giuridica delle banche dati*.

(97) Database Directive, Article 3(1).

(98) Database Directive, Article 3(1) and recital 15.

(99) *Is artificial intelligence set to become art’s next medium?*, in *Christie’s*, 20 August 2018, available at <https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx>.

(100) NUGENT, *The Painter Behind These Artworks Is an AI Program. Do They Still Count as Art?*, in *Time*, 20 August 2018.

(101) Chun, *It’s getting hard to tell if a painting was made by a computer or a human*, in *Artsy*, 21 September 2017. The research of the Art and Artificial Intelligence Laboratory (Rutgers University) is aimed at the use of AI to create proper art. Their publications are available at <https://sites.google.com/site/digihumanlab/publications>.

(102) Copyright, Designs, and Patents Act 1988, Section 9(3), 12(7), 79(2)(c), 81(2), 178, 214, 263. In Italy, whilst it is believed that machine-generated works are protected as long as they are distinct from the computer programme that generated them and are original,

human author of the work” (103) and are authored by the person who made the necessary arrangements to create the work (104). Though a rather advanced regime, the provisions on computer-generated works are not fit for strong AI and unsupervised machine learning, because they postulate the presence of a person to make the arrangements necessary for the creation of the work, who will be the author (105). Conversely, the regime of computer-generated works may fit some narrow AI applications and supervised machine learning (should the originality equation be solved). Given that in Italy there is no ad hoc provision on computer-generated works and given that the Italian copyright act does not limit the concept of author to humans (106), one may go as far as to argue that the Italian regime would be more suitable for an AI scenario than the British one (107), since it allows machines to be authors and hence owners of the works they produce (108). Needless to say that this would apply only if the originality conundrum were unravelled.

The fully dehumanised production of authorial and entrepreneurial works requires either interpretive stretches or, better, a legislative reform that clarifies the crucial points of authorship and ownership of AI works. In this field, the discussion about authorship and ownership should have an inversion. While normally one starts with authorship because ownership follows (109), with AI works, we need

the lawmaker has not taken a position on their authorship. See, e.g., GUTIERREZ, *La tutela del diritto di autore*, Milano, Giuffrè, 2008, 43 and ERCOLANI, *Computer-generated works*, in *Dir. aut.* 1998, 604.

(103) Copyright, Designs, and Patents Act 1988, Section 178.

(104) Copyright, Designs, and Patents Act 1988, Section 9(3).

(105) Copyright, Designs, and Patents Act 1988, Section 9(3). This section must be read in combination with the provision that defines computer-generated works as works generated “in circumstances such that there is *no human author* of the work” (Section 178, italics added). One can infer that this regime applies only to works where there are humans involved to make the necessary arrangements for the creation of the work. It would not apply, however, to the extremes of the spectrum: works where the human author makes the work with the aid of a machine (e.g. this author using Microsoft Word) and works created by AI with no human involved.

(106) Under Article 8(1) *legge* 22 April 1941 n. 633 (l.a.), the author is the entity (not necessarily the human being) who is indicated to be the author according to custom or, who is mentioned to be the author in the acting, execution, performance, or broadcasting of the work. Thus, it is important to read the contracts or the terms of service to understand who the author is.

(107) An *ad hoc* regime or a revision of the current general regime would be needed to accommodate the specific characteristics of the works generated by machines. For instance, machines do not die, therefore the usual duration system (seventy years after the death of the author) would be unsuitable. One could either provide *ad hoc* mechanisms (eg the British system, with the machine-generated works falling into the public domain after fifty years from the date they had been made), or the rise of machines could constitute a good opportunity to review the current system by, for instance, limiting the duration of copyright to the author’s lifetime.

(108) Given the current development of AI, it is still valid the theory of SAMUELSON, *AI-locating*, cit., 1185, whereby it is more convenient to consider the user as the original owner of the work (even though one should assess on a case-by-case basis the individual contribution of the user).

(109) For instance, in the UK, the Copyright, Designs, and Patents Act 1988 first defines the author as the person who creates the work (Section 9) and then set forth the principle that

first to decide whether or not to allow forms of ownership and, if so, how to allocate it. The question of whether or not to protect AI works is of political nature, because a public domain solution may be seen as inappropriate by some intellectual property holders, whereas allowing a generalised propertisation of AI works would be likely restrict access to knowledge and stifle creativity. At least five arguments can be brought in favour of a weak or no copyright protection for AI works. First, it is pivotal to prevent the monopolisation of culture. One need only think of the famous metaphor of the monkeys that, typing randomly for an infinite time, could write Shakespeare's complete works (110). Indeed, AI can create potentially copyright works faster than a group of monkeys working for the eternity, but unlike them the 'typing' would not be entirely random, in that the machine would learn and improve over time (111). Therefore, one day, every human being composing a song, writing a book, or making a database may risk receiving a letter of cease and desist from an AI (112). A comparative analysis provides further evidence of copyright's unfitness for AI works, with no known jurisdiction clearly allowing for copyright protection of such works and the US that changed their practice in 2017 to expressly exclude non-human authorship (113). Third, the traditional justifications

the author is the first owner (Section 1).

(110) It is usually accepted that the first formulation of the infinite monkey theorem is to be attributed to BOREL, *La mécanique statique et l'irréversibilité*, in *J. Phys. Theor. Appl.* 1913, I, 189 ff., who wanted to prove the point that the violation of laws of statistical mechanics is impossible, as opposed to improbable. However, the concept has been traced back to ARTISTOTLE'S *Metaphysics* by BORGES and HELFT, *La biblioteca total*, in *Sur* 1939, VIII, 13 ff.

(111) Many scholars are reflecting upon animal creations to understand copyright in the AI age. See, e.g., GUADAMUZ, *The monkey selfie: copyright lessons for originality in photographs and internet jurisdiction*, in

(I) internet Policy Review 2016, I, 1 ff., and ROSATI, *The Monkey Selfie case and the concept of authorship: an EU perspective*, in *JPLP* 2017, XII, 973 ff..

(112) This problem of 'copyright trolls', however, would have less negative consequences, if compared to the patent trolls (or non-practicing entities), since copyright not being a monopoly it allows for independent identical creations, principle which is expressed in the requirement for a claimant to prove a causal link in infringement proceedings. Equally, the importance to prove that the human work predates the AI one may lead to the proliferation of tamper-evident registration methods, which in turn may be enabled by blockchain technologies. On the emerging phenomenon of copyright trolls see DEBRYN, *Shedding light on copyright trolls: An analysis of mass copyright litigation in the age of statutory damages*, in *UCLA Ent. L. Rev.* 2012, 79 ff., GREENBERG, *Copyright Trolls and Presumptively Fair Uses*, in *University of Colorado Law Review* 2014, 53 ff., SAG and HASKELL, *Defense Against the Dark Arts of Copyright Trolling*, in *Iowa Law Review* 2018, 571 ff. On patent trolls, see LEMLEY and MELAMED, *Missing the forest for the trolls*, in *Columbia Law Review* 2013, VIII, 2117, CHIEN, *Startups and patent trolls*, in *Stan. Tech. L. Rev.* 2014, 461 ff., REILLY, *Patent "trolls" and claim construction*, in *Notre Dame L. Rev.* 2016, 1045 ff., COHEN, GURUN, and KOMINERS, *Patent Trolls: Evidence from Targeted Firms*, Harvard Business School Finance Working Paper No. 15-002, 8 June 2018. On blockchain-enabled copyright registration see NOTO LA DIEGA and STACEY, *Legal and regulatory issues in the blockchain: A focus on copyright law*, in RAGNEDDA and DESTAFANIS, *Blockchain and Web 3.0* (forthcoming).

(113) Under the U.S. Copyright Office, *Compendium of U.S. Copyright Office Practices § 101*, USPTO, Alexandria, 2017, 313.2, "[t]o qualify as a work of 'authorship' a work must be created by a human being".

for property rights in this field hardly apply to AI works, whose copyright protection could hardly be justified using the moral and economic foundations of copyright (114). In fact, AI does not sweat (115), nor it needs incentives to keep being creative (116). Another argument in favour of weak or no IP protection of AI works may be derived by the fact that the legislator showed that computer-generated works are less deserving of protection, if compared to fully human works. In particular, this can be seen in the reduction of the terms of protection to 50 years from the end of the year when the work was made and the exclusion of the moral rights (117). It is this paper's submission that with the further decrease in 'humanity' which one would witness with proper AI works, the protection should weaken accordingly.

The strongest argument in favour of AI works being receiving low or not copyright protection is the lack of originality, which troubled AI studies since their inception. In his pioneering *Computing Machinery and Intelligence*, Turing reformulated the question as to whether a machine can think by presenting the so-called imitation game. However, he had to respond to a number of critiques, at least one of which (118) are directly related to the problem at issue here. According to the so-called Lady Lovelace's objection (119), computers do not think because they are

(114) See the seminal studies of BREYER, *The uneasy case for copyright: A study of copyright in books, photocopies, and computer programs*, in *Harvard Law Review* 1970, 281; LANDES and POSNER, *An economic analysis of copyright law*, in *The Journal of Legal Studies* 1989, II, 325 ff.; JEHOAM, *Critical reflections on the economic importance of copyright*, in *IIC* 1989, IV, 485 ff..

(115) On the sweat of the brow doctrine applied to databases see GINSBURGM *No 'Sweat'? Copyright and Other Protection of Works of Information After Feist v. Rural Telephone*, in *Columbia Law Review* 1992, II, 338 ff.. This doctrine is rooted in a Lockean conception of property. According to LOCKE, *Two Treatises of Government*, II, *Of Civil Government*, London, 1690, 5, §25, now in *Two Treatises of Government*, II, *Of Civil Government*, London, 1821, 208, «God, who hath given the world to men in common, hath also given them reason to make use of it to the best advantage of life, and convenience... Though the earth, and all inferior creatures, be common to all men, yet every man has a property in his own person; this no body has any right to but himself. The labour of his body, and the work of his hands, we may say, are properly his. Whatsoever then he removes out of the state that nature hath provided, and left it in, he hath mixed his labour with, and joined to it something that is his own, and thereby makes it his property».

(116) Cf. ZIMMERMAN, *Authorship without Ownership: Reconsidering Incentives in the Digital Age*, in *DePaul L. Rev.* 2002-2003, 1121 ff.. The justifications centred on the incentives often overlooked that «the public interest is only truly served if copyright law provides appropriate incentives for all parties involved» (GEIGER, GRIFFITHS, HILTY, and SUTHERSANEN, *Declaration on a balanced interpretation of the "three-step test" in copyright law*, in *IIC* 2008, I, 707 ff..

(117) Copyright, Designs, and Patents Act 1988, Section 79 and 81.

(118) Another important argument points at the lack of consciousness, as beautifully expressed by JEFFERSON, *The Mind of Mechanical Man*, in *British Medical Journal*, 25 June 1949, 1(4616), 1105, 1110: "not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain". The argument, in the form of the so-called Chinese room, dates back to SEARLE, *Minds, brains, and programs*, in *Behavioral and brain sciences* 1980, III, 417 ff..

(119) According to LOVELACE, Notes by the translator, in TAYLOR (ed), *Scientific Mem-*

incapable of originality, which in turn depends mainly on the fact that they do not learn independently. Turing objects that the Countess in 1843 did not have access to the evidence that would have convinced her that machine could indeed think and, anyway, “[w]ho can be certain that ‘original work’ that he has done was not simply the growth of the seed planted in him by teaching, or the effect of following well-known general principles” (120). Turing’s counter-objection does not apply to originality as understood in copyright law. Indeed, copyright law is aware that we are *nanos gigantum humeris insidentes* (121) and that, therefore, the human creation builds on existing knowledge: what is required is only the link between the work and the author’s intellectual effort.

Grasping originality in copyright law, however, has been and still is not an easy task (122) also because its meaning must be construed in the specific meaning that it has when applied to databases (123). As correctly suggested (124), one should look separately at selective databases and databases protected by reason of the arrangement of the contents. The first ones, exemplified by a guide book (whose author has to select only certain information to communicate to the tourist), is protected because of the creative freedom in the selection of the contents (125). Data mining, crawling, and scraping will be analysed below, and it will be critically as-

oirs, 3, Taylor, London, 1843, 666, note G, the “Analytical Engine has no pretensions whatever to *originate* anything. It can do whatever we *know how to order it* to perform. It can *follow* analysis; but it has no power of *anticipating* any analytical relations or truths” (italics in the original). The Analytical Engine was Charles Babbage’s unbuilt invention considered as a forerunner of the electronic calculating computers. See FUEGI and FRANCIS, *Lovelace & Babbage and the Creation of the 1843 ‘Notes’*, in *IEEE Annals of the History of Computing* 2003, IV, 16 ff..

(120) TURING, *op. cit.*, 450. *Ibid.*, the Author reformulate the objection as ‘the machines cannot take us by surprise’, but this formulation is even less relevant from a copyright perspective.

(121) Notoriously, JOHN OF SALISBURY, *Metalogicon* (1159), trans. Doyle MacGarry, University of California Press, Oakland, 1955, 167, attributed the metaphor of the giants standing on the giants’ shoulders to Bernard DE CHARTRES.

(122) As a testament to the fact that the originality problem was never resolved see, e.g., ALGARDI, *Il plagio letterario e il carattere creativo dell’opera*, Milano, Giuffrè, 1966; OLSON, *Copyright originality*, in *Missouri Law Review* 1983, I, 29; PARCHOMOVSKY and STEIN, *Originality*, in *Virginia Law Review*, 2009, VI, 1505; ROSATI, *Originality in a Work, or a Work of Originality: The Effects of the Infopaq Decision*, in *J. Copyright Soc’y USA*, 2010, LVIII, 795; ROSATI, *Originality in EU Copyright: Full Harmonization through Case Law*, Elgar, Cheltenham-Northampton, 2013; LIU, *Of originality: originality in English copyright law: past and present*, in *EIPR*, 2014, VI, 376 ff.; MARGONI, *The harmonisation of EU copyright law: the originality standard*, in PERRY (ed), *Global Governance of Intellectual Property in the 21st Century*, Springer, Cham, 2016, 85 ff.; CASABURI, *Originalità, creatività, elaborazione creativa, citazione e plagio: Profili evolutivi*, in *Foro it.* 2017, XII, 3779; ROSATI, *Why originality in copyright is not and should not be a meaningless requirement*, in *JIPLP* 2018, VIII, 597 ff..

(123) The first scholars to stress this were L.C. UBERTAZZI, *Raccolte elettroniche di dati e diritto d’autore: prime riflessioni*, in Alpa (ed), *La tutela giuridica del software*, Giuffrè, Milano, 1984, 51 and 53 and DI CATALDO, *Banche-dati e diritto sui generis: la fattispecie costitutiva*, in *this Journal* 1997, 20 and 21.

(124) SPADA, *Banche dati e diritto d’autore (il “genere” del diritto d’autore sulle banche dati)*, in *this Journal* 1997, 5.

(125) Ivi, 6.

sessed if they can be seen as an expression of creative freedom. ‘Arrangement’, in turn, corresponds to coordination and arrangement as in Feist (126), i.e. linkage between the contents and in the order given to the them (127). If there is creative freedom in the way the contents are arranged or coordinated, then the database will be protected as a copyright work. AI, and in particular but not exclusively deep learning, can be used to find patterns that humans missed and, therefore, it could lead to a non-banal linkage between the data. However, would this be the author’s own intellectual creation? No, if the author is the developer or the owner of the AI, because it is not their creation. No, still, if the author is the AI because under current laws AI do not have personality.

In the domain of databases, the Directive covers “databases which, by reason of the selection or arrangement of their contents, constitute the author’s own intellectual creation shall be protected as such by copyright” (128). This goes beyond the concept of original as not copied, as in Paolo Spada’s intuition (129). Looking at the travaux préparatoires, one finds the express reference to the presence of a human author as a general principle, whereas legal persons as owners of databases are mere “deviations to that rules [...] merely tolerated” (130). The reference to a human author may be explained by a twofold circumstance. First, technologies already allowed the automation of databases (131). Second, databases are intrinsically less creative than traditional authorial works and clarifying ‘human’ may have acted as a delimitation of the scope of protection. While interesting, the travaux préparatoires will not play a key role in the interpretation of the Database Directive because said reference is not mirrored in any of the provisions of the directive (132).

The Court of Justice provided some guidance in *Football Dataco* by defining originality as the author’s creative ability to make free and creative choice in the selection or arrangement of the contents, thus stamping his or her personal touch on the database (133). In other words, the Court did not made the effort to provide a definition of originality that takes into account the intrinsic characteristics of the

(126) *Feist Publications, Inc. v. Rural Telephone Service Co.*, 499 U.S. 340 (1991), published also in *Dir. inf.* 1992, 111 with a comment by ZOPPINI, *Itinerari americani ed europei nella tutela delle compilazione: dagli annuari alle banche dati*. For a comparative analysis between the US and the EU, see MAZUMDER, *Database Law: Perspectives from India*, Springer, Singapore, 2016; DAVISON, *op. cit.*, 160 ff., and DERCLAYE, *The legal protection*, cit., 223 ff., DERCLAYE, *Intellectual property rights on information and market power - comparing European and American protection of databases*, in *IIC* 2007, III, 275 ff., and TABREZ and SOURAV, *Comparative Analysis of Copyright Protection of Databases: The Path to Follow*, in *Journal of Intellectual Property Rights* 2011, II.

(127) This view, first presented by SPADA, *op. loc. cit.*, has been recently developed by GUPTA, *Footprints of Feist in European Database Directive: A legal analysis of IP Law-making in Europe*, Springer, Singapore, 2017.

(128) Database Directive, Article 3.

(129) SPADA, *op. cit.*, 10.

(130) RAMALHO, *Will robot rule*, cit., 7.

(131) See, e.g., Schoen, Smith, and Buchanan, *op. cit.*, 1771.

(132) On the advantages and disadvantages in the use of such interpretive aids see, recently, PRISLAN, *Domestic explanatory documents and treaty interpretation*, in *International & Comparative Law Quarterly* 2017, IV, 923 ff..

(133) *Football Dataco*, cit., para 38.

subject matter. It merely referred to the leading cases in matters of originality (134). The ‘personal touch’ is a fuzzy concept which can be applied in a relatively easy way to works such as artistic works or traditional literary works such as books. Conversely, it is not easy to imagine how a database can exhibit its author’s personal touch. The matter may be different if the originality threshold were the traditional British standard of skill, labour, or judgement (135), because it can take a lot of skill, labour, and judgement to make a database. However, the Court of Justice expressly excluded the application of any criterion other than originality (136), and this means at least two things. First, it is immaterial whether or not the selection or arrangement of the contents includes “adding important significance” (137) thereto. Second, and more importantly, skill, labour, and judgement are irrelevant “if that labour and that skill do not express any originality in the selection or arrangement” (138). The shortcomings of the Court’s guidance are reflected in the fact that, even it has been claimed that originality standard for copyright protection of databases has been harmonised “across the EU Member States, there is evidence that the national courts are still uncertain on how to apply it” (139). Even though national variations should not be ignored, in principle the EU standard of originality leaves little scope for a copyright protection of databases and this applies all the more to AI databases. Indeed, the database is not original if its setting up “is dictated by technical considerations, rules or constraints which leave no room for creative freedom” (140). This sentence must be construed as meaning that databases whose setting up is dictated by technical considerations, rules or constraints can be original, if there is still room for the author’s creative freedom. In the current system, the epicentre is the human author, therefore only databases that are expression of a human’s intellectual creation can be protected (141). Arguably, most AI databases cannot be protected because they cannot meet the EU standard of originality (142).

(134) *Infopaq International*, cit., para. 45; Court of Justice 22 December 2010, *Bezpečnostní softwarová asociace*, case C-393/09, para. 50, in *ECR*, 2010, I, 13971; and Court of Justice, 1 December 2011, *Painer*, case C-145/10, paras. 89 and 92, in this *Journal*, 2012, 486. On the latter, with a focus on the originality conundrum, see also MICHAUX, *La notion d’originalité en droit d’auteur: une harmonisation communautaire en marche accélérée*, in *Revue de droit commercial belge* 2012, 599 ff.

(135) *The Newspaper Licensing Agency Ltd and Others v Meltwater Holding Bv and Others* CA [2011] EWCA Civ 890, [2012] RPC 1, [2012] Bus LR 53.

(136) *Football Dataco*, cit., paras. 40 and 53.

(137) *Football Dataco*, cit., para. 41.

(138) *Football Dataco*, cit., para. 42.

(139) JIIP and Technopolis Group, *Final report*, cit., vi.

(140) *Football Dataco*, cit., para 39.

(141) As noted by the Advocate General in *Painer*, cit., at para. 121, only human creations are protected, though these “can also include those for which the person employs a technical aid, such as a camera” (ibid.). However, in AI databases (and works more generally), AI is not a mere technical aid: it is the creator and/or maker itself. Conversely, if an automating technology is a mere technical aid, one could apply the regime of computer-generated works or, in its absence, extend the case law on photograph’s originality.

(142) BENTLY, SHERMAN, GANGJEE, and JOHNSON, *Intellectual Property Law*, Oxford University Press, Oxford, 2018, 118, who refer this assertion to all computer-generated works, that could however be protected by related rights or unfair competition law.

Indeed, in narrow AI scenarios, the AI is unlikely to take free choices, because the latter will in principle be entirely dictated by technical constraints. Conversely, databases created by strong AI are unlikely to be original, and hence protected, because the choice in the selection or arrangement of the contents may be free, but not creative, since the concept of creativity appears closely related to the involvement of a human author. The conclusion would be different in two scenarios. First, if AI developed to a point that we would grant legal personality and our resistance in accepting machine creativity would decrease. The Database Directive itself, at a closer look, allows Member States to recognise legal persons as owners of a database, and the argument may be put forward that this would apply to AI should its personality be recognised (143). Granting legal personality to AI, however, would not in itself unravel the originality conundrum. Second, if Brexit led to a return to the ‘skill, labour, or judgement’ standard – a return that “is both a challenge, and a customization opportunity” (144) – this would render British copyright law more AI-friendly and AI databases may be more easily protected by means of copyright. It should be noted, however, that this may be made difficult by the fact that the European standard of originality has made its way into the British statutory provision on databases (145). Therefore, a reversing case law could affect the general copyright concept of originality, but not the one applied to databases, for which there would be the need for a legislative reform. It should incidentally be noted, finally, that the fact that AI databases are in principle not covered by copyright is not likely to have significant practical consequences, since copyright does not cover most traditional databases and in light of its limited usefulness, in that it does not protect against independent creations (146).

If, contrary to this paper’s position, the option in favour of a copyright protection for AI databases should prevail, then the main ownership alternatives will be between the AI’s owner, the developer, the end-user, or some form of joint-ownership (147). The AI’s owner may be recognised as the database’s rightholder if the regime of employees’ works were applied, but said regime is designed for human employees. The developer is likely to have some rights on the code or the program more generally, but not on the final database. The end-user may have some rights, but only in the event they will contribute in a meaningful way in the selection or arrangement of the contents thus stamping their personal touch on the database. This excludes all unsupervised systems and arguably the labelling itself will not be sufficient to make out the originality requirement because the relevant process is dictated by technical constraints and is not expression of free and creative choices. Joint ownership, finally, would be a solution leading to uncertainty and not many foreseeable benefits. In light of the difficulties in the allocation of ownership and control, one may expect contracts playing a crucial role. For example, Amazon re-

(143) Database Directive, Article 4.

(144) RAMALHO and GOMEZ GARCIA, *op. cit.*, 670.

(145) Copyright Designs and Patents Act 1988, Section 3A(2). This is the only reference to the author’s own intellectual creation in the whole Act.

(146) ZOPPINI, *Commento alla Direttiva 96/9 dell’11 marzo 1996 sulla tutela giuridica delle banche dati*, in *Dir. inf.* 1996, 490, 491.

(147) DICKENSON, MORGAN, and CLARK, *op. loc. cit.*

tains ownership “[a]ll content included in or made available through any Amazon Service” (148), including inter alia data compilations and audio clips. In particular, these contents are protected “by Luxembourg and international copyright, authors’ rights and database right laws” (149). In addition, said laws cover also the “compilation of all content included in or made available through any Amazon Service” (150); these are “the exclusive property of Amazon” (151). The scope of this contractual provision is unclear, because the interactions between the end – user and Alexa may qualify as data collections and audio clips, but also as user – generated content, which is owned by the user and licensed to Amazon (152). Overall, the allocation of ownership would be one of the aspects that would need specific attention in the context of a reform allowing copyright protection for AI databases and works more generally.

Finally, the option in favour of a stronger protection for AI works may gain support only if one day AI would be granted legal personality either as a consequence of achieved conscience, or if the EU will proceed with its plans for an electronic personality (153). Whilst the public domain option seems to be preferred in the literature, it may be worth it to explore whether the sui generis right would enable AI owners to extract value from AI works, while avoiding the drawbacks of a strong proprietisation by copyright or contractual means.

5. The sui generis right is the main news in the Database Directive (154) and it has been emphatically labelled a legal monstrosity (155). It has been criticised for its anticompetitive effect, because originality and the idea-expression dichotomy do not limit the relevant monopoly and indeed the sui generis right “creates a monopoly in collections of facts and other non-copyrightable items that is difficult or sometimes even impossible to ‘invent around’” (156). Moreover, it has been argued that

(148) Amazon Conditions of Use & Sale, last updated on 24 May 2018, clause 3.

(149) Ibid.

(150) Ibid.

(151) Amazon Conditions of Use & Sale, clause 8.

(152) Ibid.

(153) On this proposal, see NOTO LA DIEGA, *The European strategy on robotics and artificial intelligence: Too much ethics, too little security*, in *European Cybersecurity Journal* 2017, II, 6 ff.. It has been argued that, should AI be granted legal personality, the rights and obligations of AI systems would be different to those of other subjects and they “could only have rights and obligations that are strictly defined by legislators” (CERKA, GRIGIENE, and SIRBIKYTE, *Is it possible to grant legal personality to artificial intelligence software systems?*, in *Computer Law & Security Review* 2017, V, 685).

(154) Databases were arguably protected by copyright even before the Database Directive. See, for instance, DERCLAYE, *The Legal Protection*, cit., 45 and BENDER, *Computer Law*, 2, 4, Bender, New York, 1978, 10.

(155) Reichman, *Legal Hybrids* between the patent and copyright paradigms, in *Colum. L. Rev.* 1994, 2432, 2496, quoted in GHIDINI, *Rethinking*, cit., 242.

(156) HUGENHOLTZ, *Abuse of Database Right Sole-source information banks under the EU Database Directive*, in LÉVÊQUE and SHELANSKI (eds), *Antitrust, patents and copyright: EU and US perspectives*, Edward Elgar, Cheltenham, 2005, 203. Cf., more generally and recently, FALCE, GHIDINI, OLIVIERI, *Informazione e Big data tra innovazione e concorrenza*, Giuffrè, 2018.

the sui generis right contradicts the principle that “does not allow exclusive rights on ‘presentation of information’” (157). The discomfort of part the literature may be explained by the fact that, unlike copyright, the sui generis right does no reward intellectual labour: here the right is a reward/incentive for those who make substantial investments in obtaining, verifying, or presenting the contents of the database (158). These critiques may be the basis for the 2004 rulings of the Court of Justice (159) that interpreted the scope of the Database Directive in an overly narrow way by stating that only the investment regarding the obtaining, verification, or presentation of existing independent materials count towards the sui generis protection; conversely, the resources used to create data are not covered. This, and not the law as stated in the Directive, constitute the main reason why the Database Directive may be unfit for AI databases. Such narrow interpretation explains why some experts (160), in view of the review of the Directive, have pointed out that it is not clear if the current definition of a database embraces AI and algorithm-generated data, and “whether they should benefit from protection under the sui generis right”

(157) GHIDINI, *Rethinking*, cit., 242, referring to the European Patent Convention, Article 52(2)(d). However, the counterargument could be put forward that said provision regards exclusively patents and is not expression of a general principle. For example, it has been argued that the exclusion can be explained by saying that the presentation of information is seen as too abstract and intellectual in nature to be patentable (BENTLY, SHERMAN, GANGJEE, and JOHNSON, *Intellectual Property Law*, Oxford, 2018, 474). This is a rule that applies to patents and arguably trade secrets (*Bailey & Williams v Levi Roots* [2011] EWHC 3098), but certainly not to other fields of intellectual property, such as copyright, where the presentation of information, though abstract and intellectual in nature, can be protected as literary work and dramatic work, should the relevant expression be original.

(158) Database Directive, Article 7(1). Even though the reference to ‘sui generis’ may suggest that this right is a right of a new kind, entirely different to copyright, this intuition would be inaccurate. Indeed, this right is not actually sui generis, if one takes into account that copyright itself has been genetically modified over the years, becoming an all-encompassing versatile and socioeconomically neutral form of protection that is available when the production of a good may be disincentived by the ease of copy made possible by new technologies and by free competition. This thesis was first formulated by SPADA, *op. cit.*, 17. Cf., more recently, HUGENHOLTZ, *Something Completely Different: Europe’s Sui Generis Database Right*, in FRANKEL and GERVAIS (eds), *The Internet and the Emerging Importance of New Forms of Intellectual Property*, Wolters Kluwer, Alphen aan den Rijn, 2016, 205.

(159) Court of Justice 9 November 2004, *British Horseracing Board v William Hill Organization*, case C-203/02, in ECR, 2004, I, 10415; Court of Justice 9 November 2004, *Fixtures Marketing v Svenska Spel*, case C-338/02, in ECR, 2004, I, 10497; *Fixtures Marketing v OPAP*, cit.; Court of Justice 9 November 2004, *Fixtures Marketing v Oy Veikkaus Ab*, case C-46/02, in ECR, 2004, I, 10365. On these rulings see RAGONESI, *Nota alle sentenze della Corte di Giustizia sulle banche dati*, in this *Journal* 2005, 575 ff., DERCLAYE, *The Court of Justice Interprets the Database sui generis Right for the First Time*, in *European Law Review* 2005, 420; MANAVELLO, *Prima decisione della Corte di Giustizia sulla protezione delle banche di dati*, in *Dir. Ind.* 2005, IV, 420; BERTANI, *Banche dati ed appropriazione delle informazioni*, in *Europa e diritto privato* 2006, 319; MASSON, *Creation of Database or Creation of Data: Crucial Choices in the Matter of Database Protection*, in *EIPR* 2006, 261; APLIN, *op. cit.*, 204.

(160) JIIP and TECHNOLIS GROUP, *Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases – Final report*, European Commission, Brussels, 2018, v.

(161). For the same reason, the stakeholders (162) lament that the Database Directive is outdated, because it does not take into account a number of technological developments, including “industry aggregation of data and big data; automatic data generation; and advanced computational methods for analysis, information and decision making” (163). Now, if the *sui generis* right is fit for AI databases, said right could play an unprecedentedly important role as the chief way to protect AI works, which cannot currently be copyrighted for lack of originality. Conversely, if AI databases fall outside the scope of the Database Directive, there is the risk of overprotection resulting from the combined effect of contracts and TPMs, which would not be limited by the exceptions laid out by the directive (164).

Before analysing the main hurdles to the *sui generis* protection of AI databases, let us briefly see what are the main features of this right. The rightholder (165) has the power to prevent extraction or re-utilisation of the whole or a substantial part of the contents of the database, to be evaluated quantitatively or qualitatively (166). In section 6 below, it will be assessed whether AI-enabled data mining can constitute unlawful extraction or re-utilisation of a substantial part of a database or if it can fall under one of the exceptions to copyright and *sui generis* right. Indeed, the lawful user can carry out the unsubstantial extraction and re-utilisation of the contents, and they can extract or re-utilise the substantial part thereof for private purposes (only if it is a non-electronic database), teaching, scientific research, and public security. The duration of protection is quite peculiar, because even if in theory the *sui generis* right expires in 15 years (167), practically it is easy to transform it into a perpetual right by changing the contents substantially, even if the change results from the accumulation of successive additions, deletions or alterations (168). Now, given that AI may render the change in a database’s contents easier, it may be argued that AI could easily trigger this provision, thus potentially giving rise to a perpetual protection of the database, covering also those contents which have not been changed (169). Should the *sui generis* be revitalised, as argued in this paper, the

(161) *Ibidem*.

(162) ‘Stakeholders’ here refers to the for-profit database makers, SMEs, startups, and the research community that gave evidence in the context of JIIP and TECHNOLPOLIS GROUP, *Final report*, cit..

(163) *Ivi*, iv.

(164) Ryanair, cit., 83. See BORGHI and KARAPAPA, *Contractual Restrictions on Lawful Use of Information: Sole-source Databases Protected by the Back Door?*, in *EIPR* 2015, VIII, 505, and CIANI, *Property rights model v Contractual approach: How protecting non-personal data in cyberspace?*, in *Dir. Com. Internaz.* 2017, IV, 831.

(165) In *Innoweb*, cit., para.36, the reference is expressly to “the person who has taken the initiative and assumed the risk of making a substantial investment” (emphasis added). The case refers also to *British Horseracing*, cit., paras. 32 and 46; *Fixtures Marketing*, cit., para. 35; and *Directmedia Publishing*, cit., para. 33.

(166) Database Directive, Article 7(1).

(167) Database Directive, Article 10(1)-(2) for *dies a quo* and regime.

(168) Database Directive, Article 10(3). See the Opinion of Advocate General Stix-Hackl delivered on 8 June 2004, paras. 139-155, in *British Horseracing*, cit., in ECR 2004, I, 10415.

(169) PILA and TORREMANS, *op. loc. cit.* are critical on the aspect of this perpetual protection that covers also old contents.

matter of the potentially eternal duration of the right should be addressed by the Court of Justice. Finally, the sui generis right applies only to EU databases (170), but agreements can be concluded by the Council upon proposal from the Commission (171). Such agreements may be necessary to protect British databases in the EU because of Brexit. The Database Directive has been transposed into UK law and in June 2018 the European Union (Withdrawal) Act has expressly saved all EU-derived legislation (172). However, much will depend on whether, at the end of the EU-UK negotiations, will be considered a third country or not (173).

From this paper's perspective, the main question is whether a substantial investment in AI can be seen as a qualifying investment in the obtaining, verifying, or presenting the contents of a database. The answer requires a closer look to the four cases decided by the Court of Justice at the end of 2004 (174). It will be shown that the decision is open to criticism and that the interpretation which scholars and the Commission give to the rulings, in the form of the spin-off theory, is inaccurate.

In British Horseracing, the dispute was over the use by William Hill of information taken from the claimant's database, for the purpose of organising betting on horse racing. The database at hand contained a large amount of information supplied by horse owners and other stakeholders of the racing industry (175). Part of the contents are the lists of horses running in the races, which are compiled thanks to a manned call centre; among other things, the operators must ascertain whether the horse can be authorised to run the race (176). The arrangement is partly automated, because there is a computer that allocates a saddle cloth number to each horse and determines the stall from which it will start (177). In terms of investment, it costed around £4 million a year to run this database (178). William Hill was a lawful user of the databases and rearranged a small part of its contents (179). The claimant, however, believed that even if the individual extracts might have been seen as non-substantial, the activity still infringed on their sui generis right because it

(170) By EU databases we mean databases whose makers or rightholders are nationals of a Member State, have their habitual residence in the EU, or if a company has the registered office, central administration or principal place of business in the EU, or, if only the registered office, also the genuine ongoing link to the economy of a Member State. See Database Directive, Article 11(1)-(2).

(171) Database Directive, Article 11(3).

(172) European Union (Withdrawal) Act 2018, Section 2. On Brexit and intellectual property much has been said, but no or little attention has been given to the protection of databases. See, e.g., TRAUB and DENNIS, *Brexit - What Could Happen to My IP Rights?*, in *Intellectual Property & Technology Law Journal* 2017, XI, 20, FARRAND, *Bold and newly Independent, or Isolated and Cast Adrift? The Implications of Brexit for Intellectual Property Law and Policy*, in *Journal of Common Market Studies*, 2017, VI, 1306; RAMALHO and GOMEZ GARCIA, *op. cit.*, 669.

(173) RAMALHO and GOMEZ GARCIA, *op. cit.*, 670.

(174) *British Horseracing v William Hill*, cit.; *Fixtures Marketing v Oy Veikkaus*, cit.; *Fixtures Marketing v Svenska Spel*, cit.; *Fixtures Marketing v OPAP*, cit..

(175) *British Horseracing v William Hill*, cit., para. 10.

(176) *Ivi*, para. 14.

(177) *Ibid.*

(178) *Ivi*, para. 15.

(179) *Ivi*, paras. 17 and 19.

would at least qualify as repeated and systematic extraction and/or re-utilization of insubstantial parts of the contents in conflict with the normal exploitation of the database (180). From an AI perspective, the main question at the centre of this preliminary ruling is what ‘investment in obtaining’ data means.

As to the ‘obtaining’ the data, the main passage of the court’s reasoning revolves around the purpose of the sui generis right as one could infer from some, perhaps overemphasised, recitals of the Database Directive (181). In particular, the purpose of the protection by the sui generis right – actually one of the purposes – 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” (182). What is presented as an inescapable conclusion, then, is that ‘investment in obtaining’ must be understood as referring to “the resources used to seek out existing independent materials and collect them in the database, and not to the resources used for the creation as such of independent materials” (183). If one dissects the recitals upon which the court based its decision, they could be used to argue opposite interpretations of the scope of the directive. First, databases are a vital tool in the development of the information market (184). This is true, but reducing the scope of the directive is unlikely to facilitate the growth of said market. Second, interestingly, the directive refers to the exponential growth in the amount of “information generated and processed” (185) and accordingly calls for significant investments in advanced information processing systems. The express reference to information that is ‘generated’ may be seen as encompassing both created and obtained data. Third, it is pointed out that these investments in information storage and processing systems need “a stable and uniform legal protection for databases” (186). It could be argued that sterilising the directive has not led to a stable protection for databases, nor stimulated the relevant industry. The only argument that might have had some merit was the one based on the recital (187) which excludes the compilation of several recordings of musical performances on a CD from the scope of the sui generis right *inter alia* because it does not represent a substantial enough investment. Whilst this might be interpreted as excluding created data, it could be also seen as meaning that (financially) trivial operations such as collections of recordings are in principle insufficient investments. It does not seem, anyway, that such a passage could justify a case law that effectively sterilises the Database

(180) Ivi, para. 20 and Database Directive, Articles 7(1) and 7(5).

(181) Database Directive, recitals 9, 10, 12, 19, and 49.

(182) *British Horseracing v William Hill*, cit., para. 31 (see also para. 30).

(183) Ivi, para. 31. The same goes for the concept of verification; indeed, according to the court, in the context of the assessment of the investment, the judge cannot take into account “[t]he resources used for verification during the stage of creation of data or other materials which are subsequently collected in a database” (ivi, para. 34). The same ratio decidendi can be found in the *Fixtures Marketing* cases at paras. 39 (*Svenska Spel*), 39-40 (*OPAP*), and paras 33-34, 41-42, 44-46, 49 (*Oy Veikkaus*).

(184) Database Directive, recital 9.

(185) Database Directive, recital 10.

(186) Database Directive, recital 12.

(187) Database Directive, recital 19.

Directive. Indeed, even though the vague wording of the directive allows for such interpretation, it cannot be said that it supports it (188). Alongside the misplaced emphasis on some recitals, there are some strong arguments against the conclusions of the Court of Justice in the 2004 rulings. First, one needs to keep in mind that one of the main goals of the directive was to stimulate investments thus bridging the gap between the EU and the US databases industries. This fact must be analysed in light of the empirical evidence clearly showing the majority of the investments made by the database owners regards data collection, rather than the setting up of the database itself (189). Second, and perhaps more importantly, it is difficult to draw a line between the concepts of creation and obtaining; this is confirmed by the fact that live football data are deemed to be ‘obtained’ in the UK (190), and ‘created’ in Germany (191). Even in the literature, there is no consensus on where to draw the line (192). Data mining itself is a good example of the untenability of dichotomy because mining leads to the discovery of correlations between existing data (193) and one could argue both ways, that this is creation of data or, as seems more reasonable, data obtaining. The untenability of the dichotomy creation/obtaining indirectly (194) recognised by the European Commission, when they observe that “in the context of automated data collection [...] it becomes increasingly difficult to distinguish between data creation and obtaining of data when there is systematic categorisation of data already by the data-collecting object” (195). Building a ratio decidendi on such weak foundations is not consistent with the principle of legal certainty and it does not reflect the versatility of AI (196), whose process of making a database cannot be compartmentalised in phases such as creating and obtaining the contents. The criticised dichotomy has profound consequences on the practical relevance of the Database Directive and on the users’ rights. Indeed, on the basis of the joint operation of the 2004 rulings restricting the scope of the directive and *Ryanair* (197) stating that there are no limits to the contractual autonomy when a database falls outside the scope of the directive, the directive has not been able to limit the proprietisation of data by contractual means, as will be further explained in section 7 below.

The sterilising effect of this case law has been worsened by the fact that the 2004 rulings have been read as if they introduced a spin-off theory, i.e. as if there

(188) Cf. Opinion of Advocate General Stix-Hackl delivered on 8 June 2004, *British Horseracing v William Hill*, case C-203/02, in ECR, 2005, I, 10515, paras. 41-46.

(189) Commission, *Evaluation*, cit., 36.

(190) *Football Dataco Ltd v Stan James Ltd* (No 2) [2013] EWCA Civ 27.

(191) Commission, *Evaluation*, cit., 25.

(192) See, for instance, the different proposals of DERCLAYE, *Databases Sui Generis Right: Should We Adopt the Spin-Off Theory?*, in *EIPR* 2004, 402, and OTTOLIA, *op. cit.*, 79.

(193) Draft Directive on Copyright in the Digital Single Market, Article 2(2).

(194) ‘Indirectly’ because the reference is to sensor-equipped, connected ‘Internet of Things’ objects, but the same applies to most AI scenarios.

(195) Commission, *Evaluation*, cit., 15.

(196) In terms, NAZEMI and PEDRAM, *Deploying Customized Data Representation and Approximate Computing in Machine Learning Applications*, in arXiv, 3 June 2018.

(197) *Ryanair*, cit., 386.

was no protection for the databases that constitute only a collateral activity of the company (198). It is this paper's contention (199) that the spin-off theory should not be interpreted as meaning that if making databases is not the main activity, then the Database Directive will not apply (200). This broad interpretation goes beyond what actually decided by the court and it affects negatively AI databases because it leads to the popular belief that "machine-generated databases [...] may largely be considered 'spin-off' databases" (201). In fact, the court clarified that the creation of a database can be "linked to the exercise of a principal activity in which the person creating the database is also the creator of the materials contained in the database" (202). One needs only to establish that (also) the obtaining, verification or presentation "required substantial investment [...] independent of the resources used to create those materials" (203). Even if, regrettably, this theory is predicated on the untenable creating/obtaining dichotomy, its narrow interpretation as proposed here leaves scope for a *sui generis* protection of AI databases. Thus, for instance, if one invested in two different AI applications, one for data mining, and the other for database making (obtaining, verifying, presenting data), the relevant AI database may be protected if the second investment is substantial, regardless of the database making activity being a primary or secondary activity (204).

AI can render big data held by companies usable by processing, structuring and optimising said data. It has been suggested (205) that since the 2004 rulings did not regard the verification and presentation of data, there would be some scope to recognise a *sui generis* right on AI databases, should the substantial investment concern such activities (206). This might ultimately "influence the legal regulation

(198) Commission, *Evaluation*, cit., 15, are arguably inaccurate when they state that "databases which are the by-products of the main activities of an economic undertaking ('spin-off' databases) are in principle not protected by the *sui generis* right, as they would not fulfil the 'substantial investment' threshold" (similarly *ivi*, 24). See also HUGENHOLTZ, *Abuse*, cit., 203, who sees the spin-off theory as a way to avoid the monopolisation of sole-source databases. However, the A. points out that this case law, even interpreted broadly, is not enough to counter the monopolisation effects of the sole-source databases not falling under the spin-off theory.

(199) See also OTTOLIA, *op. cit.*, 76, according to whom the 2004 rulings do not introduce a rigid dichotomy between companies that produce data and companies that arrange them; the fact that a database falls or not within the scope of the directive does not, indeed, depend on an abstract and general notion of data production.

(200) A more refined version of this interpretation can be found in HUGENHOLTZ, *Abuse*, cit., 203.

(201) Commission, *Evaluation*, cit., 35.

(202) British Horseracing, cit., para. 35.

(203) *Ibid.*

(204) Cf. DERCLAYE, *Databases Sui Generis Right*, cit., 402, in particular where the A. points out that "if the spin-off theory, as it seems to, refers to a broader meaning that any database which is a spin-off of another activity should not obtain protection, it goes too far".

(205) Commission, *Evaluation*, cit., 25.

(206) On AI for verification purposes see, e.g., AMRANI, LÚCIO, BIBAL, *ML + FV = \$|heartsuit \$? A Survey on the Application of Machine Learning to Formal Verification*, in arXiv 10 June 2018. As to AI for presentation purposes see inter alia SOTO, KIROS, KESELJ, MILIOS, *Machine learning meets visualization for extracting insights from text data*, in *AI Matters* 2016, II, 15 ff..

of the emerging data-driven business models building on ‘big data’ analytics of machine-generated, Internet of Things data” (207).

When it comes to understanding when an investment is substantial, the Court of Justice did not elaborate much, limiting itself to observe that the person who has taken the initiative and assumed the risk of making a substantial investment in terms of human, technical and/or financial resources in the setting up and operation of a database receives a return on his investment by protecting him against the unauthorised appropriation of the results of that investment” (208). The substantiality can be qualitative or quantitative, with the quantitative assessment referring “to quantifiable resources and the qualitative assessment to efforts which cannot be quantified, such as intellectual effort or energy” (209). In light of such a limited guidance, one needs to look at the national approaches. National rulings do not usually elaborate on the concept of substantiality of the investment. This has been explained in light of the fact that “in most cases, the investment is so enormous that there is no discussion as to whether the required level of substantiality is attained” (210). Who studied the matter more closely reached the following conclusions (211). First, ‘investment’ ought to be defined broadly, as including an effort in time, energy or money. Second, the threshold of substantiality should be set at a low level (212). This has been confirmed recently in the commissioned study in support of the evaluation of the Database Directive, where it has been pointed out that “national courts have been generous and granted protection for relatively low-level investments” (213). This conclusion has been upheld by the Commission that clarified that “[a]s a general rule, investment needs to be more than minimal, which

(207) *Ibid.*

(208) *Innoweb*, cit., para. 39, italics added, referring to *British Horseracing Board*, cit., paras. 32 and 46; *Fixtures Marketing*, cit., para. 35; and Court of Justice 9 October 2008, *Directmedia Publishing*, case C-304/07, para. 33, in ECR, 2008, I, 7565 and in this *Journal* 2009, 374 ff., with a comment by COGO. See also AREZZO, *L'estrazione non autorizzata del contenuto di una banca dati*, in *Dir. ind.* 2009, 192, and SAMMARCO, *Sull'ampiezza del diritto sui generis in relazione all'attività di estrazione del contenuto di una banca di dati non avente carattere creativo*, in *Dir. inf.* 2008, 780.

(209) Paragraphs 28 (Svenska Spel), 43 (OPAP), 38 (Veikkaus), referring to the Database Directive, recitals 7, 39, and 40th. For brevity reasons, this paper will not expand on the concepts of qualitative and quantitative substantiality of the investment, on which one can refer to DERCLAYE, *The legal protection*, cit., 91 ff..

(210) DERCLAYE, *Databases Sui Generis Right: What is a Substantial Investment? A Tentative Definition*, in *IIC* 2005, 2 ff..

(211) *Ivi*, 4.

(212) For instance, in France, see *Cour de Cassation* 23 March 2010, in *RIDA* 2010, 273 and, in Italy, Court of Rome 10 December 2009 as cited by Commission, *Evaluation*, cit., 27. Only in some instances, national courts have denied protection for lack of substantiality in the investment. See, e.g., in France, *Cour de Cassation* 19 June 2013, 12-18.623, *Réseau fleur-i v L'Agitateur floral*, as cited in JIIP and TECHNOLIS GROUP, *Final report*, cit., 7 and Trib. com. Paris 16 February 2001, *AMC Promotion v CD Publishers Construct Data Verlag GmbH*, referred to by CARON, *Liberté d'expression et liberté de la presse contre droit de propriété intellectuelle*, in *Communication Commerce Electronique* 2002, II, 25.

(213) JIIP and TECHNOLIS GROUP, *op. ult. cit.*, 8.

points towards a relatively low threshold” (214). It seems clear, therefore, that it is not the substantiality to bring AI databases out of the scope of the Database Directive, being the creating/obtaining dichotomy the actual problem. Accordingly, the proposal put forward by some database makers (215) to require merely a (non-substantial) investment would not, unlike what believed by the Commission “to widen the scope of protection, and thus potentially bringing the *sui generis* right fully into the domain of big data” (216).

The generous approach to substantiality and broad construction of investment must be kept in mind when asking oneself whether AI databases meet the relevant requirement. If the AI application is designed ad-hoc to create a certain database, than the investment in AI will be in principle sufficient to qualify for *sui generis* protection. The matter is more complex should the AI be able to create databases serially. In that event, one could argue that the requirement of the substantial investment would be met only for the first database, if the creation of the subsequent databases does not require an autonomous human, technical, or financial effort. For the first database it would be immaterial if there is human involvement because the investment might be substantial in reason of the financial cost of the AI application. Therefore, even databases whose creation were fully automated may be covered by the *sui generis* right (217). Conversely, if the AI application requires human intervention every time it makes a database (e.g. labelling in the event of supervised learning) or other efforts, one could argue that also the subsequent AI databases may be covered by the *sui generis* right.

In 2017 and 2018, the European Commission collected evidence (218) to decide whether or not to reform the Database Directive and it seems clear that the main concern was that the *sui generis* right did not fit automatically-collected or machine-generated data and the Internet of Things. In the public consultation, 42% of the respondents believed that the *sui generis* right was not appropriate for said data, that they claimed should be protected, though they did not explain why (219). The conviction that machine-generated databases are not covered by the *sui generis* is prevalent in the literature (220). This is countered by the opposite indication coming from the workshop organised in the context of the commission study in

(214) Commission, *Evaluation*, cit., 27.

(215) Ivi, 28.

(216) Ibid.

(217) This opinion does not seem shared by Commission, *Evaluation*, cit., 36, where it is submitted that if the data “not only requires automatic processing and formatting, but also manual processing and quality checks...there is case-law where relatively small investments triggered *sui generis* protection”. This interpretation is based on the criticised misconstruction of the spin-off theory, and of the creating-obtaining dichotomy.

(218) These included a public consultation, stakeholder meetings, a contracted study, which included an online survey, in-depth interviews, and a workshop (JIIP and TECHNOPO-LIS GROUP, *Final report*, cit.).

(219) Synopsis report on the responses to the public consultation activities on the evaluation of Directive 96/9/EC on the legal protection of databases, para. 1.2.5.

(220) LEISTNER, ‘Big Data and the EU Database Directive 96/9/EC: Current Law and Potential for Reform’ in LOHSSE, SCHULZE, and STAUDENMAYER (eds), *Trading Data in the Digital Economy: Legal Concepts and Tools*, Nomos, Baden-Baden, 2017, 25.

support of the evaluation of the Database Directive; indeed most “participants thought it unclear whether the sui generis right applied to machine-generated data” (221). More in favour of the fitness of the sui generis right for machine data a minority of the scholars (222) and the Bundesgerichtshof in *Autobahnmaut* (223). In that case, machine-generated toll data were held to be protected by the sui generis right because the highway company had invested substantial financial resources in the recording of pre-existing data on cars using the highway, as well as in the verification and presentation of the data by means of a computer program. This is of great importance for AI databases, not only because it shows the untenability of the creating-obtaining dichotomy, but also because an investment in some software enabling verification and presentation of the contents could be enough for the sui generis protection of AI databases.

Despite the mixed signals, the Commission’s contractors concluded that “the Database Directive does not apply to the databases generated with [...] artificial intelligence). In fact, the generation of these databases is closely interlinked with the creation of their content” (224). Even if one could accept this statement, there are at least two caveats. First, this preclusion does not stem from the Directive, but from the narrow way the Court of Justice interpreted it in order to limit perceived monopolisation risks. And indeed the Commission pointed out that “[t]he interpretation of the scope in the 2004 CJEU rulings [...] rules out concerns about the sui generis right playing an anti-competitive role” (225). Such interpretation is based on an excessive emphasis on some recitals of the directive, and on an untenable dichotomy between creating and obtaining data. Should it become clear that, as opposed to being a threat to free access to knowledge, the Database Directive can play a positive role by preventing contractual abuses, then one could expect a reversing case law that clarify the applicability of the sui generis right to AI databases. The solution would not be to abolish the sui generis right, as requested by those who observed that it did not stimulate investments (226), but to broaden it (227). Second, AI databases are not only those where the AI creates data; they include also those databases where the AI obtains, verifies, or presents the contents thereof. For the latter, the sui generis right, even in the narrow interpretation given by the Court of Justice, does apply to (some) AI databases. For instance, it may be argued that Amazon Echo’s information is a database despite this is not the main activity carried

(221) JIIP and TECHNOPSIS GROUP, *Final report*, cit., 25.

(222) Ivi, 28 criticises this interpretation of the EU case law and in particular of the *British Horseracing*, cit., and *Fixtures Marketing*, cit..

(223) *Autobahnmaut*, BGH I ZR 47/08 (25 March 2010).

(224) Ivi, ii. The same Authors, however, reformulate this position in a softer way by asserting that “[t]he Internet of Things, Artificial Intelligence, algorithm- and sensor-generated data, Big Data are all gaining increasing economic importance. It is nevertheless unclear [...] whether the current definition of a database embraces them, and, even more importantly, whether they should benefit from protection under the sui generis right” (ivi, 5).

(225) Ivi, 21, where such effect is also linked to the “prevalence of contracts” (ibid.). This rather obscure reference should not be interpreted as meaning that the prevalence of contracts in the database industry does not have anti-competitive effects.

(226) Ibid.

(227) This opinion is not isolate, see e.g. ivi, iii.

out by Amazon, if one can show that there has been a substantial investment in obtaining, verifying, and presenting the data. The information on the human contribution to this database and of other potential efforts is not in the public domain, therefore one cannot conclude in one sense or the other. However, even in the event of little or no human involvement, the substantiality threshold might still be met showing the financial cost of the development of the AI technology.

As said before, contrary to this paper's position, the dominant view would seem that the *sui generis* right is unlikely to cover AI databases and big machine data. To remedy this situation, the Commission proposed to introduce a data producer's right for non-personal or anonymised data (228). A data producer's right, which can be placed in the context of the debate on data property (229), would be "[a] right to use and authorise the use of non-personal data" (230) granted to the data producer, that is "the owner or long-term user (i.e. the lessee) of the device" (231). Thus, users would "utilise their data and thereby contribute to unlocking machine-generated data" (232).

The proposal's underpinnings might seem *prima facie* unobjectionable. Indeed, according to the Commission, since the *sui generis* right has a limited application in the context of Big Machine Data, and since the latter is fundamental in the data economy, it would follow that we need to introduce said new right. However, this double assumption is unproven and, conversely, this paper contributes to bring evidence that it is simply incorrect. Indeed, the *sui generis* right can cover AI databases and in data falling outside the directive can efficiently, if not overly, protected by means of contracts, TPMs, trade secrets, and unfair competition laws (233). In a context of such strong protection, it does not seem that there be need for further incentives in the form of new rights. Moreover, the access to such a form of protection may discourage the recourse to the *sui generis* right thus contributing to the sterilisation of the Database Directive (234). Overall, it can be said that the data producer's right would be the wrong solution to a made-up problem.

Saving the *sui generis* right, thus, would bring a threefold benefit. First and foremost, it would provide some form of protection to the AI works comprised in the database, AI works that would otherwise be in the public domain for lack of

(228) See Commission, *Free Flow of Data*, cit.

(229) See HOEREN, *Big data and the Ownership in Data: Recent Developments in Europe*, in *EIPR* 2014, XII, 751 ff.; ZECH, *A legal framework for a data economy in the European Digital Single Market: rights to use data*, in *JIPLP* 2016, VI, 460 ff.; BURNS, *Regulating machine data: less is more for global growth*, in *WIPO Magazine* 2017, VI; STALLA-BOURDILLON ET AL., *Building the European data economy. Position paper on the proposal for a new right in non-personal data*, available at http://ec.europa.eu/information_society/newsroom/image/document/2017-30/consultation_data_eco-knight_65284C58-BC45-BD3E-6F27AD94A35F71EC_46162.pdf; HUGENHOLTZ, *Against data Property: Unwelcome Guest in the House of IP*, in ÜLRICH, DRAHOS, and GHIDINI, *Kritika. Essays on Intellectual Property*, 3, Elgar, Cheltenham, 2018, 48; DREXL, *op. loc. cit.*

(230) Commission, *Free Flow of Data*, cit., 13.

(231) *Ibid.*

(232) *Ibid.*

(233) DREXL, *op. loc. ult. cit.*

(234) BURNS, *op. loc. cit.*

originality. This would allow the extraction of value from the AI works, whilst preventing their monopolisation. Second, a revamped sui generis right would prevent undue data proprietarisation by contractual means by reducing the negative effects of the Ryanair case and its enshrining of an unlimited contractual autonomy. Finally, by providing some form of protection to data, it would significantly weaken the case for a new data producer's right, which would no longer necessary for the European data economy.

6. When there is an investment in AI that is clearly used to obtain, verify, and present the contents of a database, the sui generis right is likely to apply. The main problem, as we have seen, is in the definition of 'obtaining', given that judge – made creation – obtaining dichotomy. A very good example of the untenability of the dichotomy is data mining. Data mining, whose growth is closely related to developments in AI technologies, identifies correlations between existing data (235), therefore while prima facie it may be seen as 'creating' data, arguably it 'obtains' them. The fact that one could argue both ways confirms that the dichotomy should be abandoned.

The importance of data mining has been recognised by the European Commission, that has accordingly provided an ad-hoc exception in the proposed EU reform of copyright. However, one needs to assess to what extent existing exceptions can constitute a good legal basis for data mining, in the form of a defence in proceedings for infringement of AI databases. Now, the pair 'infringement and AI databases' can be analysed from a twofold perspective. Either as infringement of the AI database or infringement carried out by the AI in making the database. The first angle will be only briefly analysed because there are no significant differences between the infringement of an AI database and of a traditional one (236). Conversely, the second scenario is of the utmost importance from this paper's perspective because if data mining is considered infringing per se, this would significantly hinder the potential of AI in impacting the database market.

The infringement of the database copyright is only partly regulated by the Database Directive; therefore, the relevant regime should be construed building on the general principles about copyright infringement, i.e. restricted act (237), causal link (238), and substantiality (239). Therefore, there is infringement when someone, without the AI database's owner's permission, reproduces, alters, distributes, or communicates the whole or a substantial part of the database (240), if the new da-

(235) Draft Copyright Directive in the Digital Single Market, Article 2(2).

(236) For a more thorough analysis of database infringement see DERCLAYE, *The legal protection*, cit., 100 ff., and JIIP and TECHNOLIS GROUP, *Legal annex*, cit., 27-43, 69-96.

(237) Database Directive, Article 5.

(238) If there is no proof of direct copying, the focus will be on proving access and that the similarities between the two databases are sufficiently numerous, close, or extensive to make it likely that the similarity is due to copying and not to coincidence. *Designers Guild v Russell Williams* [2001] FSR 113. Access should be easy to prove if the database can be used only by registered users or if some logging system is in place.

(239) *Infopaq*, cit..

(240) The exclusive rights of the owner of a database are the usual afforded by EU copy-

tabase was derived from the allegedly infringed one. The only aspect that may deserve more attention is substantiality. Indeed, this is a matter of quality rather than quantity, in the sense that “[t]he reproduction of a part which by itself has no originality will not normally be a substantial part of the copyright and therefore will not be protected” (241). Consequently, since originality has a limited scope in the area of databases, third parties will be allowed to more extensive copying of the contents – if compared to traditional authorial works – as long as they do not appropriate the original expression of the author’s ideas. Lastly, should the said requirements be made out, the lawful user could still invoke of the defences or exceptions laid out by the Database Directive (242), in particular if they can prove that the act performed was necessary to access the contents or is within their normal use, regardless of the owner’s authorisation (243). These are the only mandatory exception to the database copyright and cannot be waived contractually (244). In addition, Member States have the discretion to extend the general copyright exceptions, and to introduce exceptions for private copy (but only for non-electronic databases), public security, teaching or scientific research. The latter exception – which applies also to the sui generis right – deserves more attention because it could be used by AI databases’ owners who make their databases using data mining targeted at the contents of third parties’ databases. Such mining will be lawful only if its sole purpose is non-commercial scientific research and with due acknowledgement of the source (245). On the latter point, one should keep in mind that “mentioning the authors’ names and/or the sources, which may not always make sense for data analysis” (246). While this exception is positively less narrow than the one provided in the Draft Copyright Directive in the Digital Single Market, as seen below, the due acknowledgement, the fact that it can be waived contractually (247), and the fact that it is not mandatory may significantly reduce its practical relevance (248). In assessing the scope of this exception, moreover, one needs to keep in mind that all exceptions to the database rights are subjected to the so-called three step test and cannot, therefore, unreasonably prejudice the AI database’s owner’s legitimate inter-

right law. See Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society [2001] OJ L 167/10 (Infosoc Directive), Articles 2-4.

(241) *Ladbroke v William Hill* [1964] 1 W.L.R. 273, 293.

(242) Database Directive, Article 6(1).

(243) Database Directive, Article 6(2).

(244) Database Directive, Article 15.

(245) Database Directive, Article 6(2)(b).

(246) TRIAILLE, DE MEEÛS D’ARGENTEUIL, and DE FRANQUEN, *Study on the legal framework of text and data mining (TDM)*, Publications Office, Luxembourg, 2014, 116

(247) On the use of contracts to override copyright exceptions see, recently, ARONSSON-STORRIER, *Submission to Australian Department of Communication and the Arts, Copyright Modernisation Consultation: Contracting out of Copyright Exceptions*, in *SSRN*, 3 July 2018, available at <https://ssrn.com/abstract=3211946>.

(248) DUTILH, *op. cit.*, 545, points out that according to some stakeholders “[t]hese exceptions (education and science) should be mandatory. Because this exception is not mandatory, France, Italy and Ireland have not implemented it”. It must be said that Italy has a research exception for databases (see Article 64 *sexies*, para. 1, lett. a, l.a.).

est or conflict with the normal exploitation of the database (249). Coming to the remedies, the Database Directive mandated merely the introduction of remedies, leaving the Member States free to decide how to regulate them (250). Nearly ten years later, the Commission (251) clarified that the Enforcement Directive (252) applied to copyright, all rights related to copyright, and the sui generis rights. Therefore, the AI database owner will be able to react to the third parties' infringement availing themselves of corrective measures (253), injunctions (254), compensatory actions (255), or other remedies provided by the applicable national law (256).

While the remedies are the same, the rest of the infringement regime is different if we compare copyright and sui generis right. There is infringement of the latter in the event of unauthorised extraction or re-utilisation of the contents of a database, or of a substantial part thereof, evaluated qualitatively or quantitatively (257). A defendant extracts if they permanently or temporarily transfer the contents of a database to another medium by any means or in any form (258). In turn, they will have re-utilised the contents if they made them available to the public in any form, with the exclusion of public lending (259). Whereas the consultation of a database does not constitute in itself infringement, the transfer or making available of the contents following on-screen consultation does constitute a potential infringing act (260). From an AI perspective, it is also crucial to keep in mind that extraction and re-utilisation are construed broadly and encompass fully automated extraction and re-utilisation (261). 'Quantitative substantiality' refers to the volume of data extracted or re-utilised, "and must be assessed in relation to the volume of the contents of the whole of that database" (262). Evaluating substantiality qualitatively, in

(249) Database Directive, Article 6(3), regards the three-step test applied to the database copyright, Article 8(2) deals with the homologous provision in the sui generis domain. This may seem a two-step test, though the reference to the Berne Convention for the protection of Literary and Artistic Works is likely to mean that the list of exceptions be exhaustive (Berne Convention, Article 9(2)).

(250) Database Directive, Article 12.

(251) Statement by the Commission concerning Article 2 of Directive 2004/48/EC of the European Parliament and of the Council on the enforcement of intellectual property rights (2005/295/EC) [2005] OJ L 94/37.

(252) Directive 2004/48/EC of the European Parliament and of the Council of 29 April 2004 on the enforcement of intellectual property rights [2004] OJ L 157/45 (Enforcement Directive).

(253) Enforcement Directive, Article 10.

(254) Enforcement Directive, Article 11.

(255) Enforcement Directive, Article 13.

(256) Enforcement Directive, Article 16.

(257) Database Directive, Article 7(1).

(258) Database Directive, Article 7(2)(a).

(259) Database Directive, Article 7(2)(b), which includes distribution of copies, rental, and on-line or other forms of transmission.

(260) *Directmedia Publishing*, cit., paras. 34-36.

(261) *Innoweb*, cit., 28, and 33, referring to *British Horseracing*, cit., para. 51, and *Football Dataco*, cit., para. 20.

(262) *British Horseracing*, cit., para. 70.

turn, means looking at to the scale of the investment in the obtaining, verification or presentation, regardless of the volume (263). While these concepts seem broad enough to catch most infringing acts, it is important to keep in mind that the fact that the extraction or re-utilisation affects the value of the contents of a database is immaterial when assessing the infringement of the sui generis right (264). From this paper's perspective, it is important to say that, given the general, albeit open to criticism, trend to negate the sui generis protection for AI databases, one could foresee that only AI databases where a very meaningful investment has been proven will be accepted as being covered by said right. Therefore, if compared to traditional databases, it will be more likely that the infringement will stem for extraction or re-utilisation of a substantial part of the contents, assessed qualitatively.

The Court of Justice provided some guidance on whether 'extraction' also covered materials derived indirectly from the database, without having direct access to the database (265). Answering this question is crucial in light of the fact that after *Ryanair* it has become clear that contracts, coupled with TPMs, can offer the strongest form of protection to database owners, with the only exception of protection against third parties that are not bound by the contract. By the principle of privity, the use and dissemination of data only indirectly derived from a database is not unlawful if the latter falls outside the scope of the Database Directive, for instance because the substantial investment has been in the creation of data and not in their obtaining, verification, or presentation. On this point, the court stated that 'extraction' also covered materials that, although derived originally from a protected database, were derived only indirectly from that database. Indeed, a different solution would leave "the maker of the database without protection from unauthorised copying from a copy of the database" (266) and would "prejudice the investment of the maker of the database" (267). A strong indication seems to come by the fact that the principle of exhaustion applies in this field only in the sense of preventing control on the resale of the database, whilst the original owner can still control extraction and re-utilisation (268).

From an AI perspective, then, it is crucial to understand that provision (269), briefly recounted above, according to which there is a new database, and accordingly a new term of protection, if the event of "[a]ny substantial change, evaluated qualitatively or quantitatively, to the contents of a database, including any substantial change resulting from the accumulation of successive additions, deletions or alterations, which would result in the database being considered to be a substantial new investment, evaluated qualitatively or quantitatively" (270). Now, given that AI

(263) *Ivi*, para. 71, where the court adds that "[a] quantitatively negligible part of the contents of a database may in fact represent, in terms of obtaining, verification or presentation, significant human, technical or financial investment".

(264) Database Directive, recital 46 and *British Horseracing*, cit., para. 72.

(265) *British Horseracing*, cit., paras. 51 ff.

(266) *Ivi*, para. 52.

(267) *Ivi*, para. 53.

(268) Database Directive, Article 7(2)(b).

(269) Database Directive, Article 10(3).

(270) Database Directive, Article 10(3).

may render the change in a database's contents easier, it may be argued that AI could easily trigger this provision, thus potentially giving rise to a perpetual protection of the database, covering also those contents which have not been changed (271).

To conclude on the infringement of AI databases, this can derive from the repeated and systematic extraction or re-utilisation of insubstantial parts of the contents conflict with the normal exploitation of that database or unreasonably prejudice the legitimate interests of the maker (272). The purpose of this provision is to avoid the extraction or re-utilisation of a substantial part of the contents 'from the back-door' (273). Said repeated and systematic act is infringing only if the cumulative effect seriously prejudices the investment (274) by leading to the unauthorised "reconstitution of the database as a whole or, at the very least, of a substantial part of it" (275). One should keep in mind, moreover, that infringing activity does not have to be carried out necessarily with a view to create another database (276). Therefore, for instance, it would be illegal if through data mining insubstantial parts of an AI database are extracted thus leading to the reconstitution of the substantial part of the database.

Now, unless the requirements of provision on repeated extraction are made out, contractual clauses that restrict the lawful user's right to extract or re-utilise insubstantial parts of a database covered by a sui generis right are null and void (277). Therefore, for example, Amazon's contractual provision purporting to prevent any extraction or re-utilisation of any contents of their services is unenforceable (278). Equally, one can argue for the unenforceability of the clause that bans "data mining, robots, or similar data gathering and extraction tools (whether once or many times)" (279). It is this paper's submission that if it falls under the right to insubstantial extraction, or any binding exception, such clauses would be null and void. This user's right is particularly relevant from this paper's perspective, mostly because it covers all acts of repeated insubstantial extraction or re-utilisation are lawful, not only the non-commercial ones and no acknowledgement is required (280). This, coupled with the prevalence on contracts, make this right more appealing to data miners than the research exception, as long as they design the model in a way that target a number of different databases extracting or re-using repeatedly the contents of each without going beyond the threshold of substantiality. The main problem is likely to be how to code 'substantiality', since it is a rather vague con-

(271) PILA and TORREMANS, *op. loc. cit.* are critical on the aspect of this perpetual protection that covers also old contents.

(272) Database Directive, Article 7(5).

(273) Common Position (EC) No 20/95 adopted by the Council on 10 July 1995, OJ 1995 C 288, 14.

(274) *British Horseracing*, *cit.*, para. 86.

(275) *Ivi*, para. 87.

(276) *Ivi*, para. 86.

(277) Database Directive, Article 15.

(278) Amazon Conditions of Use and Sale, clause 3.

(279) *Ibid.*

(280) Database Directive, Article 8(1).

cept; therefore, further guidance from the Court of Justice would be much welcome (281).

The second, crucial, research question is whether data mining used to create a database is legal or not. The main obstacles seem contractual, e.g. if a website Terms of Service prevent mining (282), as well as stemming from the fact that at least part of the text and data mined may be covered by intellectual property rights (283).

In the context of the Draft Copyright Directive in the Digital Single Market, data mining has been defined as “any automated analytical technique aiming to analyse text and data in digital form in order to generate information such as patterns, trends and correlations” (284). To understand how data mining works, the relevant process can be divided in four stages (285). The first, not always present, step includes scraping (286) and crawling (287), used by the miner search for the relevant contents and retrieves the information, for instance by saving it on their own device or on the cloud. The second step is the creation of a target dataset. This may include the transformation of the contents for standardisation purposes, their enrichment with metadata, and the selection of only a part of the content deemed necessary for the analysis. The miner will extract said contents to a new dataset,

(281) On the problem of coding legal concepts, which has become urgent with the data protection by design obligation under the GDPR, see inter alia NOTO LA DIEGA, *Against the dehumanisation*, cit., 3 and passim.

(282) See, e.g., Facebook’s Automated Data Collection Terms, last updated on 15 April 2010, and Amazon’s Conditions of Use and Sale, clause 3. On this aspect, see the original work of STROWEL and DUCATO, *Limitations to text and data mining: Making the case for a right to ‘machine legibility’ of T&C and privacy policies* (forthcoming), in the context of smart disclosure systems and, in particular, of the automated analysis of contracts and privacy policy for enhancing the awareness of consumer.

(283) See the rigorous work of MONTAGNANI and AIME, *Il text and data mining e il diritto d’autore*, in this *Journal* 2017, 376 ff., to which one can refer to for a more in-depth analysis and bibliographic references.

(284) Draft Copyright Directive in the Digital Single Market, Article 2(2). The European Parliament, in the text adopted on 12 September 2018, improved the definition by pointing out that data mining is a technique which “analyses works and other subject matter in digital form in order to generate information, *including, but not limited to*, patterns, trends and correlations” (in italics the innovations). Arguably, this broader definition better reflects the heterogeneous phenomenology of data mining.

(285) CASPERS and GUIBAULT. *op. cit.*, 9.

(286) Scraping is a “process of making a semi-structured document from the Internet [...] and analyze the document to take certain data from the page to be used for other purposes” (KURNJAWATI and TRIAWAN, *Increased information retrieval capabilities on e-commerce websites using scraping techniques*, in *2017 International Conference on Sustainable Information Engineering and Technology*, IEEE, Piscataway, 2018, 226). In other words, it is a “technique for extracting data from the World Wide Web [...] and saving it to a file system or database for retrieval or subsequent analysis” (ibid.).

(287) Scraping and crawling overlap, but scraping can be done “manually by the user or automatically by the bot or the web crawler” (KURNJAWATI and TRIAWAN, *op. loc. cit.*). Moreover, a web scraper can be seen as the combination of a web crawler for crawling links and a data extractor from crawled links. MAHTO and SINGH, *A dive into Web Scraper world*, in *2016 3rd International Conference on Computing for Sustainable Global Development*, IEEE, Piscataway, 2016, 689, 690.

which they will use for the third step of the process, i.e. analysis, which is done with a mining software whose algorithm may be developed ad hoc by the miner but does not have to. The last stage is the publication of the findings of their mining. This can take many forms, from an academic paper to a proper database (288).

One could postulate that since the miners make free and creative choices in the selection of the contents, as seen in the third step of the process, they are stamping their personal touch on the database which would, therefore, be original and hence protected by copyright (289). However, the originality must regard the selection of pre-existing contents, not the creation of new ones, which is the essence of the data mining process.

Computer scientists complain that the question of the legality of data mining has not been answered with an adequate degree of certainty (290). In the US, some certainty may have been achieved with *eBay v Bidder's Edge* (291), where the judge innovatively applied the trespass of chattels to an online activity and, accordingly, granted an injunction to the notorious e-commerce portal to stop a bot from crawling its website for auction aggregation purposes. From this paper's perspective, it is important to underline that the court deemed relevant the fact that Bidder's Edge had to accept eBay's terms of service. This case, however, cannot be considered a reliable precedent because the dispute was settled out of court (292) and because subsequent cases casted some doubts on its validity (293). Other US cases uphold the legality (or at least non-illegality) of scraping because it would not breach anti-hacking laws, whilst others consider it as a copyright infringing behaviour (294).

In Europe, currently, there is no data mining exception (295), however the argument could be put forward that it could fall within the scope of existing exceptions (296). The main references are to exceptions for temporary reproduction (297), private copy (298), teaching and scientific research (299)

First, one should wonder whether data mining, in all its phases, could be re-

(288) *Ibid.*

(289) *Football Dataco*, cit., para 38.

(290) MAHTO and SINGH, *op. cit.*, 689, who deal with web scraping, but the reasoning applies to all data mining.

(291) 100 F.Supp.2d 1058 (N.D. Cal. 2000).

(292) MAHTO and SINGH, *op. loc. ult. cit.*

(293) *Intel Corp. v. Hamidi*, 30 Cal. 4th 1342 (2003), and *White Buffalo Ventures, LLC v. Univ. of Tex. at Austin*, 420 F.3d 366, 370-74 (5th. Cir. 2005).

(294) MAHTO and SINGH, *op. cit.*, 689-690. See inter alia *LinkedIn v Doe Defendants*, Case No. 5:16-cv-4463 (US District Court, Aug. 8, 2016), *LinkedIn v Robocog*, Case No. C14-00068 (WHA) (US District Court, January 6, 2014), and *Southwest Airlines Co. v. Boardfirst LLC*, Civ. Act. No. 3:06-CV-0891-B (N.D. Texas, September 12, 2007).

(295) Individual Member States, however, have introduced such an exception. See, e.g., in the UK, the Copyright, Designs and Patents Act 1988, Section 29A, as inserted by inserted by The Copyright and Rights in Performances (Research, Education, Libraries and Archives) Regulations 2014 (S.I. 2014/1372).

(296) The scope of the research exception and of the lawful user's rights under the Database Directive has been analysed above and will not be repeated here.

(297) Infosoc Directive, Article 5(1).

(298) Infosoc Directive, Article 5(2)(b).

(299) Infosoc Directive, Article 5(3)(a).

garded as mere transient or incidental reproduction solely aimed at a transmission in a network between third parties by an intermediary, or a lawful use of a work to be made, without independent economic significance. On the bright side, this exception is mandatory, therefore it is present in all Member States, though in some of them it can be waived contractually (300). The main problem here is that the exception is designed for caching and browsing activities (301), and the Court of Justice has interpreted the relevant requirements narrowly (302). As seen in the 4-step process described above, data mining can lead to copies which are not temporary or accessory (303). Moreover, while data mining is open ended, the exception applies only to specified purposes, such as the transmission in a network between third parties. More importantly, such temporary reproduction cannot “lead to a modification of that work” (304). This requirement is a gravestone to the possibility to use this exception in our context, because data mining leads to said modification most of the times, for instance if, during the analysis stage, the work is transformed so that the machine can process it (305).

Going on to the research exception, positively, unlike the homologous exception to the database rights, here acknowledgement is due only if it is possible, which is usually not the case when it comes to data mining. However, arguably, it is even less fit for data mining activities, because it is limited to the sole purpose of illustration for teaching or non-commercial scientific research (306). Moreover, Member States have discretion as to whether to implement it. The first hurdle is that the national implementations have interpreted it narrowly, as mainly encompassing the “personal use, study or (small scale) research” (307). In addition, its application to the online environment is limited (308). Similar issues characterise the private copy exception, which has been interpreted as the copies made in the family circle and it applies only to natural persons for purposes that are not even indirectly commercial,

(300) For more information on this see, e.g., KRETSCHMER, DERCLAYE, FAVALE, and WATT, *The relationship between copyright and contract law*, Intellectual Property Office Research Paper No. 2010/4, and, more recently, ARONSSON-STORRIER, *op. loc. ult. cit.*.

(301) InfoSoc Directive, recital 33.

(302) Court of Justice 5 June 2014, Public Relations Consultants Association v Newspaper Licensing Agency, case C-360/13, in this Journal, 2014, 1591, Infopaq, *cit.*, and Court of Justice 4 October 2011, Football Association Premier League v QC Leisure, case C-403/08, in ECR. 2011. I. 9083. See ALBERTI, *Radiodiffusione via satellite e clausole di esclusiva territoriale: note a margine di CEG*, 4 ottobre 2011, in *Europa e diritto privato* 2012, 256. BONADIO and SANTO, *Communication to the Public* in *FAPL v QC Leisure and Murphv v Media Protection Services (C-403/08 and C-429/08)*, in *EIPR* 2012, 277 ff., GIETZELT and UNGERER, *Die urheberrechtliche Dimension des Internetbrowsers und Caching*, in *Zeitschrift für Gemeinschaftsprivatrecht* 2014, 278 ff..

(303) Triaille, de Meeûs d'Argenteuil, and de Francquen, *op. loc. ult. cit.*

(304) *Infopaq*, *cit.*, para. 54.

(305) Triaille, de Meeûs d'Argenteuil, and de Francquen, *op. cit.*, 31-32.

(306) For a more in-depth analysis of this aspects see Guibault, Westkamp, and Rieber-Mohn, *Study on the implementation and effect in Member States' laws of Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society*, European Commission, Brussels, 2007, 49 ff.

(307) CASPERS and GUIBAULT, *op. cit.*, 34.

(308) GEIGER, GRIFFITHS, HILTY, and SUTHERSANEN, *op. loc. ult. cit.*

and require compensation (309). Moreover, it is optional, as evidenced by the fact that the UK no longer has a private copy exception (310). These exceptions would hardly apply to most data mining activities.

Overall, such exceptions may not confer effective rights to the consumers: they are narrow, usually optional, overridable by means of contracts or TPMs, inconsistently implemented, and, more importantly, do not cover most data mining phases and activities (311).

In light of the methodological option of leaving the policy considerations to a minimum while focusing on existing laws, a couple of words must be spent for the proposed new text and data mining exception as provided by the Draft Directive on Copyright in the Digital Single Market (312). The most positive and relevant innovation is that the exception is mandatory (313) and not overridable by means of contracts (314). The second aspect to be noted from this paper's perspective is that it applies expressly to general copyright, database copyright, sui generis right, and the proposed publishers' right (315). The main limitation is that only research organisations can avail themselves of this exception, only for scientific research purposes, only if they had lawful access to the works, and it does to cover the re-utilisation of the contents of a database (316). If this provision does not change, it

(309) VALENTI, *sub art. 12 l.a.*, in MARCHETTI e UBERTAZZI, *Commentario breve alle leggi su proprietà intellettuale e concorrenza*, Cedam, Padova, 2016, 1514.

(310) In the UK, the private copy exception has been quashed in 2015 and there is no indication of a possible reinstatement, despite the hopes expressed in NOTO LA DIEGA, *In light of the ends. Copyright hysteresis and private copy exception after the British Academy of Songwriters, Composers and Authors (BASCA) and others v Secretary of State for Business, Innovation and Skills case*, in *Diritto Mercato tecnologia*, 2015, II, 1 ff..

(311) STROWEL and DUCATO, *op. cit.*, 10 and HELBERGER ET AL., *Digital Content Contracts for Consumers*, in *J Consum Policy* 2013, I, 37.

(312) Draft Directive on Copyright in the Digital Single Market, Article 3.

(313) Draft Directive on Copyright in the Digital Single Market, Article 3(1).

(314) Draft Directive on Copyright in the Digital Single Market, Article 3(2).

(315) Draft Directive on Copyright in the Digital Single Market, Article 3(1). The proposal of a new publisher's right for the digital use of their press publications (Article 11) has given rise to a heated debate and criticism, which is very well represented by RICOLFI, XALABARDER, and VAN ECHOUUD, *Academics against Press Publishers' Right, Statement from 169 EU academics* (24 April 2018), available at <https://www.ivir.nl/academics-against-press-publishers-right/>; STALLA-BOURDILLON ET AL., *Open Letter to the European Commission – On the Importance of Preserving the Consistency and Integrity of the EU Acquis Relating to Content Monitoring within the Information Society*, in *SSRN*, 19 October 2016, available at <https://ssrn.com/abstract=2850483>; BENTLY and KRETSCHMER, *Strengthening the position of press publishers and authors and performers in the Copyright Directive*, European Parliament, Strasbourg, 2017, HILTY and MOSCON (eds), *Modernisation of the EU Copyright Rules - Position Statement of the Max Planck Institute for Innovation and Competition* (28 September 2017), available at <https://www.ip.mpg.de/de/projekte/details/modernisierung-des-eu-urheberrechts.html>; and SENFTLEBEN ET AL., *The Recommendation on Measures to Safeguard Fundamental Rights and the Open Internet in the Framework of the EU Copyright Reform*, in *EIPR* 2018, III, 149 ff. For a more optimistic take, see the rather isolated MELZI D'ERIL and VIGEVANI, *La buona informazione che garantisce diritti*, in *Il Sole 24 Ore*, 14 September 2018.

(316) Draft Directive on Copyright in the Digital Single Market, Article 3(1).

will be as unimportant – from a data mining perspective – as the old research exception (317), with the only advantage of not being expressly limited to non-commercial purposes (318), and to be binding, and mandatory. The European Parliament has further weakened this provision by suggesting that the not only the data mining but also the research to which the former is preparatory must be carried out by said research institutions. Arguably, this means that universities and research centres will not be able to rely on the text and data mining exception, should they decide to commercialise their data. Something similar applies to educational establishments and cultural heritage institutions. The Council and the Parliament have opened to the possibility that these subjects, and not only research organisms, can avail themselves of this exception, but only if an undertaking controlling them does not benefit from the exception (319). The latest text, then, mandates some form of TPMs or security measures by requiring that the reproductions and extractions made for text and data mining must be stored securely. This may be seen as part of the broader trend towards the technological enforcement of intellectual property and its ratio may be to prevent re-utilisation or further dissemination of the data, in a way that would unreasonably infringe the owner's rights. More interestingly, the reform attempts to prevent the abuse of TPMs by providing that the rightholders can put in place measures to ensure the databases' and network's security and integrity, but this cannot go beyond what is necessary to achieve said objective (320). Given that the over-protection of databases derives mainly from contracts and TPMs, the binding nature of this exception and its limitation on TPMs is likely to play a positive role in rebalancing the users-rightholders equilibrium.

Overall, the proposed text and data mining exception is too timid an attempt to address an activity, like data mining, that is becoming pervasive and from which the future of research may depend (321). Conversely, the right to extract or re-utilise insubstantial parts of a database covered by a *sui generis* right may be a better fit for data mining processes, because alongside being binding and not requiring acknowl-

(317) Interestingly, the version adopted by the European Parliament on 12 September 2018 allows Member States to “continue to provide text and data mining exceptions in accordance with point (a) of Article 5(3) of Directive 2001/29/EC” (Article 3(4)). It must be said, however, that this provision must be interpreted as meaning that the new exception does not replace the research exception, but it complements it. Conversely, it should not be interpreted as meaning that if the Member States have implemented a research exception they do not need to implement the text and data mining one.

(318) Since the research exception covers only the use “for the *sole purpose* of...scientific research” (Infosoc Directive, Article 5(3)(a)), the reference merely to research in the text and data mining exception is to be interpreted as meaning that the exception can be invoked in mixed projects where research is coupled with other objectives. Cf. TRIAILLE, DE MEEUS D'ARGENTEUIL, and DE FRANQUEN, *op. cit.*, 116.

(319) This is the meaning this author attributes to the new final sentence of Article 3(1) of the Draft Directive on Copyright in the Digital Single Market, as voted by the Parliament on 12 September 2018. However, the wording of the provision is so obscure that the proposed interpretation may be entirely wrong.

(320) Draft Directive on Copyright in the Digital Single Market, Article 3(3), left untouched as of 17 September 2018.

(321) Triaille, de Meeûs d'Argenteuil, and de Francquen, *op. cit.*, 114.

edgment, it is not limited to research purposes and to research organisms. Even the research exception to the *sui generis* act is better than the research exception to copyright, because it covers also those activities where ‘research’ is not the sole purposes. These are further reasons to revitalise the *sui generis* right (322). It must be kept in mind, finally, that even in the event that data mining were to be considered as an infringing act, the resulting AI database may still qualify for protection in at least (323) three scenarios. First, there has been a separate investment in the obtaining, creating, presenting of the contents. Second, data mining itself is seen as ‘obtaining’ data. Third, a reversing case law abandons the creating-obtaining dichotomy.

7. The existence of an ad-hoc instrument such as the Database Directive should not obfuscate the fact that databases were and are protected by a wide array of legal tools (324). In particular, this section will focus on TPMs, contracts, and unfair competition, whose joint operation might lead to an over-protection of databases. Empirical evidence and doctrinal studies (325) support the view that these legal regimes are more important than copyright and *sui generis* right in the age of Big Machine Data (Table n. 1). Other legal regimes that might apply and that will not be analysed in this paper include patents, trade marks, design rights, the protection of national treasures, laws on security, confidentiality, data protection, privacy, and access to public information (326). Even if confidentiality is mentioned by the Database Directive as one of the regimes that are not affected by the directive (327), it is this author’s conviction that trade secrets are not a suitable tool for the protection of databases, which are designed to be made available to the public (328).

(322) This is not to say, however, that changing the proposed text and data mining exception is not of the utmost importance. It is hoped that the trilogue will lead to amendments to the text and data mining exception that will get rid of the problematic aspects mentioned above. This is because data mining may lead to the infringement not only of the *sui generis* right, but also of other intellectual property rights on the contents of the database.

(323) Consistently with this paper’s premises, we are not considering copyright protection as a possible scenario, though in the future this might need some further thoughts, should the human element be considered necessary to make out the originality requirement, or should the requirement were altogether abandoned.

(324) For a comprehensive, though partly outdated, overview see DERCLAYE, *The Legal Protection of Databases*, Edward Elgar, Cheltenham, 2008, *passim*. More recently, and less comprehensively, see DREXL, *Designing Competitive Markets for Industrial Data – Between Propertisation and Access*, in *JIPITEC* 2017, VIII, 257. More generally, see JIIP and TECHNOPSIS GROUP, *Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases – Annex 1: In-depth analysis of the Database Directive, article by article*, European Commission, Brussels, 2018 (hereinafter ‘Legal annex’).

(325) DREXL, *op. cit.*, para 42 and *passim*, and Legal annex, *cit.*, 134.

(326) These regimes are not affected by the Database Directive, which provides their continued application under Article 13.

(327) Database Directive, Article 13.

(328) See DREXL, *op. cit.*, for the trade secrets protection. However, Drexl’s focus is on machine data and raw data more generally, not to proper databases. Indeed, a database is designed for its content to be available and retrievable, whereas trade secrets regard information that is not readily accessible and subject to steps to keep it secret. On the incompatibility

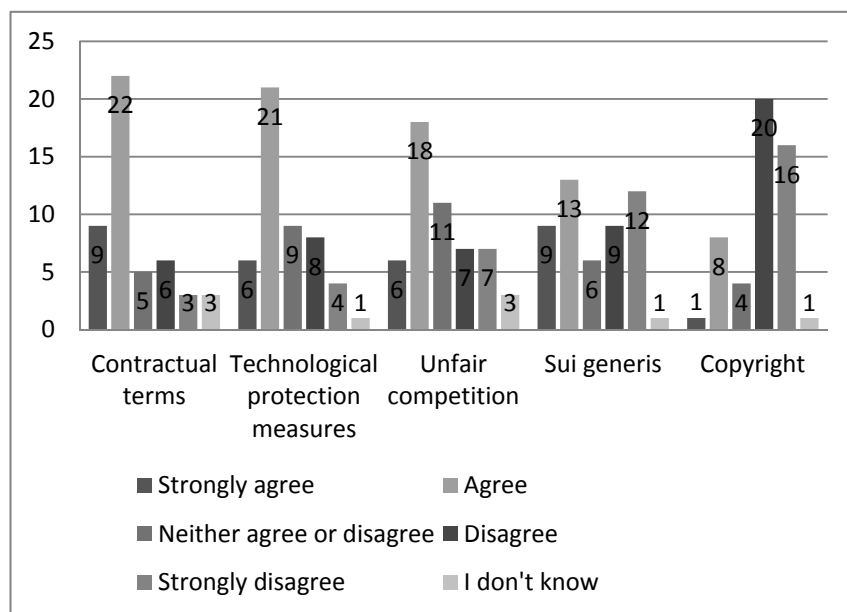


Table n. 1. Experts' answers to the question 'Do you consider that the databases that gather vast amount of data with the help of emerging/advanced technologies (e.g. sensor technologies) should benefit from the following means of protection against unauthorised use?' (329).

Starting off with the TPMs (330), it is useful to keep in mind that the Directive on the Information Society professedly leaves intact and in no way affects the legal protection of databases (331). Nonetheless, it adds two layers of protection regard-

between databases and trade secrets, see Trib. Bologna sez. Impresa civ 4 July 2017 n. 1371, unpublished. There is no awareness of such incompatibility in Chalton, *The legal protection of databases*, Thorogood, London, 2001, 86-87. On the definition of trade secrets, see 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, Article 2(1). At the time of writing, 16 Member States have transposed this directive into national legislation. See, e.g., in Italy, the *decreto legislativo* 11 May 2018 n. 63 modifying the *decreto legislativo* 10 February 2005 n. 30 (Industrial Property Code) and, in the UK, the Trade Secrets (Enforcement, etc.) Regulations 2018, S.I. 2018 n. 597. Bulgaria, Germany, Estonia, Greece, Spain, France, Cyprus, Latvia, Luxembourg, The Netherlands, Austria, Portugal, Romania, and Slovenia have failed to transpose it.

(329) Legal annex, cit., 134, but the table has been made by this author.

(330) TPMs deserve a separate focus because they can be used to restrict access also to content not covered by intellectual property rights. See, e.g., HUGENHOLTZ, *Abuse*, cit., 219 and DERCLAYE, *The legal protection*, cit., 191.

(331) Infosoc Directive, Article 1(2)(e), recital 20.

ing Digital Rights Management (DRMs) and TPMs (332). They are of great importance because data producers can use such measures to over-protect their databases “over and above the protection granted by the sui generis right” (333). First, Member States must provide for the protection against any person that knowingly (334) distributes, imports for distribution, broadcasts, communicates or makes available to the public any copyright material and databases covered by the sui generis right from which the DRM “has been removed or altered without authority” (335). Second, Member States must prevent the circumvention of any TPMs, “which the person concerned carries out in the knowledge, or with reasonable grounds to know, that he or she is pursuing that objective” (336).

AI is very relevant when it comes to TPMs and DRM for a twofold reason. On the one hand, the subjective element in the actions against the removal of DRM and against the circumvention of TPMs might render the application of this provision problematic when the circumvention measures are fully automated, especially with strong AI, because it would be difficult to show that the infringement had been carried out or facilitated knowingly. Moreover, it is important to keep in mind that, as stated in *Nintendo v PC Box* (337), in assessing the purpose of potentially circumventing devices, products or components, national courts may examine how often

(332) The international basis of such measures is the WIPO Copyright Treaty, Article 11. Many systems have similar provisions in place. See, e.g., the US Digital Millennium Copyright Act (17 U.S. Code § 1201) and the Swiss Bundesgesetz über das Urheberrecht und verwandte Schutzrechte, Article 39a. See, inter alia, BECHTOLD, *Digital Rights Management in the United States and Europe*, in *American Journal of Comparative Law*, 2004, 323 ff., GASSER, *Legal Frameworks and Technological Protection of Digital Content: Moving Forward Towards a Best Practice Model*, Berkman Center Research Publication No. 2006-04; KERR, *Digital Locks and the Automation of Virtue*, in GEIST (ed), *From "Radical Extremism" to "Balanced Copyright": Canadian Copyright and the Digital Agenda*, Irwin Law, Toronto, 2010, 247, and IWAHASHI, *How to Circumvent Technological Protection Measures Without Violating the DMCA: An Examination of Technological Protection Measures Under Current Legal Standards*, in *Berkeley Technology Law Journal*, 2011, 491 ff.

(333) DERCLAYE, *The legal protection*, cit., 191.

(334) The person must know or have reasonable reasons to know that this way they are inducing, enabling, facilitating or concealing an infringement of copyright or sui generis right. Infosoc Directive, Article 7(1).

(335) Infosoc Directive, Article 7(1)(b). Rights-management information is defined as “any information provided by rightholders which identifies the work or other subject-matter referred to in this Directive or covered by the sui generis right provided for in Chapter III of Directive 96/9/EC, the author or any other rightholder, or information about the terms and conditions of use of the work or other subject-matter, and any numbers or codes that represent such information” (Infosoc Directive, Article 7(2)).

(336) Infosoc Directive, Article 6(1), italics added. This directive explicitly defines TPMs as “any technology, device or component that, in the normal course of its operation, is designed to prevent or restrict acts, in respect of works or other subject-matter, which are not authorised by the rightholder of any copyright or *any right related to copyright as provided for by law or the sui generis right provided for in Chapter III of Directive 96/9/EC*” (Article 6(3), italics added).

(337) Court of Justice 25 January 2014, *Nintendo v PC Box*, case C-355/12, in EIPR, 2014, 335, with a comment by MINERO, Videogames, consoles and technological measures: the *Nintendo v PC Box* and 9Net Case. See also RENDAS, *Lex Specialis(sima): Videogames and Technological Protection Measures in EU Copyright Law*, ivi, 2015, 39.

they “are in fact used in disregard of copyright and how often they are used for purposes which do not infringe copyright” (338). While AI in general is a versatile tool that lends itself to manifold uses, it cannot be ruled out that specific AI applications might be deemed to be illegally circumventing TPMs. Moreover, the Infosoc Directive (339) rules out the manufacture, import, etc. of products that are promoted as circumventing, have only a limited commercially significant purpose or use other than to circumvent TPMs, or are primarily designed to enable or facilitate such circumvention. The more we move towards strong and general AI, intrinsically multi-purpose, the more unlikely will be the application of this regime to AI circumvention measures. Second, and more importantly, AI can be used not only as a circumvention measure, but also as a TPM in itself. AI is increasingly used to prevent infringement in a way that does not cope well with the copyright exceptions and limitations (340) and with the principle of exhaustion (341) leading to over-protection (342) and in a way that hinders cultural diversity (343). The problems with TPMs and exceptions are exacerbated by AI and other technologies such as blockchain, but they predate them (344). AI-enabled TPMs and DRMs are just the confirmation that the (binary) code may as well be seen as a form of law (345), but the law cannot be reduced to code (346), because it mirrors the complexity of its language and of the politics that produces it (347).

(338) Nintendo v PC Box, cit., para. 39.

(339) Infosoc Directive, Article 6(2).

(340) It is for the national court to determine whether a fair balance has been struck and whether other measures “could cause less interference with the activities of third parties or limitations to those activities, while still providing comparable protection of the rightholder’s rights” (ibidem). The problem of TPMs and exceptions will be partly addressed by the Draft Copyright Directive in the Digital Single Market, which makes certain exceptions mandatory and not contractually-overridable.

(341) On the overridability of the principle of exhaustion by means of TPMs see DERCLAYE, *The legal protection*, cit., 211.

(342) KARAGANIS and URBAN, *op. loc. ult. cit.* and PEREL and ELKIN-KOREN, *op. loc. ult. cit.*.

(343) JACQUES, GARSTKA, HVIID, and STREET, *Automated anti-piracy systems as copyright enforcement mechanism: a need to consider cultural diversity*, in *EIPR* 2018, IV, 218.

(344) See, e.g., BRAUN, *The interface between the protection of technological measures and the exercise of exceptions to copyright and related rights: comparing the situation in the United States and the European Community*, in *EIPR* 2003, XI, 496 ff.. On blockchain, TPMs, and DRMs, see DE FILIPPI, and HASSAN, *Blockchain Technology as a Regulatory Technology: From Code is Law to Law is Code*, in *First Monday* 2016, XII, that underlines how DRM can “prevent users from legitimately accessing or reproducing copies of a work, since the code rarely differentiates among the different types of users”.

(345) The obvious reference is to LESSIG, *Code v2*, Basic Books, New York, 2006. The idea of code as law has received a number of articulated objections, some of which very well articulated in O’HARA, *Smart Contracts – Dumb idea*, in *IEEE Internet Computing* 2017, II, 97 ff..

(346) For an example of this reductionist view see CASEY and NIBLETT, *The death of rules and standards*, in *Ind. LJ* 2016, 1401. An excellent critical response has been given by MICHAELS, *Abstract Innovation, Virtual Ideas, and Artificial Legal Thought*, in *Mar. J. Bus. & Tech. L.* (forthcoming) 22 ff. of the manuscript.

(347) More on these aspects in NOTO LA DIEGA, *Against the dehumanisation*, cit., 3 ff..

Interestingly, while all general provisions on TPMs and DRMs apply to databases *sic et simpliciter*, the obligation for Member States to take appropriate measures to ensure that rightholders make available to the beneficiary of an exception or limitation provided for in national law (348) the means of benefiting from that exception or limitation is only to be applied in the context of the Database Directive only “*mutatis mutandis*” (349). It is not entirely clear which part of the relevant regime needs tweaking. This author’s conjecture is that the normative fragment that does not apply to databases is the one that provides the non-application of the said regime on exceptions if the work is “made available to the public on agreed contractual terms in such a way that members of the public may access them from a place and at a time individually chosen by them” (350). Indeed, the application of this provision to databases would risk bringing most of them outside of its scope.

While it has been argued (351) that TPMs do not necessarily lead to over-protection, this cannot be said for sole-source databases (352) and other situations in which there is not or little competition in the market, which makes the digital lock up very likely (353). In highly competitive markets, a competitor could exploit the abuse of TPMs from a competitor to sell a more ‘open’ database, which may attract part of the market (354). There is no evidence on whether the market of AI databases be oligopolistic because given that there is no registration requirement one can hardly depict the relevant market with accuracy (355). However, the majority of the competition law literature is pointing out how AI decreases competition, to the point that it will “end competition as we know it” (356). Therefore, it is

(348) Infosoc Directive, Articles 5(2)(a)-(e), (3)(a), (3)(b) and (3)(e).

(349) Infosoc Directive, Article 6(4).

(350) Infosoc Directive, Article 6(4).

(351) DERCLAYE, *The legal protection*, cit., 193, which seems to limit this view to multiple source databases.

(352) A database is ‘sole-source’, if it is the only available source for certain data. Its antonym is ‘multiple-source databases’.

(353) STROWEL, *La protection des mesures techniques: une couche en trop? Quelques remarques à propos du texte de Kamiel Koelman*, in *Auteurs & Média*, 2001, 90; STROWEL *L’émergence d’un droit d’accès en droit d’auteur? Quelques réflexions sur le devenir du droit d’auteur*, in DOUTRELEPONT and DUBUISSON (eds), *Le droit d’auteur adapté à l’univers numérique*, Bruylant, Brussels, 2008, 61; ROTHCHILD, *Economic analysis of technological protection measures*, in *Oregon Law Review*, 2005, 489, 561, SAMUELSON, *Intellectual Property and the digital economy: why the anti-circumvention provisions need to be revisited*, in *Berkeley Technology Law Journal*, 1999, 519.

(354) SAMUELSON, *op. ult. cit.*, 566.

(355) Therefore, the most recent empirical analysis of the database industry in Europe relies on the GALE Directory of Databases. JIIP and TECHNOLIS GROUP, *Economic analysis*, cit., 3. However, it is meaningful that at the workshop organised to collect evidence in the context of said study, it has been noted that data “is in the hands of few players who will exercise undue monopoly power for information that should be open data” (*ibid.*, 27), and this opinion is shared by two of the surveyed experts.

(356) EZRACHI and STUCKE, *Virtual Competition: The Promise and Perils of the Algorithm-Driven Economy*, Harvard University Press, Cambridge (Ma), 2016, 218. Along the same lines, see MEHRA, *Antitrust and the Robo-Seller: Competition in the Time of Algorithms*, in *Minn. L. Rev.* 2016, 1323 ff., SURBLYTE, *Data-Driven Economy and Artificial Intelligence:*

not unreasonable to imagine that the invisible hand of the market will not fix the power abuses made possible by the digital locks, which leads us to talk about the competition law tools that can protect databases. Competition law plays a key role because in the information society most services are data-fuelled and, accordingly, the heart of the matter is becoming “control of data, i.e. information, as the source of ‘dominant positions’” (357).

The Database Directive itself recognises that (then not harmonised (358)) unfair-competition legislation covers databases and can prevent the extraction and re-utilisation of the contents thereof (359). It is moreover confirmed that the right to prohibit extraction and/or re-utilisation “relates not only to the manufacture of a parasitical competing product but also to any user who, through his acts, causes significant detriment, evaluated qualitatively or quantitatively, to the investment” (360). Equally, said directive expressly recognises that the sui generis right lends itself to anti-competitive abuses, in particular when the rightholder is a dominant enterprise; therefore, the Database Directive is “without prejudice to the application of Community or national competition rules” (361). Quite appositely, the Proposal (362) of the Directive included compulsory licensing provisions on fair and non-discriminatory terms (363), which have been struck out during the adoption procedure. What is left is the European Commission’s duty to submit a triennial report to verify “especially whether the application of [the sui generis] right has led to abuse of a dominant position or other interference with free competition” (364). Given that in the decades following the adoption of the directive, the report has been published only once (365), one may infer that addressing the anti-competitive conse-

Emerging Competition Law Issues, in *WuW* 2017, 120 ff., CALO, *Digital Market Manipulation*, in *George Washington Law Review* 2014, 995 ff.. Contra, PETIT, *Antitrust and Artificial Intelligence: A research Agenda*, in *Journal of European Competition Law & Practice* 2017, VI, 361 ff.

(357) GHIDINI, *Rethinking*, cit., 244.

(358) See now Directive 2005/29/EC of the European Parliament and of the Council of 11 May 2005 concerning unfair business-to-consumer commercial practices in the internal market and amending Council Directive 84/450/EEC, Directives 97/7/EC, 98/27/EC and 2002/65/EC of the European Parliament and of the Council and Regulation (EC) No 2006/2004 of the European Parliament and of the Council (Unfair Commercial Practices Directive’) [2005] OJ L 149/ 22.

(359) Database Directive, recital 6. On the database rights as barriers to entry see PEZZOLI, *Big Data e antitrust: un’occasione per tornare ad occuparci di struttura?*, in FALCE, GHIDINI, OLIVIERI, *op. cit.*, ch. 12, and FALCE, *Copyrights on data and competition policy in the digital single market strategy*, in *Italian Antitrust Review* 2018, 1, 32.

(360) Database Directive, recital 42, italics added.

(361) Database Directive, recital 47. See also Article 13

(362) Proposal for a Council Directive on the legal protection of databases (92/C 156 /03) COM(92) 24 final (Draft Database Directive).

(363) Article 8(2) of the Draft Database Directive provided that “[t]he right to extract and re-utilize the contents of a database shall also be licensed on fair and non-discriminatory terms if the database is made publicly available by a public body which is either established to assemble or disclose information pursuant to legislation, or is under a general duty to do so”. See also the original recitals 31-35.

(364) Database Directive, Article 16(3).

(365) DUTILH, *The implementation and application of Directive 96/9/EC on the legal*

quences of the database rights is not a top priority for the Commission (366).

When the Database Directive was drafted, the link between database rights and competition was considerably stronger. Indeed, the *sui generis* right was “very close to an unfair competition action for slavish imitation or parasitism” (367). This is in line with the fact that unfair competition laws act as an “incubator for new types of rights to emerge, which are later-on integrated into the corpus of traditional intellectual property laws or are transformed into rights *sui generis*” (368). The fact that the final version of the directive abandoned said approach in favour of a proprietary one does not mean, however, the competition law becomes irrelevant; indeed, it can be used both to protect the users against abuses of their database rights (369) and to protect the rightholders against unfair practices that do not qualify as infringement.

The only time the Court of Justice dealt with databases from a competition perspective was, incidentally, in *Compass-Datenbank* (370), which regards a case of alleged abuse of the *sui generis* right. In that context, the Court held that a public authority does not exercise an economic activity when it stores, in a database, data which businesses are statutorily obliged to report, allows interested parties to search for that data, and provides them with print-outs; thereof said public authority must not “be regarded, in the course of that activity, as an undertaking, within the mean-

protection of databases, European Commission, Brussels, 2001.

(366) In 2005, HUGENHOLTZ, *Abuse*, cit., 219 hypothesised that the report had “been postponed indefinitely”.

(367) JIIP and TECHNOLIS GROUP, *Legal annex*, cit., 120.

(368) KUR, *What to Protect, and How? Unfair Competition, Intellectual Property, or Protection Sui Generis*, in LEE, WESTKAMP, KUR, and OHLY (eds), *Intellectual property, unfair competition and publicity: convergences and development*, Elgar, Cheltenham, 2014, 11.

(369) On the abuse of the database rights, see FALCE, *Copyrights*, cit., 41, and HUGENHOLTZ, *Abuse*, cit., 219. On the abuse of copyright as a conceptualisation of the public interest defence, see the interesting BURRELL and COLEMAN, *Copyright Exceptions: The Digital Impact*, Cambridge University Press, Cambridge, 2005, 287. On the *abus de droit* applied to the intellectual property rights see Caron, *Abus de droit et droit d'auteur*. Litec. 1998, MOYSE, *Abus et propriété intellectuelle ou du bon usage des droits*, in SCASSA, GOUDREAU, SAGINUR, DOAGOO (eds), *Intellectual Property for the 21st Century: Multidisciplinary Perspectives on Intellectual Property Law*, Irwin Law, Toronto, 2014, 114, ZENKER, *Kartellrecht und Rechtsmissbrauch*, Nomos, Baden-Baden, 2018, and, incidentally, RICOLFI, *Diritto d'autore ed abuso di posizione dominante*, in *Dir. aut.* 2002, II, 215 who refers to MARZANO, *Diritto d'autore ed antitrust tra mercati concorrenziali e network economies*, in *Dir. aut.* 1998, 430 and PATTERSON, *Copyright Misuse and Modified Copyleft: New Solutions to the Challenges of Internet Standardization*, in *Michigan Law Review*, 2000, 1351. On the *abus de droit* in general see, among the recent works, FURGIUELE, *Abuso del diritto. Significato e valore di una tecnica argomentativa in diversi settori dell'ordinamento*, Edizioni Scientifiche Italiane, Naples, 2017, and CAPOTORTI, *L'abuso del diritto nell'ordinamento dell'Unione Europea*, Doctor thesis – University of Milan, 2017.

(370) Court of Justice 12 July 2012. *Compass-Datenbank*. case C-138/11. in Europe. 2012. X. 40 with a comment by IDOT. *Champ d'application matériel*. See LUNDQVIST, “Turning Government Data Into Gold”: *The Interface Between EU Competition Law and the Public Sector Information Directive*, in *International Review of Intellectual Property and Competition Law* 2013, 79 ff. and ROBIN, *Prérogative de puissance publique n'est pas activité économique en droit de la concurrence*, in *Revue Lamy de la Concurrence* 2013, 28 ff..

ing of Article 102 TFEU” (371) on the abuse of dominant position. Moreover, the prohibition of any use based on the sui generis right, or on the exercise of any other intellectual property right, still is not enough to qualify the activity as economic (372). Whilst this ruling is of some importance from a competition law perspective, the same cannot be said from a genuine database perspective (373). From the latter angle, this ruling has little relevance. Indeed, private undertakings relying on database rights are likely to fall under Article 102 TFEU, should all the requirements be met. An example is the *Nuovoimaie* case (374), where the Italian Competition Authority found that the company, a dominant operator in the copyright law related rights management and intermediation market, had abused its position by inter alia denying new entrants access to the general archive of works and artists. Consequently, the authority accepted *Nuovoimaie*'s commitments to license access to the database either freely to the database updated as of mid-March 2014, or access to the full database against an annual license fee of 4.5% of the total royalties managed (375). This is in line with the case law (376) according to which if intellectual property owners, in exercising their exclusive rights, threaten competition and consumer choice, can be held liable for abuse of dominant position (377).

Even those who argue against a sui generis right on machine-generated data

(371) *Compass-Datenbank*, cit., para. 53.

(372) *Ivi*, para. 51.

(373) In the lack of ad-hoc guidance, it is suggested to keep in mind that the main purpose of the sui generis right is to promote investment, not simply to reward it; therefore, any abusive conduct that would “clearly contravene the stated purpose of the database right would [...] run the risk of being disqualified as being anticompetitive” (HUGENHOLTZ, *Abuse*, cit., 218).

(374) Italian Competition Authority (*Autorità Garante della Concorrenza e del Mercato*) 22 March 2017 n. A489.

(375) A more recent example is the inquiry launched by the Italian Competition Authority against the main operators of the distribution and sale of electricity, whose strategy involved possible commercial exploitation of the database and billing data of standard offer customers. *Autorità Garante della Concorrenza e del Mercato* 4 May 2017 nn. A511, A512, A513, available at <http://www.agcm.it/stampa/comunicati/8752-istruttoria-nei-confronti-di-enel-a2a-e-acee-per-condotte-anticorcorrenziali-nel-mercato-della-vendita-di-energia-elettrica.html>, accessed 6 September 2018. The authority has not reached a decision yet.

(376) Court of Justice, 8 September 2016, *Lundbeck*, case T-472/13, in *Competition Law Insight*, 2016, X, 12, with a comment by COLE and ROBERT, *A landmark judgment: The General Court has affirmed the Lundbeck pay-for-delay decision*; Court of Justice 16 July 2015, *Huawei v ZTE*, case C-170/13, in *GRUR Int*, 2015, 781, with a comment by HILTY and SŁOWINSKI, *Standardessentielle Patente – Perspektiven außerhalb des Kartellrechts*; Final report of the Hearing Officer in Case COMP/38.636 – *Rambus* [2010] OJ C 30/15; Court of Justice 1 July 2010, *AstraZeneca*, case T-321/05, in *World Competition*, 2011, II, 245 with a comment by MAGGIOLINO and MONTAGNANI, *Astrazeneca's Abuse of IPR-Related Procedures: A Hypothesis of Anti-Trust Offence, Abuse of Rights, and IPR Misuse*; Court of Justice 17 September 2007, *Microsoft*, case T-201/04, in *ECR*, 2007, II, 3601; Commission Decision of 21 December 1988 relating to a proceeding under Article 86 of the EEC Treaty (IV/31.851 - *Magill*) [1989] OJ L 78/43.

(377) On the topic, in general, see RICOLFI, cit. and CAPUANO, *Abuso di posizione dominante e proprietà intellettuale nel diritto dell'Unione europea*, Editoriale Scientifica, Naples, 2012. Cf. KERBER, *Digital Markets, Data, and Privacy: Competition Law, Consumer Law and Data Protection*, in FALCE, GHIDINI, OLIVIERI, *op. cit.*, ch. 1.

accept that its anti-competitive consequences could be tempered by invoking anti-trust remedies and in particular the abuse of dominant position.³⁷⁸ Positively, competition authorities are starting to look into personal and non-personal data (379), and antitrust categories such as the essential facility doctrine may play a key role in avoiding proprietary excesses (380). In addition, since the Court of Justice has recognised that the *abus de droit* (abuse of right) is a general principle of EU Law (381), database owners abusive practices could be countered also in the absence of a dominant position (382).

From the opposite perspective – competition as a protection for the rightholders – the main reference is to the slavish imitation as a form of unfair competition, which remedies a risk of confusion separate³⁷⁹ (383) from the mere reproduction. There is slavish imitation if a database is insubstantially copied or otherwise exploited in their distinctive elements (384), when there is no technical reason to copy it or exploit it in order to profit from the research, development and marketing of the competitor (385). In *Football Dataco* (386), the Court of Justice had been asked whether the Database Directive precluded national rights in the nature of copyright in databases other than those provided for by directive itself. An adequate answer could have clarified the relationship between the database rights and unfair competition law. However, unfortunately, the Court interpreted the question narrowly as if referred merely to the question whether national laws can subject the database copyright to criteria others than originality (to which the answer is no) (387). Nonetheless, *Football Dataco* has been construed broadly as meaning that “it is not possible

(378) FALCE, *Copyrights*, cit., 41 argues against the *sui generis* protection for machine generated data, but with the caveat that “should the *sui generis* protection framework be retained, one could suggest to rely on the role antitrust norms can play as procompetitive antibodies.” According to GHIDINI, *Rethinking*, cit., 243, unfair competition would be a better form of protection for non-personal data, if compared to the *sui generis* right.

(379) GHIDINI, *Rethinking*, cit., 243.

(380) Ivi, 244. The idea of big data as essential facility has been developed by GRAEF, EU Competition Law, Data Protection and Online Platforms, Wolters Kluwer, 2016.

(381) *Halifax e a.*, C-255/02, EU:C:2006:121, 68; *SICES e a.*, C-155/13, EU:C:2014:145, 29, quoted in FALCE, *op. ult. cit.*, 41, fn 43.

(382) *Ibidem*.

(383) JIIP and TECHNOLIS GROUP, *Legal annex*, cit., 120.

(384) Therefore, if the database does not have elements that act as a badge of origin or if the new database contains such alterations that exclude consumers’ confusion, then the relevant remedy should not be available. See, e.g., Tribunale di Napoli 4 March 2014, in *Redazione Giuffrè*³ and Trib. Milan, sez. *Impresa* 24 December 2013, in Riv. dir. ind., 2014, II, 41 with a comment by CAPRA.

(385) This definition is adapted from the one provided by STECKLER, *Unfair trade practices under German law: “slavish imitation” of commercial and industrial activities*, in *EIPR* 1996, VII, 390. More recently see SUJECKI, *Slavish imitation and trade mark protection: a Dutch perspective*, in *EIPR* 2011, XII, 743 ff.. More generally, see LA VILLA, *Imitazione servile e forme di mercato*, Giuffrè, Milan, 1976; DI CATALDO, *L’imitazione servile*, Giuffrè, Milan, 1979, and ARCIDIACONO, *Parassitismo e imitazione servile non confusoria*, Giappichelli, Turin, 2017.

(386) *Football Dataco*, cit., para. 24(2).

(387) Ivi, paras. 47 and 52.

to cumulate slavish imitation or parasitism with the sui generis right” (388). This interpretation seems to go too far, because the paragraph of the ruling it refers to simply states that the Database Directive harmonised the criteria for copyright protection “as is apparent from recital 60” (389), which points out how said criteria are harmonised without affecting the term of protection. This interpretation, alongside going against the clear meaning of *Football Dataco*, does not take into account that the Database Directive is expressly “are without prejudice to the application of Community or national competition rules” (390). A stronger argument in favour of the non-cumulation of sui generis right and parasitism is that the Directive considered that existing legislation protecting databases, with different attributes (391), negatively affects the functioning of the internal market (392). However, apart from the fact that the directive expressly states the continued application of competition law (393), it is clear that the main reference when talking about obstacles to the free movement of databases is copyright (394), and it is clearly provided that different legislations “not adversely affecting the functioning of the internal market or the development of an information market within the Community need not be removed or prevented from arising” (395). On top of that, the differences between the unfair commercial practices laws have been reduced (396) after the adoption of the Unfair Commercial Practices Directive (397). Therefore, the cumulation of parasitism and sui generis right cannot be ruled out. And indeed, if one looks at the national legal systems, many Member States “still cumulate slavish imitation with the sui generis right and/or copyright” (398), despite most of the relevant literature being against it. A good example is France (399), where notwithstanding the Cour de Cassation not upholding the sui generis-parasitism overlap (400), French first instance courts are split on the issue (401) and recent Cour de Cassation decisions allow the over-

(388) JIIP and TECHNOLIS GROUP, *Legal annex*, cit., 120, interpreting *Football Dataco*, cit., para. 49.

(389) *Football Dataco*, cit., para. 49.

(390) Database Directive, recital 47. See also Article 13.

(391) Database Directive, recital 1.

(392) Database Directive, recital 2.

(393) Database Directive, Article 13.

(394) *Football Dataco*, cit., para. 48, Database Directive, recital 4.

(395) Database Directive, recital 2.

(396) Reduced and not eliminated because the directive applies only to business-to-consumer relationships, and also because several aspects are left unharmonised.

(397) This directive has been transposed in all Member States. In Italy, for instance, see *decreto legislativo* 6 September 2005 n. 206 (*Codice del consumo*), Articles 18-27 quater.

(398) *Legal annex*, cit., 121.

(399) Similarly, in Spain, *Tribunal Supremo* 30 January 2008 n. 14, in ORTEGA DOMÉNECH, *El derecho de autor en la Jurisprudencia del Tribunal Supremo*, Reus, Madrid, 2013, 27, in Germany BGH Bundergerichtshof 6 May 1999, I ZR 199/96, and, in Italy, for instance, Trib. Milan sez. *impresa* 1 August 2016, in this *Journal* 2017, 1815; Corte app. Bologna 10 February 2017 n. 356, *Giurisprudenza delle Imprese*.

(400) Court of Cassation 12 November 2015 n. 14-14501, *Pressimmo on Ligne v. Yakaz*, unpublished, but available at <https://www.legifrance.gouv.fr/affichJuriJudi.do?idTexte=JURITEXT000031478862>, accessed 6 September 2018.

(401) *Legal annex*, cit., 121.

lap between slavish imitation and other intellectual property rights (402). Now, while a final conclusion on the issue cannot be reached, it would seem that the recourse to the unfair commercial practices regime is unlikely to lead to over-protection of databases because most makers are unfamiliar with it (403) and unfair competition is not used much in database-related litigation (404), unlike contracts that are the crucial element in many database disputes.

Even when a database is not protected by copyright and sui generis right – better, above all when it is not – contracts and TPMs are used to restrict access to databases in a way which is problematic particularly when it comes to the “de facto monopolization of data by sole-source database producers” (405). Possible solutions include a compulsory license and an obligation on the part of provider “to actually deliver the data under fair and non-discriminatory terms” (406), like in the realms of standard essential patents (407) and telecommunications law (408).

Indeed, the detrimental and over-protective consequences of the use of contracts in the protection of database was made clear in *Ryanair v PR Aviation* (409), which is particularly relevant from an AI perspective because it regards an automated meta-search engine. The defendant was a website operator allowing consumers to search through the flights of low-cost air companies, compare conditions, and book a flight. Its meta-search engine gets the data in an automated way from a dataset linked to the Ryanair website. The defendant’s screen scraping, i.e. the automated extraction of data from a website (410), was in violation of the Terms & Conditions (411) that put in place an exclusive distribution system and prevent unauthorised websites to sell Ryanair flights. The use of the website was limited to private non-commercial purposes.

Now, the defendant’s arguments, upheld by the referring court, revolved around that the national implementation of the Database Directive provided some

(402) DERCLAYE and LEISTNER, *Intellectual Property Overlaps: A European Perspective*, Hart, Oxford, 2011, 173.

(403) Nearly 60% of the database makers surveyed in the context of the JIIP and Technopolis Group, *Legal annex*, cit., 123, answered “I do not know” to the question “How would you compare the protection of your databases through unfair competition law with their protection via the sui generis right?”.

(404) Sixty-five percent of the respondents affirmed that they never encountered legal proceedings where unfair competition law was used to protect databases. *Ibidem*.

(405) HUGENHOLTZ, *Abuse*, cit., 219.

(406) *Ibidem*.

(407) See, recently, BOSWORTH, MANGUM, and MATOLO, *FRAND Commitments and Royalties for Standard Essential Patents*, in BHARADWAJ DEVAIAH, and GUPTA (eds), *Complications and Quandaries in the ICT Sector*, Springer, Singapore, 2018, 19 ff.

(408) Court of Justice 25 November 2004, *KPN v. OPTA*, case C-109/03, in ECR, 2004, I, 11273. See Pace, *Comunicazioni elettroniche, servizio di repertorizzazione e superdominanza*, in *Europa e diritto privato* 2006, 851.

(409) *Ryanair*, cit., 312.

(410) Cf. CASPERS and GUIBAULT, *Baseline Report of Policies and Barriers of TDM in Europe*, Future TDM, Wien, 2016, 9.

(411) The current version of the Terms & Conditions, as updated on 5 September 2018, no longer contains such provisions. See Ryanair General Terms & Conditions of Carriage, available at <https://www.ryanair.com/gb/en/useful-info/help-centre/terms-and-conditions>.

limitations to the contractual autonomy. In particular, the rightholder cannot prevent the lawful user from accessing the contents and making normal use of them (412) and this exception cannot be contractually overridden (413). However, the referring court noted that the database *de quo* was not protected because of lack of originality and of substantial investment. Therefore, the question to the Court of Justice was whether the scope of the Database Directive covers unprotected databases and, consequently, whether the limits on contractual freedom resulting from the non contractually-overridable exceptions and users' rights (414) apply also to such databases (415).

According to the Court, it is immaterial that Ryanair's database matches the definition of database given in the directive. The latter provides for two different sets of rights and obligations that apply only if the relevant criteria are met (respectively, originality and substantial investment). Therefore, in the Court's reasoning, if the database maker does not have a right under the directive, then exceptions cannot be invoked against him or her. The conclusion (416) is that the Database Directive is not applicable to a database which is not protected either by copyright or by the *sui generis* right, therefore there are no limits to the rightholder's freedom to lay down contractual limitations on its use by third parties, because the exceptions provided by the directive will not apply.

The reasoning is weak for three interwoven reasons. First, Article 1 does not merely define databases, as held in *Ryanair*. It expressly deals with the scope of the directive. Any database in any form will fall within the scope of the directive, as long as the materials are independent, arranged systematically or methodically, and individually accessible. Therefore, since *Ryanair*'s website falls within the scope, there is no reason not to recognise the applicability of the provision on the binding nature of the exceptions. The difficulty of which provision to apply, whether the one on copyright or the one on the *sui generis* right, is only a practical one that could be overcome for instance finding the highest common denominator between Article 6 and Article 8, which does not seem complicated since the provisions are indeed quite similar. A different solution would, indeed, be unreasonable because it would lead to recognise a stronger protection to those databases where the author did not put in place any intellectual effort or any meaningful investment. Moreover, the main justification of the Database Directive is to stimulate investments in the database industry to bridge the gap between the US and the EU market. This goal cannot be achieved applying *Ryanair*, because it is in the database makers' interest not to cre-

(412) Dutch Copyright Act (*Auteurswet*), Article 24a(1), which corresponds to Article 6(1) of the Database Directive. The Court of Justice does not refer to the homologous provision applicable to the *sui generis* right, recognising the lawful user's right to the insubstantial extraction and re-utilisation of the contents of the database (Database Directive, Article 8).

(413) Dutch Copyright Act, Article 24a(3), corresponding to Article 15 of the Database Directive.

(414) The wording is slightly different even though the content is similar. On the one hand, there are exceptions to the database copyright (Article 6), on the other hand, rights and obligations of the lawful user of databases covered by *sui generis* right (Article 8).

(415) *Ryanair*, cit., para. 28.

(416) *Ryanair*, cit., para. 49.

ate original databases and not to invest significantly in obtaining, verifying, and presenting contents. Thus, they will be outside the scope of the Database Directive and will be able to restrict the users' rights without limits. While arguably Ryanair makes of contracts the key tool to protect databases, database makers should be wary of the intrinsic limitations of contract law, the main being the principle of privity (417). As a rule, obligations cannot be imposed on those who are not party to a contract; therefore, once extracted, the further dissemination of the information from third parties cannot be prevented (418).

It is important to keep in mind Ryanair when analysing British Horseracing and Fixtures Marketing, which will be done in section 5 below. Suffice it to say that the trend towards the narrowing of the scope of the Database Directive may have the unforeseen outcome of worsening the problem of the monopolisation of information and raw data, which can be easily achieved if the contractual freedom is without limits (419). Said trend may be seen as slowed by Verlag Esterbauer and Apis-Hristovich, whose broad definition of database partly offsets the problems arising from the joint operation of Ryanair and Fixtures Marketing. A reform of the Directive should, therefore, either clarify and broaden its scope of application or move the provisions on the exceptions in Chapter 4 on the common provisions (420). More ambitious plans may involve a harmonisation of contract law, that may focus on intellectual property contracts or encompass the main principles of contract law, now that with Brexit there is more homogeneity between the contract law traditions of the Member States.

All this is relevant also from an AI perspective. Indeed, since AI databases are unlikely to be considered original and to be protected by a *sui generis* right, as will be clarified below, this means that there is an incentive for database owners to invest in AI databases rather than in traditional ones because the former will be protected more strongly by means of contracts (421) and TPMs (422). The European Commission, in its evaluation of the Database Directive, observed that "the *sui generis* right is generally ignored in contractual frameworks" (423). However, this

(417) Privity is a common law principle, but similar concepts are present in most jurisdictions, as expression of the principle *res inter alios acta aliis nec nocet nec prodest*. See, recently, RĂDULESCU, *The principle of relative effect of contracts. A historical view and aspects of comparative law*, in *Challenges of the Knowledge Society* 2018, XII, 292.

(418) SYNODINOU, *Ryanair Ltd v. PR Aviation BV: contracts, rights and users in a low cost database law*, in *Kluwer Copyright Blog*, 26 January 2015, available at <http://copyrightblog.kluweriplaw.com/2015/01/26/ryanair-ltd-v-pr-aviation-bv-contracts-rights-and-users-in-a-low-cost-database-law/>.

(419) Cf. Commission, *Evaluation*, cit., 25 where it is stated that "the reduced scope may come under pressure in the future, potentially producing unexpected results in relation to machine- and sensor-generated 'big data'".

(420) Cf. JIIP and TECHNOPSIS GROUP, *Legal annex*, cit., 117. As pointed out by DREXL, *op. cit.*, para. 184, the legislature could "promote access through unwaivable exceptions and limitation as part of a comprehensive legislation of data ownership".

(421) In terms, Drexl, *op. cit.*, para 42, who underlines that "factual control over data can enable the data holder to commercialise that data without additional legal protection by relying on contract law".

(422) Ivi, para. 183.

(423) Commission, *Evaluation*, cit., 17.

should not be interpreted as meaning that contracts are not playing a key role in the data proprietisation, but only that the “the sui generis right does not seem to be widely used as a licensing tool” (424).

In conclusion, the powerful players of the data economy are commodifying our data with a multi-pronged strategy whose main elements are uncircumventable TPM (425), as well as Terms of Service and privacy policies that most people do not read, let alone understand or negotiate (426). The contractual limitations provided by the Database Directive, therefore, may play a vital role in better balancing the competing interests of the database industry and of the public.

8. The conclusions of this paper are fourfold. First, in order to grasp AI databases it is pivotal to understand that AI is an umbrella term encompassing a number of different technologies, where the degree of autonomy and ‘intelligence’ varies greatly. General or strong AI has not been achieved yet, but AI applications become increasingly refined and complex, therefore exploring proper AI works is no longer a work of science fiction, but one of the most pressing issues that intellectual property lawyers are urged to grapple with.

Second, AI is a formidable engine for text and data mining activities which, in turn, are playing a crucial role in the advancement of the database industry and of research worldwide. Existing copyright exceptions may cover some steps of the mining process, but overall are not fit for purpose and can be circumvented by contracts. The EU reform of copyright in the Digital Single Market, currently in the trilogue phase, provides a text and data mining exception that, unfortunately, is too weak of an attempt, especially because it is limited to research organisms and research purposes. The fact that data mining may not always have a legal basis does not mean, however, that the resulting database would not be protected under the Database Directive, should the relevant requirements be met.

Third, current EU copyright laws as interpreted by the Court of Justice do not allow copyright on AI works, including databases. The main, although not only, hurdle is the originality conundrum. Originality means that the work must be the author’s own intellectual creation. With AI works, if one considers the human as the author, it cannot be said it is the latter’s own intellectual creation, because they could not impress their personal touch on the work (427). The AI itself cannot be considered as author, let alone owner, mainly stemming from issues of lack of personality that seem far from being resolved. This, alongside other arguments, is the basis of the firm conviction that AI works should receive a low protection that could take the form of the sui generis right.

Finally, and more importantly, contrary to popular belief, AI databases can be

(424) Ibid..

(425) See COLSTON, *Protecting Databases – A call for regulation*, 2007, IXX, 85.

(426) See, e.g., NOTO LA DIEGA, *Uber law and awareness by design. An empirical study on online platforms and dehumanised negotiations*, in *Revue européenne de droit de la consommation* 2016, II, 383 ff., and DUCATO, *House of Terms: Fixing the Information Paradigm with Legal Design*, in *BILETA* 2018, Aberdeen, 10-11 April 2018.

(427) This paper build on the conviction the machine-generated works are not the same thing as AI works.

covered by a sui generis right. AI is likely to lead to a renovated, and perhaps unprecedented, importance of the sui generis right as the preferential form of protection for AI works. Indeed, AI works, not in themselves copyrightable, could be protected if organised in a database. The contestation that the sui generis right is not fit for AI is not based on the directive, but on the restrictive interpretation that the Court of Justice gave to its scope in its 2004 rulings. Those rulings were based on the wrong assumption that there is a dichotomy between creating data and obtaining data, whereas this distinction no longer holds in the data economy where the binary between creating and obtaining data has long been disrupted, as shown by how data mining works. The sterilising effect of the 2004 ruling may be reduced by the overcoming of the common misunderstanding that the court introduced a spin-off theory interpreted as meaning that if making databases is not the main activity of a company (by-product databases), then this database is not covered by a sui generis right. In fact, spin-off databases can also be protected if the owner proves a substantial investment, quantitatively or qualitatively. AI is a flexible tool and lends itself for the optimisation of the processes of obtaining, verifying, and presenting contents. It seems particularly likely that if there is a substantial investment in an AI application developed for the setting up of a database, the latter will be covered by a sui generis right.

In conclusion, countering the narrow interpretation of the scope of the Database Directive given by the Court of Justice may have three positive consequences. First, it would provide a form of protection to AI works otherwise in the public domain, thus striking a balance between the interests to the commercialisation of AI works and to the access to knowledge. Second, it would reduce the negative consequences of *Ryanair*, which allowed the abuse of contracts to achieve over-protection of data and databases. Those who feared that the sui generis right would have led to a disastrous monopolisation of information cannot but notice that in the data economy, the propertisation of data is the combined effect of contracts, TPMs, and trade secrets, which are leading to an overprotection of data. Third, it would nip in the bud unfortunate proposals of a new data producer's right, that would no longer be necessary – if it ever was – because the sui generis right would provide a sufficient protection. Therefore, instead of the abolition of the Database Directive, there are strong arguments to relaunch the sui generis right, which thanks to AI will eventually abandon the peripheries of the intellectual property realm.

Abstract

This paper deals with those databases where Artificial Intelligence technologies are used to obtain, verify, or present the database's contents ('AI databases'). The overarching research question is whether AI databases can be protected under the copyright and sui generis regimes provided by the Database Directive. The alleged inadequacy of the sui generis right for the data economy and, in particular, for machine-generated data led the European Parliament to call on the Commission to abolish said right and the Commission to propose the introduction of a data producer's right as a new property that would have done what the sui generis right had been unable to. It is this paper's contention that, contrary to popular belief, the sui generis right is fit for AI databases and that a different solution would lead to an overprotection of said subject matter by contractual means. The sui generis right

may be the best, if not the only, way to protect AI 'authorial' works. Indeed, even if AI works currently fall outside the scope of copyright law for lack of originality, they could nonetheless be protected if part of a database. Thus, thanks to AI, the sui generis right may become more important than it ever was.