



IAIC



DGBIC



CREDA

DIRITTO MERCATO TECNOLOGIA

FONDATA E DIRETTA DA

Alberto M. Gambino

COMITATO DI DIREZIONE

Valeria Falce, Giusella Finocchiaro, Oreste Pollicino,
Giorgio Resta, Salvatore Sica

15 febbraio 2021

AI-Generated Database

Paolo Burdese

COMITATO SCIENTIFICO

Guido Alpa, Fernando Bocchini, Giovanni Comandè, Gianluca Contaldi,
Vincenzo Di Cataldo, Giorgio Floridia, Gianpiero Gamaleri, Gustavo Ghidini,
Andrea Guaccero, Mario Libertini, Francesco Macario, Roberto Mastroianni,
Giorgio Meo, Cesare Mirabelli, Enrico Moscati, Alberto Musso, Luca Nivarra,
Gustavo Olivieri, Cristoforo Osti, Roberto Pardolesi, Giuliana Scognamiglio,
Giuseppe Sena, Vincenzo Zeno-Zencovich, Andrea Zoppini

E

Margarita Castilla Barea, Cristophe Geiger, Reto Hilty, Ian Kerr, Jay P. Kesan,
David Lametti, Fiona MacMillan, Maximiliano Marzetti, Ana Ramalho,
Maria Páz Garcia Rubio, Patrick Van Eecke, Hong Xue



Nuova
Editrice
Universitaria

La rivista è stata fondata nel 2009 da Alberto M. Gambino ed è oggi pubblicata dall'Accademia Italiana del Codice di Internet (IAIC) sotto gli auspici del Ministero dei beni e delle attività culturali e del turismo - Direzione generale biblioteche e istituti culturali (DGBIC) e dell'Università Europea di Roma con il Centro di Ricerca di Eccellenza del Diritto d'Autore (CREDA). Tutti i diritti sono dell'IAIC.

Comitato dei Valutazione Scientifica

EMANUELA AREZZO (Un. Teramo), EMANUELE BILOTTI (Un. Europea di Roma), FERNANDO BOCCHINI (Un. Federico II), ROBERTO BOCCHINI (Un. Parthenope), ORESTE CALLIANO (Un. Torino), LOREDANA CARPENTIERI (Un. Parthenope), LUCIANA D'ACUNTO (Un. Federico II), VIRGILIO D'ANTONIO (Un. Salerno), FRANCESCO DI CIOMMO (Luiss), PHILIPP FABBIO (Un. Reggio Calabria), MARILENA FILIPPELLI (Un. Tuscia), CESARE GALLI (Un. Parma), MARCO MAUGERI (Un. Europea di Roma), ENRICO MINERVINI (Seconda Un.), MARIA CECILIA PAGLIETTI (Un. Roma Tre), ANNA PAPA (Un. Parthenope), ANDREA RENDA (Un. Cattolica), ANNARITA RICCI (Un. Chieti), FRANCESCO RICCI (Un. LUM), GIOVANNI MARIA RICCIO (Un. Salerno), CRISTINA SCHEPISI (Un. Parthenope), BENEDETTA SIRGIOVANNI (Un. Tor Vergata), GIORGIO SPEDICATO (Un. Bologna), ANTONELLA TARTAGLIA POLCINI (Un. Sannio), RAFFAELE TREQUATTRINI (Un. Cassino), DANIELA VALENTINO (Un. Salerno), FILIPPO VARI (Un. Europea di Roma), ALESSIO ZACCARIA (Un. Verona).

Norme di autodisciplina

1. La pubblicazione dei contributi sulla rivista "Diritto Mercato Tecnologia" è subordinata alla presentazione da parte di almeno un membro del Comitato di Direzione o del Comitato Scientifico e al giudizio positivo di almeno un membro del Comitato per la Valutazione Scientifica, scelto per rotazione all'interno del medesimo, tenuto conto dell'area tematica del contributo. I contributi in lingua diversa dall'italiano potranno essere affidati per il referaggio ai componenti del Comitato Scientifico Internazionale. In caso di pareri contrastanti il Comitato di Direzione assume la responsabilità circa la pubblicazione.

2. Il singolo contributo è inviato al valutatore senza notizia dell'identità dell'autore.

3. L'identità del valutatore è coperta da anonimato.

4. Nel caso che il valutatore esprima un giudizio positivo condizionato a revisione o modifica del contributo, il Comitato di Direzione autorizza la pubblicazione solo a seguito dell'adeguamento del saggio.

La Rivista adotta un Codice etico e di buone prassi della pubblicazione scientifica conforme agli standard elaborati dal Committee on Publication Ethics (COPE): Best Practice Guidelines for Journal Editors.

Comitato di Redazione – www.dimt.it – dimt@unier.it

ALESSANDRO ALBANESE GINAMMI, MARCO BASSINI, CHANTAL BOMPREZZI, FRANCESCA CORRADO, CATERINA ESPOSITO, GIORGIO GIANNONE CODIGLIONE, FERNANDA FAINI, MONICA LA PIETRA, SILVIA MARTINELLI, DAVIDE MULA (Coordinatore), ALESSIO PERSIANI, ROSARIA PETTI, MARTINA PROVENZANO (Vice-Coordinatore), MATILDE RATTI, CECILIA SERTOLI, SILVIA SCALZINI, ANDREA STAZI (Coordinatore)

Sede della Redazione

Accademia Italiana del Codice di Internet, Via dei Tre Orologi 14/a, 00197 Roma, tel. 06.3083855, fax 06.3070483, www.iaic.it, info@iaic.it

AI-GENERATED DATABASES

*Do the creation/obtaining dichotomy and the substantial investment requirement exclude the sui generis right provided for under the EU Database Directive?
Reflections and proposals**

Paolo Burdese

ABSTRACT: *The starting point of this analysis is the position of the European Commission, which, in its 2018 evaluation of the Directive, clearly stated that the sui generis right does not apply broadly to the data economy, including Artificial Intelligence (AI). This position was justified by the spin-off theory and the lack of substantiality of the investment involved in creating the databases concerned.*

According to Article 7 of the Database Directive, in order to benefit from sui generis right protection, the maker of a database has to make a qualitatively or quantitatively substantial investment in either obtaining, verifying or presenting the contents of the database. While the CJEU and domestic courts agreed that the threshold for the substantiality of the investment should be low, major issues arose in connection with the aim of this investment. Indeed, due to judgments handed down by the CJEU in 2004, investment in creating the content of a database cannot be interpreted as “obtaining” such materials and, therefore, is not relevant. This interpretation created the so-called creation/obtaining dichotomy. However, while these judgments significantly reduced the number of protectable databases, the CJEU clearly rejected the spin-off theory.

Several years later, the CJEU’s judgment in the Ryanair case made the framework even more complex, the court ruling that the legitimate user of a database not protected under Article 7 of the Directive cannot benefit from the rights granted under Articles 8 and 15. This decision gave rise to a true

* The author is indebted to Prof. Guido Noto La Diega, Prof. Alessandro Cogo and Dr. Bryan Khan for their helpful suggestions. Opinions and errors are solely the author’s.

paradox, because an unprotected database can benefit from stronger protection by contractual arrangement.

Looking beyond the legal framework, this paper challenges the view of the European Commission and asserts that, as a matter of principle, AI-generated databases can be protected under the sui generis right. Furthermore, the fact that this kind of database is usually generated by data-recording or data-mining processes, which involves obtaining rather than creating, means that the creation/obtaining dichotomy is no longer tenable and, therefore, should be abandoned. Moreover, rejecting this dichotomy would also mitigate the negative outcomes of the Ryanair decision.

However, it seems clear that granting easier protection in this way might generate serious access-related issues, especially where sole-source databases are concerned. What is needed is a balance that, echoing the 1992 Proposal of the Directive, could be achieved by introducing compulsory licensing provisions.

SUMMARY: 1. Introduction – 2. Core AI-related concepts and the importance of AI-generated databases within the EU framework – 3. Data lakes and databases. What is relevant for our analysis? – 4. Obtaining, verification, and presentation of a database’s contents. The issues with the creating/obtaining dichotomy – 5. The (not-so) substantial investment threshold – 6. Sneaking through the window: the *Ryanair* case – 7. Application of the requirements to AI-generated databases – 8. Policy considerations – 9. Conclusion

1. Introduction

Data is central to the digital transformation of the economy and society.¹ Considering the volume of data already produced and the volume expected

¹ European Commission, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – A European strategy for data* [COM(2020) 66 final], Brussels, 2020, 1.

in the future,² as well as the projected value of the data economy (as much as €829 billion in 2025),³ it is a matter of urgency to assess whether the provisions of the EU’s legal framework are adequate. The main purpose of this paper is to investigate whether AI-generated databases are protected under the sui generis right granted by Directive 96/9 (hereinafter the “Database Directive”).⁴ Even though some commentators have called (and are still calling) for it to be abolished,⁵ as we shall see, granting such protection could be the best solution, especially considering the alternatives, which would most likely lead to over-protection.

The starting point of this analysis is the position of the European Commission, which, in its 2018 evaluation of the Directive, clearly stated that the sui generis right does not apply broadly to the data economy, including Artificial Intelligence (AI).⁶ After a brief analysis of the core AI-related concepts, and consideration of the importance of AI-generated databases, it will be necessary to verify whether they meet the requirements set out in the Database Directive.

In particular, while the notion of a database does not seem to be problematic, there are concerns relating to the substantial investment in obtaining, verifying or presenting required by Article 7 of the Directive. These concerns relate mainly to the narrow scope of application of the protection at issue, determined by CJEU judgements handed down in 2004,⁷

² Up to 175 zettabytes (one zettabytes equals 10^{21} bytes) in 2025. See *ivi*, 2.

³ See European Commission, *European data strategy – Making the EU a role model for a society empowered by data*, available at <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy>, last accessed 29 April 2020.

⁴ 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, Articles 7-11.

⁵ V. Falce, *L’insostenibile leggerezza” delle regole sulle banche dati nell’Unione dell’innovazione*, in *Diritto Industriale* 2008, 399-400.

⁶ European Commission, Executive Summary of the Evaluation of Directive 96/9/EC on the legal protection of databases [SWD (2018) 146 final], Brussels, 2018, 2

⁷ European Court of Justice, 9 November 2004, case C-203/02, *British Horseracing Board v. William Hill Organization*, “British Horseracing case”; European Court of Justice, 9 November 2004, case C- 338/02, *Fixtures Marketing v. Svenska Spel*, “Fixtures-Svenska case”; European Court of Justice, 9 November 2004, case C- 46/02, *Fixtures Marketing v. Oy Vekkaus AB*, “Fixtures-Oy case”; European Court of Justice, 9

which nullified the relevance of the investment involved in creating the data collected in a database.

Finally, considering the possible contractual protection of databases that do not meet these requirements (considerably strengthened by the CJEU), legal changes will be proposed as policy recommendations.

2. Core AI-related concepts and the importance of AI-generated databases within the EU framework

Before examining the requirements of Article 7 of the Database Directive, it may be useful to say a few words about the importance of AI and possible applications of AI-generated databases.

The term “Artificial Intelligence” was coined in the 1955 by John McCarthy,⁸ according to whom it means “*the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable*”.⁹ One of the clearest current definitions of AI is provided by the ISO, which defines it as “*the capability of a functional unit to perform functions that are generally associated with human intelligence such as reasoning and learning*”.¹⁰

Nowadays, a distinction is made between narrow AI, aimed at developing software “*to solve limited practical problems*” and strong AI

November 2004, case 444/02, Fixtures Marketing v. Organismos prognostikon agonon podosfairou, “Fixtures-OPAP case”, 20 ff.

⁸ See J. McCarthy, *A proposal for the Dartmouth Summer Research Program on Artificial Intelligence (31 August 1955)*, available at <https://www.aaai.org/ojs/index.php/aimagazine/article/view/1904/1802>, last accessed 29 April 2020.

⁹ J. McCarthy, *What is Artificial Intelligence?*, available at <http://jmc.stanford.edu/articles/whatisai.html>, last accessed 29 April 2020, 2.

¹⁰ ISO, *ISO/IEC 2382-28:1995(en) Information technology — Vocabulary — Part 28: Artificial intelligence — Basic concepts and expert systems – Voice 28.01.02 Artificial Intelligence*, available at <https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:-28:ed-1:v1:en>, last accessed 21 May 2020.

(also known as Artificial General Intelligence – AGI), which “*seeks to engineer human-level general intelligence based theoretical models*”.¹¹ At the current stage of development, even though AGI now seems to be an achievable goal,¹² narrow AI is deemed to be “*dominant in the field because it [has been] successful in solving useful practical problems*”.¹³

For the purposes of this paper, particular attention will be devoted to the way in which a machine is able to learn and, in particular the role of human intervention during such training. Indeed, machine learning can be supervised, unsupervised or semi-supervised. In the first scenario, the data set “ingested” by the machine consists of data that has been labelled¹⁴ by a human operator. In the second, the machine “learns”, i.e. experiences the world by processing unlabelled data. In the third hypothesis, there is a mix of the first two scenarios.¹⁵

The importance of AI for databases is undisputed.¹⁶ Indeed, the crucial role of AI in shaping the EU framework has been recently reaffirmed by the European Commission,¹⁷ and the impact of the data-driven economy¹⁸ on

¹¹ Both definitions are provided by T. Goertzel, *The path to more General Artificial Intelligence*, in *Journal of Experimental & Theoretical Artificial Intelligence* 2014, available at https://www.researchgate.net/publication/263429725_The_path_to_more_general_artificial_intelligence/link/5a6a17580f7e9b01f3efc656/download, last accessed 30 April 2020, 1.

¹² G. Noto La Diega, *Artificial Intelligence and databases in the age of big machine data*, in *AIDA* 2018, 98. See also T. Goertzel, *op. cit.*, 2, which argues that “*the next stage of development of AI, for the next decade and more likely for the next twenty-five years, will be increasingly dependent on contributions from strong AI*”. Even if, according to R. Fjelland, *Why general artificial intelligence will not be realized*, in *Humanit. Soc. Sci. Commun.*, 2020, available at <https://doi.org/10.1057/s41599-020-0494-4>, “*the goal cannot in principle be realized, and [...] the project is a dead end*”.

¹³ T. Goertzel, *op. cit.*, 1.

¹⁴ For instance, the human operator labels as “car” a picture of a car.

¹⁵ For a “lawyer-friendly” explanation of machine learning (supervised, unsupervised and semi-supervised), see G. Noto La Diega, *op. cit.*, 98-100 and footnotes thereof, which provide more technical insights. See also A. Ottolia, *Big Data e Innovazione Computazionale*, G. Giappichelli Editore, Turin, 2017, 10-13.

¹⁶ See G. Noto La Diega, *op. cit.*, 100, and for a detailed list of technical publications on the integration of AI and databases see footnotes 56-60 thereof.

¹⁷ See European Commission, *White paper on Artificial Intelligence – A European approach to excellence and trust* [COM (2020) 65 final], 2020, Brussels.

¹⁸ Broad term that also encompasses AI.

databases has been clearly demonstrated in the studies supporting the evaluation of the Database Directive.¹⁹ From a purely quantitative perspective, the application of AI technologies to databases is a game-changer, because it allows the system to collect, process and store a tremendous amount of data. Furthermore, from a qualitative perspective, an AI system would be able to find new correlations between data, thus creating more complex databases.

3. Data lakes and databases. What is relevant for our analysis?

Under Article 1(2) of the Database Directive, a database is “*a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means*”.

This concept is very different from the concept of data lakes. Indeed, while the latter term refers to mere flows of data indistinctly collected,²⁰ the former, even if broadly interpreted,²¹ requires the materials to be arranged in a systematic or methodical way and to be individually accessible.

In the AI field, most of the issues under discussion seem to relate to data lakes.²² However, they are not taken into account for the purposes of this paper, which focuses on databases as the output of AI systems activities.

4. Obtaining, verification, and presentation of a database’s contents. The issues with the creating-obtaining dichotomy

Under Article 7(1) of the Database Directive: “*Member States shall provide for a right for the maker of the database which shows that there has*

¹⁹ See JIIP and Technopolis Group, *Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases*, Luxembourg, 2018, 25 ff. It should be pointed out that, in this section of the Study, the focus is on web scraping and sensor-generated data. However, both technologies can be integrated with AI systems.

²⁰ See A. Ottolia, *op. cit.*, 73 ff.

²¹ E. Derclaye, *The Legal Protection of Databases: a Comparative Analysis*, Edward Elgar, Cheltenham, 2008, 54; and Fixtures-OPAP, 20 ff.

²² A. Ottolia, *op. cit.*, 75.

been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of the database.”

While the meaning of the expression “qualitatively and/or quantitatively a substantial investment” will be analyzed in the next section, here the focus will be on the “obtaining, verification or presentation of the contents”, and in particular on the so-called creating/obtaining dichotomy.

This dichotomy arises from four 2004 CJEU judgments: *British Horseracing*; *Fixtures-Svenska*; *Fixtures-Oy*; *Fixtures-OPAP*.²³ In all four cases, the Court stated that, in order to assess whether the maker of the database has invested substantially in obtaining the database’s contents, the resources used to create the elements collected should not be taken into account.²⁴ In reaching this conclusion, the CJEU relied on several recitals of the Directive, seeing as decisive the fact that, according to Recitals 9, 10 and 12, the aim of the Directive is to incentivize and protect investments in the development of storage and processing systems.²⁵ Furthermore, in Recital 39 the Court found another argument in support of its thesis. According to Recital 39, “*this Directive seeks to safeguard [...] against misappropriation of the results of the financial and professional investment made in obtaining and collection the contents [...]*”.²⁶

Before examining the aftermath of these decisions, it should be pointed out that the conclusion reached by the Advocate General – Ms. Stix-Hackl – was different. She made a distinction between the mere creation of data and instances in which such creation coincided with the processing of it, holding that in the latter case, when the two activities are not separable, the protection applies.²⁷

²³ See footnote 7.

²⁴ *British Horseracing*, § 31, *Fixtures-Svenska*, § 24, *Fixtures-Oy*, § 34, *Fixtures-OPAP*, § 40.

²⁵ *British Horseracing*, § 30, *Fixtures-Svenska*, § 23, *Fixtures-Oy*, § 33, *Fixtures-OPAP*, § 39.

²⁶ *British Horseracing*, § 32, *Fixtures-Svenska*, § 25, *Fixtures-Oy*, § 35, *Fixtures-OPAP*, § 41.

²⁷ *British Horseracing* (Opinion of A.G.), 46 and 49, *Fixtures-Svenska* (Opinion of A.G.), 56 and 59, *Fixtures-Oy* (Opinion of A.G.), 66 and 69, *Fixtures-OPAP* (Opinion of A.G.), 72 and 75.

Academic commentary on the consequences of these decisions presents a wide range of opinions. While some authors see the CJEU interpretation as positive because of its anti-monopolistic effect where novel data is concerned,²⁸ others do not consider these interventions to be effective in restricting the monopolistic effect of the sui generis right.²⁹

Moreover, some authors immediately pointed out that this construction could be easily circumvented. Indeed, creators of the data concerned may adopt several tricks in order to “obtain” a protected outcome.³⁰ In particular, database-makers could introduce procedural systems to enable a clear delineation between investment related to the creation of the data and investment related to its presentation or verification.³¹ Other possible ways of circumventing the dichotomy include the use of technological protection measures (TPM)³² or other means of denying access to the data, such as contractual terms. Furthermore, data producers could assign the data – possibly to a sister company – before making it publicly available, allowing

²⁸ For example, see P.B. Hugenholtz, *Program schedules, event data and telephone subscriber listings under the Database Directive: the spin-off doctrine in the Netherlands and elsewhere in Europe*, Paper presented at the Fordham School of Law 11th Annual Conference on International Intellectual Property Law and Policy, 2003, available at <https://www.ivir.nl/publicaties/download/spinofffordham.pdf>, last accessed 4 May 2020, 7; E. Derclaye, *Databases sui generis right: should we adopt the spin off theory?*, in *EIPR* 2004, 412; and E. Derclaye, *The legal protection of databases: a comparative analysis*, *cit.*, 94.

²⁹ V. Falce, *op. cit.*, 398. In particular, according to this author, the narrow line between the creation and the obtaining of data presents a clear risk that the protection could be extended to the data as such. Commentators who have seen in the decisions at issue a way of preventing the possible anti-monopolistic effect of the sui generis right have also recognized this difficulty, see e.g. M.J. Davison and P.B. Hugenholtz, *Football fixtures, horse races and spin-offs: the ECJ domesticates the database right*, in *EIPR* 2005, available at https://www.ivir.nl/publicaties/download/EIPR_2005_3_databaseright.pdf, last accessed 4 May 2020, 5.

³⁰ J. Gaster, “*Obtinere*” of data in the eyes of the ECJ - How to interpret the Database Directive after *British Horseracing Board Ltd et al. v. William Hill Organisation Ltd*, in *Computer und Recht: International* 2005, 135. See also, M.J. Davison and P.B. Hugenholtz, *op. cit.*, 6.

³¹ M.J. Davison and P.B. Hugenholtz, *op. cit.*, 6.

³² *Ibidem*.

the assignee to make a substantial investment in obtaining it (the same reasoning could apply to the mere outsourcing of the creation of data).³³

Finally, other commentators have strongly criticized this construction on the basis of the recitals upon which the CJEU built its conclusions. Indeed, the recitals could be read as leading to the opposite interpretation.³⁴ While under Recital 9 databases are deemed to be vital tools in the development of the information market, reduction of the scope of their protection actually hampers the development of this market.³⁵ Secondly, Recital 10 refers to the growth in the amount of information generated and processed, while calling for significant investment in advanced information-processing systems.³⁶ Thirdly, as pointed out in Recital 12, the investment in information storage and processing systems requires that there be uniform legal protection of databases in the interests of stability. From this point of view, the 2004 CJEU interventions have effectively “sterilized” the Directive, thus not granting the above-mentioned stable protection.³⁷ It should also be pointed

³³ Ivi, 7. However, according to E. Derclaye, *The legal protection of databases: a comparative analysis*, cit., 96-97, this “getting around”, does not trigger protection, because the investment should be directed at producing a database (see also T.F. Aplin, *Copyright in the digital society – The challenges of multimedia*, Hart Publishing, Oxford, 2005, 67-68). Despite the prestige and authority of the author and her opinion, we respectfully disagree, because this could prevent application of the protection in the cases in which the database-maker needs to acquire from the same source, against a payment, some of the data constituting its databases. Moreover, even if we accept this argument, the question remains: how do you draw a line between the acquisition of a set of data and the acquisition of an already made database?

³⁴ Some scholars, even before the 2004 judgments, highlighted that the distinction between the production of data and its acquisition is not reflected in the Directive. Thus, “costs incurred by generating information are analogously to be considered as “obtaining”, even if this activity as such is not sufficient to trigger the protection. See G. Westkamp, *EU database protection for information uses under an intellectual property scheme: has the time arrived for a flexible assessment of the European Database Directive?*, Paper presented at the Fordham School of Law 11th Annual Conference on International Intellectual Property Law and Policy, 2003, available at <https://ssrn.com/abstract=1115432>, last accessed 5 May 2020, 7.

³⁵ Ivi, 117.

³⁶ *Ibidem*.

³⁷ G. Noto La Diega, *op. cit.*, 118.

out that, while the main goal of the Directive was to stimulate investment in this market, the majority of investments are in data creation.³⁸

One of the most controversial issues associated with the created/obtained dichotomy concerns so-called recorded data, i.e. “*data [that] occur in nature or in time and are generally recorded by instruments of measure in order for them to be intelligible by man. Examples include results of sport competitions, meteorological, astronomical data and genomic data*”.³⁹ Here again, opinions differ. While some commentators argue that this data is created,⁴⁰ others have suggested that it is both created and collected,⁴¹ and others that it is only obtained.⁴² This latter position has been indirectly adopted by the CJEU in *Verlag Esterbauer*.⁴³ In this case, Freistaat Bayern and the company Verlag Esterbauer GmbH were in dispute over the protection under the Database Directive of a topographical map. It is significant that neither the parties nor the judge who referred the question to

³⁸ JIIP and Technopolis Group, *op. cit.*, 26.

³⁹ E. Derclaye, *The legal protection of databases: a comparative analysis, cit.*, 98-99.

⁴⁰ See M.J. Davison and P.B. Hugenholtz, *op. cit.*, 6. According to these two authors, because of the strict approach adopted by the CJEU in 2004, recorded data should be considered as created and not obtained, because they are just representations of the phenomena and not the phenomena themselves. See also J. Drexler, *Design competitive markets for industrial data – Between proprietisation and access*, in *JIPITEC*, 2017, 268.

⁴¹ E. Derclaye, *The legal protection of databases: a comparative analysis, cit.*, 99. Here, the author seems to recognize that investment related to the recording of such data could be considered as triggering the *sui generis* protection. Indeed, according to her view, they pre-exist in nature and anyone could record them, without creating a monopoly. However, she also points out that there could be the risk of a monopoly or dominant position being established when the cost of for recording the data is very high, thus enabling only a few entities to collect them.

⁴² See G. Noto La Diega, *op. cit.*, 118; M. Leistner, *Big data and the EU Database Directive 96/9/EC: current law and potential reform*, in S. Lohsse, R. Shulze and D. Staudenmyer (eds.), *Trading data in the digital economy: legal concepts and tools*, Nomos, Baden-Baden, 2017, also available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3245937, last accessed 5 May 2020, 2.

⁴³ European Court of Justice, October 29, 2015, case C-490/14, *Freistaat Bayern v. Verlag Esterbauer GmbH*, “*Verlag Esterbauer case*”. Furthermore, the same conclusion has been reached by the German Supreme Court, see BGH, March 25, 2010, I ZR 47/08, “*Autobahnmaut case*” and the EWCA, February 6, 2013, *Football Dataco Ltd v. Stan James Ltd (No 2)*. In the latter case, the Court considered live football data to be obtained. However, on another occasion, a German court reached the opposite conclusion, see G. Noto La Diega, *op. cit.*, 118 fn. 191.

the CJEU suggested that topographical data might be considered as created rather than obtained. In addition, the data involved in a recording is pre-existing information, and this operation, finalized to make it intelligible, would seem to be nothing more than the processing of information that the Directive aims to promote. Furthermore, this conclusion is supported by the great difference between created data (such as the match fixtures organized by national football associations, which do not exist before they are arranged) and recorded data (such as geographical or astronomical coordinates, which exist independently of their recording).

Moving on to the requirement set forth in Article 7 relating to the “verifying” and “presenting” of the elements of the database, the CJEU has provided guidance on both terms. Indeed, while verifying has been read as “ensuring [...] the reliability of the information contained in that database, to monitor the accuracy of the materials collected when the database was created and during its operation”,⁴⁴ presenting has been interpreted as “giving the database its function of processing information, that is to say those used for the systematic or methodical arrangement of the materials contained in that database and the organization of their individual accessibility”.⁴⁵ Regarding cases in which the data has been created, some authors have identified a certain difficulty in assessing the existence of independent investment in verifying the data, on the one hand, and presenting it, on the other.⁴⁶

5. The (not-so) substantial investment threshold

Under Article 7, for the *sui generis* right to apply, investment in obtaining, verifying or presenting the database’s elements has to be qualitatively or quantitatively substantial.

⁴⁴ Fixtures-Svenska, 27; Fixtures-Oy, 37; Fixtures-OPAP, 43.

⁴⁵ Fixtures-Svenska, 27, Fixtures-Oy, 37, Fixtures-OPAP, 43.

⁴⁶ See, *inter alia*, E. Derclaye, *The legal protection of databases: a comparative analysis, cit.*, 95. In particular, the author highlights the impossibility of drawing a line between investment in creation and in verification of content. It is easier to recognize such investment in presenting, especially when the data have been created in a disorganized fashion. However, in many cases, the created data are “*somewhat organized*”.

Examining the recitals related to this requirement,⁴⁷ it should first be pointed out that the investment can be financial, technical and/or “human”.⁴⁸

Secondly, regarding the threshold above which an investment may be deemed substantial, majority opinion tends to set a low bar. Indeed, domestic courts have on several occasions defined as “substantial” investment in databases that required just a few hours of work⁴⁹ or that were made by simply listing 1,650 email addresses.⁵⁰ Indeed, even though there are some divergent positions,⁵¹ setting a low threshold would seem to be more aligned with the aims of the Directive, as applying a high threshold would result in protecting bigger databases rather than smaller ones.⁵² Given that the objective of the European legislator is to promote the market in information, this divide would work in the opposite direction. Moreover, smaller databases contribute to the development of the said market. If the bar is set high, as has been correctly pointed out, many makers would prefer to wait for their databases to become “costly” enough before releasing them.⁵³

⁴⁷ Recitals 7, 39 and 40 of Database Directive.

⁴⁸ E. Derclaye, *The legal protection of databases: a comparative analysis*, cit., 74 and the case law referred in footnote 192. In particular, the author refers to relational investment, i.e. the investment involved in maintaining the team building the database.

⁴⁹ *Ivi*, 80, referring to the case of T. Com. Nanterre, May 16, 2000, PR Line v. Newsinvest; and G. Noto La Diega, *op. cit.*, 120, fn. 212, referring to Court of Rome, December 10, 2009.

⁵⁰ E. Derclaye, *The legal protection of databases: a comparative analysis*, cit., 74, referring to the case of TGI Paris, April 25, 2003, Sonacotra v. Syndicat Sud Sonacotra.

⁵¹ See, inter alia, G. Westkamp, *Protecting databases under US and European law – methodical approaches to the protection of investments between unfair competition and intellectual property concepts*, in IIC 2003, 782. According to this author, one of the main arguments in support of a high threshold is that it “would lessen the danger of information monopolies by creating a broader public domain”. However, as pointed out in E. Derclaye, *The legal protection of databases: a comparative analysis*, cit., 87, making a database from pre-existing elements does not create a monopoly because the data are still collectable.

⁵² E. Derclaye, *The legal protection of databases: a comparative analysis*, cit., 88

⁵³ This view is shared by many commentators. See for instance E. Derclay, *The legal protection of databases: a comparative analysis*, cit., 88; and M. Leistner, *Legal protection for the database maker: initial experience from a German point of view*, in IIC 2002, 450. These arguments were also used by the Administrative Court of Rostock,

It is important also to state that this requirement must be applied in an absolute way, i.e. by setting a *de minimis* standard. Given that the threshold for the substantiality of the investment should be low, denying the relative character of the condition at issue and excluding the relevance of related factors (e.g. the size of the company or the usual level of investment in its particular field) would not limit smaller companies' access to protection. At the same time, this approach would lead to uniform application of this criterion, thus ensuring the stability of the legal basis for the sui generis right.⁵⁴

Finally, the investment must be substantial from a quantitative or qualitative perspective. The former refers to quantifiable resources (i.e. time and money), the latter to intellectual effort or energy.⁵⁵ The relevant aspect of this part of the Article 7 is that these requirements can also be alternative. Thus, a database that has required only a qualitatively substantial investment could be protected even in the absence of significant financial commitment. This aside, these two terms are not particularly important, considering that they regard the investment itself and not the amount or quality of the elements of the database.⁵⁶

see AC Rostock, February 20, 2001, Datenbankeigenschaft von Hyperlinksammlungen. Finally, the Advocate General, in three of her 2004 opinions, also expressed preference for a low threshold, holding, inter alia, that requiring a high investment would undermine the aim of the Directive, which is to incentivize investment in the field at issue. See *Fixtures-Svenska* (Opinion of A.G.), 39; *Fixtures-Oy* (Opinion of A.G.), 49; *Fixtures-OPAP* (Opinion of A.G.), 55.

⁵⁴ See E. Derclaye, *The legal protection of databases: a comparative analysis, cit.*, 83 and 84. However, the Advocate General, even though recognizing the need to assess a *de minimis* standard in an absolute way, has clearly favoured a relative approach, specifying that evaluation of the substantiality of an investment should be carried out considering the scale, nature and contents of the database, and the sector involved. See *Fixtures-Svenska* (Opinion of A.G.), 35, 38 and 39; *Fixtures-Oy* (Opinion of A.G.), 45, 48 and 49; *Fixtures-OPAP* (Opinion of A.G.), 51, 54 and 55.

⁵⁵ See *Fixtures-Svenska*, 28, *Fixtures-Oy*, 38 and *Fixtures-OPAP*, 43.

⁵⁶ E. Derclaye, *The legal protection of databases: a comparative analysis, cit.*, 91.

6. Sneaking through the window. The *Ryanair* case

Before applying the requirements described in the previous sections, it is important to highlight what the future holds for databases that do not comply with the provisions of Article 7.

Prima facie we might think that they would be simply unprotected but, at the same time, subject to the other relevant provisions of the Directive. In particular, we might believe that Articles 8 and 15 would still be applicable. Under Article 8(1) of the Database Directive, the maker of a database made available to the public cannot prevent a lawful user from extracting and/or re-utilizing insubstantial parts of it, for any purposes. Moreover, according to Article 15 of the Directive, any contractual term contrary to the above-mentioned provision is void. However, as stated by the CJEU in the *Ryanair* case⁵⁷, Articles 8 and 15 of the Directive do not apply to databases that are not protected under Article 7.⁵⁸ This heavily criticized⁵⁹ ruling has given rise to a paradox,⁶⁰ because it means that the maker of an unprotected database can enjoy greater protection than that accorded to the maker of a protected database. Obviously, this assumption is not absolute; on the contrary, it depends significantly on the bargaining power of the maker of the

⁵⁷ European Court of Justice, January 15, 2015, case C-30/14, *Ryanair Ltd v. PR Aviation BV*, “*Ryanair case*”.

⁵⁸ *Ryanair*, 39. The CJEU has justified this conclusion on the grounds that it is a literal and a systematic interpretation of the Directive. Indeed, the Court relied on the fact that Article 8 of the Directive falls within the chapter on the sui generis right. See M. Borghi and S. Karapapa, *Contractual restrictions on lawful use of information: sole-source databases protected by the back door?*, in *EIPR* 2015, also available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2862248, 4.

Moreover, as pointed out in Recital 40, the Directive is built on a balance between the rights of the database-maker and those of lawful users. Consequently, in the absence of any such right, the above-mentioned limitations do not apply.

⁵⁹ M. Borghi and S. Karapapa, *op. cit.*, 10-12. One of the strongest arguments of the authors, according to whom Articles 8 and 15 of the Database Directive should be applicable to all databases, is that these provisions continue to apply to databases even after the term of protection has expired.

⁶⁰ European Commission, *2018 Evaluation, cit.*, 32 and JIIP and Technopolis Group, *op. cit.*, 22.

unprotected database. This paradox is therefore more likely to favour big players, with all the attendant consequences.

Certainly, it is true that “*the significance of the decision probably depends on how far a particular country’s law treats the contractual terms imposed by the website owner as binding on users*”.⁶¹ However, the paradoxical risk of overprotection of unprotected database is real.⁶²

7. Application of the requirements to AI-generated databases

Keeping in mind the construction of the above-mentioned requirements, it is now possible to assess whether the European Commission’s exclusion of AI-generated databases from the scope of the sui generis right is correct or not.

According to the European Commission, one of the main reasons why the sui generis right does not fit AI-generated databases is the so-called spin-off theory.⁶³ According to this theory, databases that are by-products of the main activity of the maker should not enjoy protection under Article 7 of the

⁶¹ European Commission, *2018 Evaluation, cit.*, 32 and JIIP and Technopolis Group, *op. cit.*, 22.

⁶² European Commission, *2018 Evaluation, cit.*, 32 and JIIP and Technopolis Group, *op. cit.*, 22. Indeed, in order to show that the Ryanair-related risks are less real than they would appear, the European Commission referred to the ruling handed down by the Court of Appeal of Amsterdam, November 22, 2016, *Pearson v. Bar Software*, which stated that contractual provisions prohibiting the copying or duplication of any part of the databases without the right holder’s permission were unenforceable. However, the reason behind this a decision is to be found in the wording of the clause. A different wording could easily avoid this outcome. See, L. Trapman, *Pearson v. Bar Software: on the use of unprotected databases*, available at <http://copyrightblog.kluweriplaw.com/2017/01/31/pearson-v-bar-software-use-unprotected-databases/>, last accessed 9 May 2020.

⁶³ European Commission, *Executive summary of the Evaluation of Directive 96/9/EC, cit.*, 2. However, European Commission, *Evaluation of the Directive 96/9/EC on the legal protection of databases* [SWD(2018) 147 final], Brussels, 2018 (hereinafter “*2018 Evaluation*”), 15 seems – wrongly in our view – to ascribe this unfitness to the absence of substantial investment, which in turn is deemed to derive from the *spin-off* theory.

Directive.⁶⁴ However, the European Commission's conclusion is not acceptable for two main reasons.

Firstly, as correctly pointed out by several commentators,⁶⁵ the CJEU rejected this theory, clearly stating that it is not the link between the database and the principal activity of the maker that precludes protection, but the absence of the requirements set by Article 7 of the Directive.⁶⁶

Secondly, the European Commission's position does not take account of the fact that, even if application of the theory at issue is the favored option, there might be companies whose main activity is to create and market AI-generated databases.⁶⁷ Indeed, the spin-off theory seems to focus on the link between the database and the main activity of its maker, not on the way in which the data is collected, verified or presented. Therefore, if the above-mentioned activities comply with the requirements for protection, in a scenario of this kind it is not possible to exclude the sui generis right.

Having rejected the spin-off theory, we conclude that the evaluation should be carried out in the light of the "substantial investment in obtaining, presenting or verifying" requirement. Following the two-part approach adopted in the previous section, we will first examine the issues related to investment in obtaining, presenting or verifying, and then the substantiality of such investment.

Regarding the first aspect, we can begin by examining the case in which AI generates a database, creating and collecting the data by processes that are indistinguishable. In this case, at least *de lege lata*, according to the

⁶⁴ For the origins of this theory, which was developed in the Netherlands, and several cases in which Dutch and other European courts have considered it, see P.B. Hugenholtz, *Program schedules, event data and telephone subscriber listings under the Database Directive: the spin-off doctrine in the Netherlands and elsewhere in Europe*, *cit.*, 2 ff. and E. Derclaye, *Database sui generis right: should we adopt the spin off theory?*, *cit.*, 402-406.

⁶⁵ See, inter alia, A. Cogo, *Fixtures-Oy – Comment to the judgement*, in *AIDA* 2005, 416; M. Leistner, *Big data and the EU Database Directive 96/9/EC: current law and potential reform*, *cit.*, 2, fn. 10; E. Derclaye, *The legal protection of databases: a comparative analysis*, *cit.*, 94; J. Gaster, *op. cit.*, 133.

⁶⁶ British Horseracing, 35, *Fixtures-Svenska*, 29, *Fixtures-Oy*, 39 and *Fixtures-OPAP*, 44.

⁶⁷ This is not an unlikely hypothesis. Think, for instance, of services of the kind offered on Bdex (see <https://www.bdex.com/>).

construction established by the CJEU in 2004, the issue of protection hardly arises. Indeed, as pointed out by the Court, when data is created by the same entity that generates the database, its collection does not require any particular investment.⁶⁸ Furthermore, in this scenario it does not seem possible to identify any separable investment in verifying the data.⁶⁹ However, a small “opening” for protection might be found where investment in presenting the data is concerned. Indeed, even though data is generally created in an organized way,⁷⁰ when an AI system is developed without features determining the organization of the outcomes, data might be produced in a disorganized manner. Thus, development of another AI algorithm to manage the presentation of such data could be seen as a substantial investment.

Moreover, differentiation of the algorithms – one for the creation of the contents (for instance through data mining) and one for obtaining, verifying or, in particular, presenting the data created by the former – could be a way of circumventing the narrow interpretation provided by the CJEU.⁷¹ Similarly, such circumvention could be achieved by assigning the data to a sister company.

Where so-called recorded data is concerned, a different scenario (again related to the first aspect of the requirement) applies. As stated above, it does not seem possible to consider such data as created rather than obtained. In the AI field, we can identify data recorded by sensors, and then processed by the AI software, and data obtained through data mining processes.⁷² In both cases, the investment would be in obtaining the data, thus triggering protection provided the substantiality threshold was met. In particular, given that the issues related to sensor-recorded data have been examined above, data mining can best be analyzed here because it is strictly related to AI. This process⁷³ consists in analysing “*text and data in digital form in order to*

⁶⁸ Fixtures-Svenska, 33; Fixtures-Oy, 44; Fixtures-OPAP, 49.

⁶⁹ E. Derclaye, *The legal protection of databases: a comparative analysis*, cit., 95.

⁷⁰ *Ibidem*.

⁷¹ G. Noto La Diega, *op. cit.*, 119.

⁷² *Ivi* 118. I

⁷³ The process could be summarized in these four stages: i) scraping and crawling in order to search the relevant contents and retrieve the information; ii) creation of a target

generate information which includes but is not limited to patterns, trends and correlations”.⁷⁴ Even though the wording of the DSM Directive refers to the generation of information, we believe that, for our purposes, as stated above, it has to be interpreted as obtained. Indeed, if “*this process leads to the discovery of correlations between existing data*”,⁷⁵ then it could be equated (*mutatis mutandis*, of course, and with an incredibly higher level of complexity) to the measuring of the distance between two places. Like distance, such correlation is something that pre-exists its collection and therefore cannot be deemed to be created.

Furthermore, even if such data were created and not obtained, we should consider the possibility of a substantial investment in verifying or – more likely – presenting the data. For instance, especially in the case of sensor-recorded data, it is possible that the data would be collected in a disorganized way. The same is true of the data flow deriving from the use of social networks.⁷⁶

Moving on to the substantiality threshold (which has to be low), development of the algorithm and the system itself requires an investment in time and money, and eventually – if a development team has to be created – in “human relationships”. Also critical for the functioning of an AI system is its training. As we have seen above, this activity may or may not entail human intervention (e.g. in the labelling process). Again, this human intervention amounts to an investment. Where this requirement is concerned, the main problem seems to be in cases in which AI is capable of producing more than one database. Indeed, as correctly pointed out,⁷⁷ if subsequent production in such cases did not require any new investment, protection would not be granted. If creation of a new database requires a further

dataset; iii) analysis; iv) presentation of the findings. See *Ivi*, 129-130 and footnotes 285-288 for technical references.

⁷⁴ Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, 2019, OJ L 130/92 (hereinafter “*DSM Directive*”), Article 2(2).

⁷⁵ G. Noto La Diega, *op. cit.*, 124.

⁷⁶ A. Ottolia, *op. cit.*, 75.

⁷⁷ *Ivi*, 121.

“labelling session”, and if the machine does not carry out such activity automatically, it is possible to claim this as a new investment.

8. Policy considerations

Even if it were clear that, despite the evaluation of the European Commission, AI-generated databases are protectable under Article 7 of the Database Directive, the next important question is whether the legal framework is able to afford sufficient and balanced protection. Unfortunately, the answer seems to be negative.

Indeed, as recorded by the study in support of the 2018 Evaluation, nearly half of the database-makers surveyed reported that more than half of the contents of their databases consisted of generated data.⁷⁸ Going by the CJEU’s interpretation, not all these databases would be protected by the *sui generis* right.

Furthermore, we cannot ignore the position of those who consider that recorded data should be qualified as created data, rather than obtained.⁷⁹ Indeed, this qualification would necessarily also extend to the outcome of data mining processes. The result would be quite dramatic for the protection of AI-generated databases.

Thus, this twofold reduction – one actual and one potential – would not lead to a broader set of data freely available but, on the contrary, would trigger the negative consequences of the contractual protection resulting from the Ryanair case. Indeed, as in the above-mentioned case, if the requirements set forth in Article 7 were not met, the safeguards for lawful users provided under Article 8(1) and “strengthened” by Article 15 would not apply. Furthermore, there might be the risk of the Ryanair rationale also being applied, by analogy, to Articles 3, 5, 6, and 7 of the DSM Directive. These provisions have introduced three new exceptions to the right conferred under Article 7 of the Database Directive: text and data mining for the

⁷⁸ JIIP and Technopolis Group, *op. cit.*, 26.

⁷⁹ M.J. Davison and P.B. Hugenholtz, *op. cit.*, 6.

purpose of scientific research (Article 3); use of databases for digital and cross-border teaching activities (Article 5); and preservation of cultural heritage (Article 6). Furthermore, under Article 7 of the DSM Directive, such exceptions cannot be prevented by contractual means.

Given this risk, it is our conviction that there are two possible solutions, both implying a reform of the Database Directive.

The first is intended to counter the overprotection generated by the *Ryanair* decision. In this case, a simple addition to Article 8 of the Directive, extending its effect to all databases falling within the definition set forth in Article 1(2), might be sufficient. Indeed, in this case the applicability of Articles 8 and 15 would be detached from protection of the database. Such provisions might then be applicable, in an analogous way, to the new set of exceptions set forth in the DSM Directive.

However, an intervention of this kind would only partly solve the problem. AI-generated databases (and more generally Big Data) pose other challenges, related mainly to the creation/obtaining dichotomy, on the one hand, and to access to data, on the other.

In particular, as stated above, the issues deriving from the dichotomy created in 2004 by the CJEU are threefold. Firstly, almost half of database-makers generate more than half of the contents of their databases.⁸⁰ Secondly, even if we believe that such content is “obtained”, the qualification of sensor-generated and mined data is controversial, and therefore the creation/obtaining dichotomy at issue is no longer tenable.⁸¹ Thirdly, as we have seen, the dichotomy is easily circumvented. Therefore, broadening the scope of Article 7, and adopting a solution reflecting the one proposed by the AG in the 2004 cases, could solve several problems. It would solve the issue of the status of sensor-generated and mined data, qualifying the investment involved in its obtention as relevant under Article 7. In addition, even in the absence of an intervention to counter the negative consequences of the *Ryanair* judgment, granting protection to a larger number of databases would guarantee a broader application of Articles 8 and

⁸⁰ JIIP and Technopolis Group, *op. cit.*, 26.

⁸¹ G. Noto La Diega, *op. cit.*, 118.

15 of the Directive and of the new exceptions introduced by the DSM Directive.⁸²

An intervention of this kind seems to be the best solution. Indeed, while it is not possible to maintain the Directive in its present form for the above-mentioned reasons, and because – as we have to keep in mind – it was designed for a different environment, alternatives such as a *sui generis* right for data are strongly criticized by many scholars.⁸³

Of course, this solution would entail a risk for data access, especially in the case of sole-source databases. However, it would not affect the creation of data itself, but only cases in which creation and collection are not separable. Furthermore, this risk could be overcome by introducing compulsory licenses,⁸⁴ which the right holder would have to grant under fair, reasonable and non-discriminatory terms. This solution was firstly put forward in the 1992 Proposal⁸⁵ regarding sole source databases,⁸⁶ and needs to be re-proposed in order to achieve a better balance between the easier protection arising from the removal of the creation/obtaining dichotomy and access issues in all those cases in which it is not possible to collect data in other ways. As suggested in 1992, the content of these licenses should encompass the extraction and reutilization of the contents of the database for commercial purposes. However, as argued long ago,⁸⁷ a system of this kind should not lead to duplicative use of the content of the database but should only cover adding-value uses or, at least, uses in sufficiently different

⁸² The use of ownership – or, as in the scenario considered by this paper, the extension of the scope of application of an existing right – as a technique for controlling and balancing situations in which there are imbalances in bargaining power is not new. See, for instance, J. Drexl, *op. cit.*, 275.

⁸³ See, for a deeper analysis, J. Drexl, *op. cit.* and P.B. Hugenholtz, *Data property: unwelcomed guest in the house of IP*, available at https://www.ivir.nl/publicaties/download/Data_property_Muenster.pdf, last accessed 22 May 2020.

⁸⁴ JIIP and Technopolis Group, *op. cit.*, 35 ff.; and M. Leistner, *Big data and the EU Database Directive 96/9/EC: current law and potential reform*, *cit.*, 11.

⁸⁵ European Commission, *Proposal for a Council Directive on the legal protection of databases* [COM(1992) 24 final].

⁸⁶ *Ivi*, Article 8.

⁸⁷ See J. Ginsburg, *Creation and commercial value: copyright protection of works of information*, in *Colum. L. Rev.*, 1990, 1930.

contexts.⁸⁸ Finally, the provisions related to compulsory licensing should also take into account the possible existence of trade secrets, as well as the legal framework guaranteeing data protection and privacy.

In addition, as suggested by other scholars, issues relating to access could be solved by bringing the entirety of a database within the scope of Article 8,⁸⁹ not only the insubstantial parts of it. Finally, the EU legislator should carefully consider the call⁹⁰ for alignment of the exceptions provided under the Database Directive and those granted under the InfoSoc Directive.⁹¹ The fact is that the exceptions currently set forth in Article 9 of the Database Directive – even if we add in those introduced by the DSM Directive – do not adequately take into account the central role played by data. First, therefore, it would be necessary to abolish the distinction – made in Article 9(a) of the Directive – between electronic and non-electronic databases where extraction for private purposes is concerned. Moreover, all the exceptions provided in Article 9 of the Directive apply only to lawful users. Therefore, as correctly pointed out,⁹² this limitation should be abandoned, because it appears to be inconsistent with the copyright framework. Another reason for such an extension is that databases might be instrumentalized as a way of protecting the AI-generated works collected within them.⁹³ This being the case, allowing users to rely on the huge set of exceptions provided under Article 3 of the InfoSoc Directive would open up far wider access to these works.

⁸⁸ JIIP and Technopolis Group, *op. cit.*, 42.

⁸⁹ M. Leistner, *Big data and the EU Database Directive 96/9/EC: current law and potential reform*, *cit.*, 9. Furthermore, considering that this provision seems to mirror Article 7 of the Directive, granting an extension of this kind would differentiate between legitimate users and others, who are able to extract insubstantial parts as well. See M. Borghi and S. Karapapa, *op. cit.*, 9. This would be a further incentive for users to negotiate a license with the right holder.

⁹⁰ M. Leistner, *Big data and the EU Database Directive 96/9/EC: current law and potential reform*, *cit.*, 13-14.

⁹¹ Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society, 2001, OJ L 167/10

⁹² M. Leistner, *Big data and the EU Database Directive 96/9/EC: current law and potential reform*, *cit.*, 13.

⁹³ G. Noto La Diega, *op. cit.*, 95.

9. Conclusions

Given the considerations expressed in the previous paragraphs, we can conclude that AI-generated databases could benefit from the *sui generis* protection granted under Article 7 of the Database Directive.

Despite the creation/obtaining dichotomy and the issue of uncertain status, we can see how the conclusions reached by the European Commission in its 2018 Evaluation are based on wrong assumptions: the spin-off theory and the lack of substantial investment involved. However, the CJEU has rejected the former argument and, as we have seen, there is likely to be qualitatively or quantitatively substantial investment in the field of AI-generated databases.

Nevertheless, given the challenges posed by the Big Data era and the dangers arising from the *Ryanair* judgment, the creation/obtaining dichotomy should be abandoned in favour of a broader approach that also grants protection in cases where investment in the creation and the collection of data is inseparable.

Finally, this greater protection should be balanced by the requirement of compulsory licenses, extending the activities permitted to the lawful user under Article 8 of the Directive to the entirety of databases, and by aligning the exceptions provided under Article 9 of the Directive with the ones granted by the InfoSoc Directive.

Bibliography

Books

T.F. Aplin, *Copyright in the digital society – The challenges of multimedia*, Hart Publishing, Oxford, 2005.

E. Derclaye, *The legal protection of databases: a comparative analysis*, Edward Elgar, Cheltenham, 2008.

P.B. Hugenholtz, *Data property: unwelcomed guest in the house of IP*, in H. Ullrich, P. Drahos, and G. Ghidini, *Kritika: essays on intellectual property – Volume 3*, Edward Elgar, Cheltenham, 2018, also available at https://www.ivir.nl/publicaties/download/Data_property_Muenster.pdf, last accessed 22 May 2020.

M. Leistner, *Big data and the EU Database Directive 96/9/EC: current law and potential reform*, in S. Lohsse, R. Shculze and D. Staudenmyer (eds.), *Trading data in the digital economy: legal concepts and tools*, Nomos, Baden-Baden, 2017, also available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3245937, last accessed 5 May 2020.

A. Ottolia, *Big data e innovazione computazionale*, G. Giappichelli Editore, Torino, 2017.

Articles

A. Cogo, *Fixtures-Oy – Comment to the judgement*, in *AIDA* 2005, 407 ff.

M. Borghi and S. Karapapa, *Contractual restrictions on lawful use of information: sole-source databases protected by the back door?*, in *EIPR*, 2015, also available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2862248, last accessed 5 May 2020.

M.J. Davison and P.B. Hugenholtz, *Football fixtures, horse races and spin-offs: the ECJ domesticates the database right*, in *EIPR* 2005, available at https://www.ivir.nl/publicaties/download/EIPR_2005_3_databaseright.pdf, last accessed 4 May 2020.

E. Derclaye, *Databases sui generis right: should we adopt the spin off theory?*, in *EIPR* 2004, 402 ff.

J. Drexel, *Designing competitive markets for industrial data – Between proprietisation and access*, in *JIPTEC*, 2017, 257 ff.

J. Gaster, “*Obtinere*” of Data in the Eyes of the ECJ - How to interpret the Database Directive after *British Horseracing Board Ltd et al. v. William Hill Organisation Ltd*, in *Computer und Recht: International* 2005, 129 ff.

J. Ginsburg, *Creation and commercial value: copyright protection of works of information*, in *Colum. L. Rev.*, 1990, 1865 ff.

T. Goertzel, *The path to more General Artificial Intelligence*, in *Journal of Experimental & Theoretical Artificial Intelligence* 2014, available at https://www.researchgate.net/publication/263429725_The_path_to_more_general_artificial_intelligence/link/5a6a17580f7e9b01f3efc656/download, last accessed 30 April 2020.

P.B. Hugenholtz, *Program schedules, event data and telephone subscriber listings under the Database Directive: the spin-off doctrine in the Netherlands and elsewhere in Europe*, Paper presented at the Fordham School of Law 11th Annual Conference on International Intellectual Property Law and Policy, 2003, available at <https://www.ivir.nl/publicaties/download/spinofffordham.pdf>, last accessed 4 May 2020.

M. Leistner, *Legal protection for the database maker: initial experience from a German point of view*, in *IIC* 2002, 439 ff.

R. Marvin, *AI Databases, what they are and why your business should care*, in *PCMag UK*, 26 October 2017, available at <https://www.pcmag.com/news/ai-databases-what-they-are-and-why-your-business-should-care>, last accessed 29 April 2020.

J. McCarthy, *A proposal for the Dartmouth Summer Research Program on Artificial Intelligence (31 August 1955)*, available at <https://www.aaai.org/ojs/index.php/aimagazine/article/view/1904/1802>, last accessed 29 April 2020.

J. McCarthy, *What is Artificial Intelligence?*, available at <http://jmc.stanford.edu/articles/whatisai.html>, last accessed 29 April 2020.

G. Noto La Diega, *Artificial Intelligence and databases in the age of big machine data*, in *AIDA* 2018, pp. 93 ff.

R. Fjelland, *Why general artificial intelligence will not be realized*, in *Humanit. Soc. Sci. Commun.*, 2020, available at <https://doi.org/10.1057/s41599-020-0494-4>.

L. Trapman, *Pearson v. Bar Software: on the use of unprotected databases*, available at <http://copyrightblog.kluweriplaw.com/2017/01/31/pearson-v-bar-software-use-unprotected-databases/>, last accessed 9 May 2020.

G. Westkamp, *EU Database protection for information uses under an intellectual property scheme: has the time arrived for a flexible assessment of the European Database Directive?*, Paper presented at the Fordham School of Law 11th Annual Conference on International Intellectual Property Law and Policy, 2003, available at <https://ssrn.com/abstract=1115432>, last accessed 5 May 2020.

G. Westkamp, *Protecting databases under US and European law – methodical approaches to the protection of investments between unfair competition and intellectual property concepts*, in *IIC* 2003, 772 ff.

Laws and Treaties

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.

Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society, 2001, OJ L 167/10.

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

Cases

European Court of Justice

European Court of Justice, 9 November 2004 case C-203/02, *British Horseracing Board v. William Hill Organization*, “British Horseracing case”.

European Court of Justice, 9 November 2004, case C- 338/02, Fixtures Marketing v. Svenska Spel, “Fixtures-Svenska case”.

European Court of Justice, 9 November 2004, case C- 46/02, Fixtures Marketing v. Oy Vekkaus AB, “Fixtures-Oy case”.

European Court of Justice, 9 November 2004, case 444/02, Fixtures Marketing v. Organismos prognostikon agonon podosfairou, “Fixtures-OPAP case”.

European Court of Justice, 15 January 2015, case C-30/14, Ryanair Ltd v. PR Aviation BV, “Ryanair case”.

European Court of Justice, 29 October 2015, case C-490/14, Freistaat Bayern v. Verlag Esterbauer GmbH, “Verlag Esterbauer case”.

Domestic Courts

France

T. Com. Nanterre, 16 May 2000, PR Line v. Newsinvest.

TGI Paris, 25 April 2003, Sonacotra v. Syndicat Sud Sonacotra.

Germany

AC Rostock, 20 February 2001, Datenbankeigenschaft von Hyperlinksammlungen.

BGH, 20 March 2010, I ZR 47/08, “Autobahnmaut case”.

Netherlands

Court of Appeal of Amsterdam November 22, 2016, Pearson v. Bar Software.

Other Documents

European Commission, *Proposal for a Council Directive on the legal protection of databases* [COM(1992) 24 final].

JIIP and Technopolis Group, *Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases*, Luxembourg, 2018.

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

European Commission, Evaluation of the Directive 96/9/EC on the legal protection of databases [SWD(2018) 147 final], Brussels, 2018.

European Commission, *White paper on Artificial Intelligence – A European approach to excellence and trust* [COM(2020) 65 final], Brussels, 2020.

European Commission, *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – A European strategy for data* [COM(2020) 66 final], Brussels, 2020.

European Commission, *European data strategy – Making the EU a role model for a society empowered by data*, available at <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy>, last accessed 29 April 2020.

ISO, *ISO/IEC 2382-28:1995(en) Information technology — Vocabulary — Part 28: Artificial intelligence — Basic concepts and expert systems — Voice 28.01.02 Artificial Intelligence*, available at <https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:-28:ed-1:v1:en>, last accessed 21 May 2020.

DIRITTO MERCATO TECNOLOGIA

Numeri Speciali

- 2016 LO STATUTO ETICO GIURIDICO DEI CAMPIONI BIOLOGICI UMANI
a cura di Dario Farace
- 2017 IL MERCATO UNICO DIGITALE
a cura di Gianluca Contaldi
- 2018 LA RICERCA SU MATERIALI BIOLOGICI DI ORIGINE UMANA:
GIURISTI E SCIENZIATI A CONFRONTO
a cura di Alberto M. Gambino, Carlo Petrini e Giorgio Resta
- 2019 LA TASSAZIONE DELL'ECONOMIA DIGITALE TRA SVILUPPI RECENTI
E PROSPETTIVE FUTURE.
a cura di Alessio Persiani

La rivista “Diritto Mercato Tecnologia” intende fornire un costante supporto di aggiornamento agli studiosi e agli operatori professionali nel nuovo scenario socio-economico originato dall’interrelazione tra diritto, mercato e tecnologia, in prospettiva interdisciplinare e comparatistica. A tal fine approfondisce, attraverso studi nei settori privatistici e comparatistici, tematiche afferenti in particolare alla proprietà intellettuale, al diritto antitrust e della concorrenza, alle pratiche commerciali e alla tutela dei consumatori, al biodiritto e alle biotecnologie, al diritto delle comunicazioni elettroniche, ai diritti della persona e alle responsabilità in rete.

