

American University Washington College of Law

## Digital Commons @ American University Washington College of Law

---

Joint PIJIP/TLS Research Paper Series

---

4-2021

### The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the failed Text and Data Mining limitations in the CSDM Directive

Christophe Geiger

*Centre for International Intellectual Property Studies (CEIPI), University of Strasbourg,*  
christophe.geiger@ceipi.edu

Follow this and additional works at: <https://digitalcommons.wcl.american.edu/research>



Part of the [Intellectual Property Law Commons](#), and the [International Trade Law Commons](#)

---

#### Recommended Citation

Geiger, Christophe. "The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the failed Text and Data Mining limitations in the CSDM Directive." (2021) PIJIP/TLS Research Paper Series no. 66. <https://digitalcommons.wcl.american.edu/research/66>

This Article is brought to you for free and open access by the Program on Information Justice and Intellectual Property and Technology, Law, & Security Program at Digital Commons @ American University Washington College of Law. It has been accepted for inclusion in Joint PIJIP/TLS Research Paper Series by an authorized administrator of Digital Commons @ American University Washington College of Law. For more information, please contact [DCRepository@wcl.american.edu](mailto:DCRepository@wcl.american.edu).

# THE MISSING GOAL-SCORERS IN THE ARTIFICIAL INTELLIGENCE TEAM: OF BIG DATA, THE FUNDAMENTAL RIGHT TO RESEARCH AND THE FAILED TEXT AND DATA MINING LIMITATIONS IN THE CSDM DIRECTIVE

*Christophe Geiger<sup>1</sup>*

## ABSTRACT

This article argues that recent strategies of the European Union in the field of Artificial Intelligence (AI) resemble a football team missing a goal-scorer to win any of the competitions with other jurisdictions having more flexible limitations to copyright, in particular with those allowing robust text and data mining (TDM) activities. It analyses the TDM limitations newly introduced in EU copyright law by the Directive on Copyright in the Digital Single Market to show that these provisions not only fail to take duly into account the right to research grounded in the fundamental right to information, but also will not allow the European Union to provide a competitive environment for the development of AI and data-driven innovations. As a conclusion, the article calls for a prompt revision of the copyright framework for TDM activities at EU and international levels, combined with an implementation of the directive by Member States that would be compliant with the fundamental rights framework of the EU and the objective advanced by European policy makers.

---

<sup>1</sup> Professor of law at the Centre for International Intellectual Property Studies (CEIPI), University of Strasbourg; Affiliated Senior Researcher, Max Planck Institute for Innovation and Competition (Munich).

On February 19, 2020, the European Commission announced an ambitious digital strategy for the European Union, setting out the objectives to be achieved in two communications – one on Europe’s digital future<sup>2</sup> and one on data<sup>3</sup> – both complemented by a White Paper on artificial intelligence<sup>4</sup>. Reading these strategic documents, one could hope that a sound and considered legal framework for the digital environment would finally be elaborated, and that EU policies in this area will in the future constitute more than a simple patchwork of sectorial interests. Indeed, with regard to the revolution that digital technology has brought to a large number of sectors within the EU, it reads: “This substantive societal transformation calls for a profound reflection at all levels of society as to how Europe can best meet, and continue to meet, these risks and challenges. It will require a huge effort, but Europe undoubtedly has the means to bring about this better digital future for everyone”<sup>5</sup>. Or: “Creating a Europe fit for the digital age is a complex puzzle with many interconnected pieces; as with any puzzle, the whole picture cannot be seen without putting all the pieces together”<sup>6</sup>. The will to finally modernize and adapt the legal framework on IP to the challenges posed by the digital environment has been reiterated in the action plan of the European Commission on IP, published at the end of November 2020, where it was emphasized that “the technological revolution – the data economy and society, the turn to artificial intelligence (AI), the growing importance of new technologies such as blockchain, 3D-printing and the Internet of Things (IoT) as well as the development of new business models such as the platform economy, and the data and circular economy - *offers a unique window of opportunity to modernize our approach to protecting our intangible assets*”<sup>7</sup>.

In short: A horizontal reflection on digital issues is therefore desired by the European Commission before (re)-defining the rules of the game through appropriate legislation. However, on closer inspection, neither in the proposed action plan, nor in the implementation agenda for these strategies, does there appear any questioning of the free spaces left by IP law in order to allow the development of a balanced digital ecosystem in the EU<sup>8</sup>. The issue has also been totally ignored by the Proposal for a

---

<sup>2</sup> Communication from the Commission, “Shaping Europe’s digital future”, Brussels, 19.02.2020, COM(2020) 67 final.

<sup>3</sup> Communication from the Commission, “A European strategy for data”, Brussels, 19.02.2020, COM(2020) 66 final.

<sup>4</sup> European Commission, “White Paper on Artificial Intelligence - A European approach to excellence and trust”, Brussels, 19.02.2020, COM(2020) 65 final.

<sup>5</sup> Communication from the Commission, “Shaping Europe’s digital future”, *supra*, p. 1.

<sup>6</sup> *Ibid.*, p. 3.

<sup>7</sup> Communication from the Commission, “Making the most of the EU’s innovative potential, An intellectual property action plan to support the EU’s recovery and resilience”, 25 Nov. 2020, COM(2020) 760 final, p. 2 (emphasis added).

<sup>8</sup> For an early critique of the lack of ambition of the EU legislator in the field of copyright, see the foundational reflections in: P.Bernt Hugenholtz (ed.), *The Future of Copyright in a Digital Environment*, Den Haag, Kluwer, 1996, in particular the chapter by the editor himself, “Adapting Copyright to the Information Superhighway”, p. 81 et

Regulation on a European approach for Artificial Intelligence published on April 21, 2021, despite the fact that one of the main goal put forward by the European Commission is to “ensure legal certainty to facilitate investment and innovation in AI”<sup>9</sup>. Unless the Commission considers that the question has already been settled, in particular by the directive of 17<sup>th</sup> April 2019 on Copyright and related rights in the Digital Single Market<sup>10</sup> (CDSM-Directive), it is difficult to understand how a horizontal reflection on digital innovation can be conducted without including a large part of digital law such as copyright law (including its exceptions and limitations) and intellectual property in general<sup>11</sup>. Worse, it is likely that the solutions that have been adopted beforehand in the CDSM-Directive are in direct contradiction with the objectives now displayed, and that - as a result - the many ambitions newly put forward are likely to remain a dead letter.

Many examples could be given, but we would like to highlight just one in this contribution: a proactive policy in the field of artificial intelligence and digital innovation undoubtedly requires proposing a modern legal framework which recognizes the risk that the development of artificial intelligence within the European Union may encounter obstacles or unjustified prohibitions<sup>12</sup>. In this context, it is known that in order to allow machine learning, which is essential to artificial

---

sq.

<sup>9</sup> Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, Brussels, 21.4.2021 COM (2021) 206 final.

<sup>10</sup> 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 L130/92-125 (CDSM Directive). For comments, see the various Opinions drafted by the European Copyright Society, at <https://europeancopyrightsociety.org/>.

<sup>11</sup> On the crucial importance of exceptions and limitations in EU copyright law (in particular for follow-on innovation and creativity), see i.e. P.Bernt Hugenholtz, “Fierce Creatures, Copyright Exemptions: Towards Extinction?”, IFLA/IMPRIMATUR, *Rights, Limitations and Exceptions: Striking a Proper Balance*, Amsterdam Oct. 1997, reprinted in: David Vaver (ed.), *Intellectual Property Rights: Critical Concepts in Law*, Vol. 2, London: Routledge, 2006, 231; Christophe Geiger, “Promoting Creativity through Copyright Limitations, Reflections on the Concept of Exclusivity in Copyright Law”, *Vanderbilt Journal of Entertainment & Technology Law* 2010, Vol. 12, Issue 3, p. 515; Christophe Geiger & Franciska Schönherr, “Defining the Scope of Protection of Copyright in the EU: The Need to Reconsider the Acquis regarding Limitations and Exceptions”, in: Tatiana-Eleni Synodinou (ed.), *Codification of European Copyright Law, Challenges and Perspectives*, The Hague/London/New York: Kluwer Law International, 2012, p. 133. And the various excellent contributions in Ruth L. Okediji (ed.), *Copyright Law in an Age of Limitations and Exceptions*, Cambridge: CUP, 2017.

<sup>12</sup> On the legal challenges posed by AI in the EU, see excellent the report drafted P. Bernt Hugenholtz, João Pedro Quintais & Daniel Gervais for the European Commission: IViR and JIPP, *Trends and Developments in Artificial Intelligence Challenges to the Intellectual Property Rights Framework*, Final Report, Sept. 2020, available at: <https://ec.europa.eu/digital-single-market/en/news/trends-and-developments-artificial-intelligence-challenges-intellectual-property-rights-0>, leaving out however “legal issues concerning the input of protected subject matter into AI systems (e.g. for text-and-data mining)” from the scope of the analysis.

intelligence, it is necessary to have robust exceptions for text and data mining so that the machine can reproduce, store and process existing data and propose new solutions<sup>13</sup>. The question is also very topical in the field of public health, because it is obvious that very large stocks of data on COVID-19 have to be processed and analyzed at a global level in order to allow the development of new treatments<sup>14</sup>.

Text and data mining has in fact become a fundamental tool for research, whether carried out by public bodies or by private actors<sup>15</sup>.

<sup>13</sup> See Philipp Hacker, “A Legal Framework for AI Training Data”, 13 *Law, Innovation and Technology* (2021, forthcoming), showing that “training data for AI do not only play a key role in the development of AI applications, but are currently only inadequately captured by EU”; Mauritz Kop, “The Right to Process Data for Machine Learning Purposes in the EU”, *Harvard Journal of Law & Technology*, Volume 34, Digest Spring 2021, p. 1. On the importance of text and data mining for artificial intelligence, see e.g. Christophe Geiger, “The Artificial Intelligence and Data-led Revolution of Copyright and its Wider Implications”, Digital Encounters, Final report, available at:

[https://www.fidefundacion.es/docs/GlobalDigitalEncounters/Encounter\\_8\\_Report\\_Final.pdf](https://www.fidefundacion.es/docs/GlobalDigitalEncounters/Encounter_8_Report_Final.pdf); Rosana Ducato & Alain Strowel, “Limitations to text and Data Mining and Consumer Empowerment: Making the Case for a Right to ‘Machine Legibility’”, *IIC* 2019, Vol. 50, No. 6, p. 649; Nicolas Binctin, “TDM : un enjeu de l’intelligence artificielle”, *RIDA* 2019, No. 262, p. 7; Theodoros Chiou, “Copyright lessons on Machine learning : what impact on algorithmic art ?”, *JIPITEC* 2019, n° 10, p. 398; Alexandra Bensamoun, Joelle Farchy & Paul-François Schira, *Intelligence artificielle et Culture*, Report for the CSPLA, January 2020, p. 54; Christophe Geiger, Giancarlo Frosio & Oleksandr Bulayenko, “Crafting a Text and Data Mining Exception for Machine Learning and Big Data in the Digital Single Market”, in: Xavier Seuba, Christophe Geiger & Julien Pénin (eds.), *Intellectual Property and Digital Trade in the Age of Artificial Intelligence and Big Data*, Collection CEIPI/ICTSD, *Global Perspectives and Challenges for the Intellectual Property System*, No. 5, Geneva/Strasbourg, 2018, p. 95.

<sup>14</sup> On the link between the two topics, see in particular Teresa Hackett, “COVID and Copyright: The Right to Research”, 17 August 2020, < [www.eifl.net](http://www.eifl.net) >; Sean Flynn, Christophe Geiger and Joao Pedro Quintais (with the collaboration of T. Margoni, M. Sag, L. Guibault, M. Carroll), “Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action”, *EIPR* 2020, No. 7, p. 393. See also the Statement on Copyright and Proposal of a Waiver from Certain Provisions of the Trade Related Aspects of Intellectual Property Rights (TRIPS) Agreement for the Prevention, Containment and Treatment of COVID-19 (IP/C/W/669), 22 March 2021, endorsed by 250 organizations and prominent researchers calling for the reduction of copyright barriers to COVID-19 prevention, containment and treatment, available at <http://infojustice.org/archives/43020>: “In too many countries, researchers lack the rights they need to use the most advanced research methodologies, such as text and data mining, to help find and develop treatments to COVID-19. Indeed, the virus itself was discovered by a text and data mining research project that would not be lawful in many countries”.

<sup>15</sup> See Recital 8 of the CDSM directive of 17 April 2019, according to which “there is widespread acknowledgment that text and data mining can, in particular, benefit the research community and, in so doing, support innovation”. See also Thomas Margoni, “Text and Data Mining in Intellectual Property Law: Towards an Autonomous Classification of Computational Legal Methods”, *CREATE working paper* 01/2020: “The impact that TDM may have on science, arts and humanities is invaluable. This is because by identifying the correlations and patterns that are often concealed to the eye of a human observer due to the amount, complexity, or variety of data surveyed, TDM allows for the discovery of concepts or the formulation of correlations that would have otherwise

However, as it has been argued by Bernt Hugenholtz and others<sup>16</sup>, the solution proposed by the 2019 directive is largely unsuitable, because it is far too restrictive with regard to the exceptions for text and data mining. Thus, to use a ‘footballistic’ image, the Artificial Intelligence strategy of the European Union might resemble a football team that would be left without any strikers to score successfully and to win any of the competitions with other jurisdictions which may have the advantage of more flexible legal provisions allowing broader TDM activities.

Why is this the case? Let us quickly come back to the two exceptions introduced into EU law by the 2019 directive.

First of all, it should be noted that the European Commission clearly did not understand at the beginning the importance of the text and data mining exceptions for the development of artificial intelligence. Admittedly, the proposal for a directive of 14<sup>th</sup> September 2016<sup>17</sup> provided for a mandatory exception<sup>18</sup> allowing text and data mining which could not be ruled out by contract<sup>19</sup>. However, it was limited to “reproductions and extractions made by research organisations in order to carry out text and data mining of works or other subject-matter to which they have lawful access for the purposes of scientific research”<sup>20</sup>. These beneficiary organizations were understood restrictively, since recital 11 specified that “research organisations across Member States generally have in common that they act either on a not for profit basis or in the

---

remained concealed or undiscovered. Considering this point of view, it can be effectively argued that TDM creates new knowledge from old data”.

<sup>16</sup> For a (critical) analysis see also P. Bernt Hugenholtz, “The New Copyright Directive: Text and Data Mining (Articles 3 and 4)”, *Kluwer Copyright Blog*, July 24 2019, <http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/>; Christophe Geiger, Giancarlo Frosio & Oleksandr Bulayenko, “Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU”, in: Conception Saiz Garcia & Raquel Evangelio Llorca (eds.), *Propiedad intelectual y mercado único digital europeo*, Valencia: Tirant lo Blanch, 2019, p. 27; Kop, *supra*, at p. 7.

<sup>17</sup> See the Proposal for a directive of the European Parliament and of the Council on copyright in the Digital Single Market, 14 September 2016, Brussels, COM(2016) 593 final, 2016/0280 (COD).

<sup>18</sup> What, unlike the 2001 directive, constitutes an improvement, as the list of 20 *optional* exceptions and limitations had only a limited impact in terms of harmonization. See on this issue e.g. Christophe Geiger & Franciska Schönherr, “The Information Society Directive” (with updates from Stavroula Karapapa), in: Irini Stamatoudi & Paul Torremans (eds.), *EU Copyright Law*, 2<sup>nd</sup> ed., Cheltenham, UK / Northampton, MA: Edward Elgar, 2021, p. 279.

<sup>19</sup> According to the recitals of the Proposal for a directive, the exception was intended to address the great legal uncertainty surrounding text and data mining activities (recital 10), because “text and data mining may involve acts protected by copyright and/or by the sui generis database right, notably the reproduction of works or other subject-matter and/or the extraction of contents from a database. Where there is no exception or limitation which applies, an authorisation to undertake such acts would be required from rightsholders” (recital 8).

<sup>20</sup> Art. 3, para. 1 of the Proposal.

context of a public-interest mission recognised by the State. Such a public-interest mission may, for example, be reflected through public funding or through provisions in national laws or public contracts<sup>21</sup>. This covered only research organizations and public universities (or those performing a public service mission).

But what about the significant research activities carried out by start-ups operating in the digital environment, which are the source of important innovations, particularly in the field of artificial intelligence, and whose potential in terms of growth has been strongly advanced in the Commission's new strategy?<sup>22</sup> Start-ups were not taken into consideration, and therefore their data mining activities remain subject to the exclusive right. Along these lines, what about public-private partnerships, so strongly encouraged when submitting applications in the context of calls for tenders to the European Union? Admittedly, recital 10 stated in a relatively vague manner that "research organisations should also benefit from the exception when they engage into public-private partnerships", without however explaining how this should be the case. Not to mention individual researchers without an affiliation to an institution, whose activities are not covered either, and journalists working by definition for private structures, who will also not benefit from the exception.

This situation is highly problematic in terms of fundamental rights, and in particular with regard to the right to research. Indeed, the right to research has a strong human rights foundation, and is protected at international, European and national levels<sup>23</sup>. Based on the right to

---

<sup>21</sup>See also article 2 ("Definitions") of the Proposal for a directive: "'research organization' means a university, a research institute or any other organization the primary goal of which is to conduct scientific research or to conduct scientific research and provide educational services: (a) on a non-for-profit basis or by reinvesting all the profits in its scientific research; or (b) pursuant to a public interest mission recognized by a Member State; in such a way that the access to the results generated by the scientific research cannot be enjoyed on a preferential basis by an undertaking exercising a decisive influence upon such organization". See also article 2, paragraph 3) according to which: "'cultural heritage institution' means a publicly accessible library or museum, an archive or a film or audio heritage institution".

<sup>22</sup> See "White Paper on Artificial Intelligence - A European approach to excellence and trust", *supra*, p. 4, according to which: "Europe is well placed to benefit from the potential of AI, not only as a user but also as a creator and a producer of this technology. It has excellent research centers, *innovative start-ups*, a world-leading position in robotics and competitive manufacturing and services sectors, from automotive to healthcare, energy, financial services and agriculture" (emphasis added).

<sup>23</sup> According to Article 19 of the Universal Declaration of Human Rights of 1948 "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and *to seek, receive and impart information and ideas through any media and regardless of frontiers*". Likewise, Article 19. 2 of the International Covenant on Civil and Political Rights of 1966 specifies that "everyone shall have the right to freedom of expression; this right shall include *freedom to seek, receive and impart information and ideas of all kinds*, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his

information, it includes an active right to search for effective and objective information by the use of existing sources, which implies in the digital environment to be able to use lawfully text and data mining (TDM) techniques to conduct research<sup>24</sup>. For journalists, it is in particular of crucial importance since certain investigations are only possible today through large-scale text and data mining. As an example, it should be recalled that the “Panama Papers” scandal, which highlighted a large-scale tax evasion of politicians, billionaires, celebrities and high-level sportsmen, could only be revealed by searching and mining information, via automated search techniques, from more than a million off-shore bank documents.

By restricting the scope of the exception in such a way, the Commission therefore risked creating an ineffective and therefore rapidly obsolete provision, in particular regarding the development of artificial intelligence, but also with regards to other activities of essential research and innovation not conducted by public bodies. This is especially so since the proposal for the directive only allowed research organizations to search texts and data “on works or other subject-matter to which they have lawful access for the purposes of scientific research”, which seemed to exclude a large part of online research since the legality of the sources would be uncertain<sup>25</sup>.

For all these reasons, the directive proposal was subject to much criticism during the parliamentary phase<sup>26</sup>, leading to changes in the final

---

choice” (emphasis added). In Europe, the right to information is derived from Art. 10(1) ECHR and Art. 11 of the Charter of Fundamental Rights of the European Union, and is included in several national constitutions such Art. 5(1) of the German Basic law, Art. 16(3) of the Federal Swiss Constitution and Art. 11 of the French Declaration of Human rights. In the EU, a “right to research” could find support also in Art. 13 of the Charter of Fundamental Rights (Freedom of the arts and sciences), which stipulates that “the arts and scientific research shall be free of constraint. Academic freedom shall be respected.” See e.g. Flynn, Geiger and Quintais, *supra*.

<sup>24</sup> The right to information consists of a passive freedom of receiving existing information, and an active right to search for effective and objective information by the use of existing sources. On the fundamental right to information in particular in the context of copyright law, see detailed Christophe Geiger, *Droit d’auteur et droit du public à l’information, approche de droit comparé*, Paris: Litec, 2004; “Author’s Right, Copyright and the Public’s Right to Information: A Complex Relationship”, in: F. Macmillan (ed.), *New Directions in Copyright Law*, Vol. 5, Cheltenham, UK: Edward Elgar, 2007, 24. And of course Bernt Hugenholtz’s, *Auteursrecht op informatie*, Deventer: Kluwer, 1989, initiating a discussion on the interface of copyright and freedom of information.

<sup>25</sup> The lawful source requirement can be, for example, found in the French exception of text and data mining, introduced into the Intellectual Property Code by the law “for a digital republic” of 7 October 2016 (Art. 122-5 10° CPI, and for databases, Art. L. 342-3, 5). It should be noted, however, that Germany, which also introduced such an exception in its copyright law in 2017, did not add this additional criterion (UrhG, par. 60d (1)).

<sup>26</sup> See in this regard Christophe Geiger, Giancarlo Frosio & Oleksandr Bulayenko, The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in

text of the directive<sup>27</sup>. As a result, the circle of beneficiaries of the exception has been extended to “cultural heritage institutions”, which according to recital 13 of the directive includes mainly libraries, museums and archives<sup>28</sup>. In addition, the possibility to store works for search and mining purposes has been added, which is important as research takes time: it must therefore be possible to get back to the works carrying the data without having to reproduce them every time. Finally, a new exception has been introduced in the body of the directive, according to which “Member States shall provide for an exception or limitation to the rights (...) for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining”<sup>29</sup>.

The objective of the introduction of a new provision was clearly to respond to the criticism addressed to the initial text and to allow Member States to provide for an exception for cases not covered by article 3 of the directive, in particular in order to authorize text and data mining for entities excluded from the scope of article 3. This clearly follows from recital 18, which states that “in addition to their significance in the context of scientific research, text and data mining techniques are widely used both by private and public entities to analyze large amounts of data in different areas of life and for various purposes, including for government services, complex business decisions and the development of new

---

the Digital Single Market - Legal Aspects, Study for the Directorate-General for Internal Policies of the Union, Department of Citizens’ Rights and Constitutional Affairs, European Parliament, February 2018. For a critical evaluation of the directive proposal, see also Christophe Geiger, Giancarlo Frosio & Oleksandr Bulayenko, “Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?”, *IIC* 2018, Vol. 49, No. 7, p. 814, and from the same authors: “The EU Commission’s Proposal to Reform Copyright Limitations: A Good but Far Too Timid Step in the Right Direction”, *EIPR* 2018, Vol. 40, p. 4; European Copyright Society, General Opinion on the EU Copyright Reform Package, 24 January 2017, p. 5; Reto M. Hilty & Heiko Richter, in: Reto M. Hilty & Valeria Moscon (eds.), *Modernisation of the EU Copyright Rules*, Position Statement, MPI for Innovation and Competition Research Paper No. 17-12, p. 25 et sq.; Nicolas Jondet, “L’exception pour le data mining dans le projet de directive sur le droit d’auteur - Pourquoi l’Union européenne doit aller plus loin que les législations des Etats membres”, *Prop. intell.* 2018, No. 67, p. 25.

<sup>27</sup> For a commentary on the final provisions in the directive Geiger, Frosio and Bulayenko, “Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU”, supra, p. 27; P. Bernt Hugenholtz, “Artikelen 3 en 4 DSM-richtlijn: tekst- en datamining”, *AMI* 2019 (5), p. 167; Ducato & Strowel, supra, p. 649; Christophe Alleaume, “Les exceptions en faveur de l’enseignement et de la recherche scientifique”, *Comm. com. électr.* 2019, No. 10, p. 9; Carine Bernault, “Exceptions - Fouille de texte”, *Prop. intell.* 2019, No. 72, p. 39; L. Ballet, “L’exception TDM dans la directive droit d’auteur, une volonté européenne”, *Dalloz IP/IT* 2019/5, p. 304.

<sup>28</sup> The full text of recital 13 is the following: “Cultural heritage institutions should be understood as covering publicly accessible libraries and museums regardless of the type of works or other subject matter that they hold in their permanent collections, as well as archives, film or audio heritage institutions. They should also be understood to include, inter alia, national libraries and national archives, and, as far as their archives and publicly accessible libraries are concerned, educational establishments, research organisations and public sector broadcasting organisations”.

<sup>29</sup> Art. 4.

applications or technologies”.

This new provision would be a major advance if this exception had not been accompanied by an opt-out mechanism, allowing rightholders to expressly reserve the use of works “in an appropriate manner, such as machine-readable means in the case of content made publicly available online”<sup>30</sup>. By conditioning the new exception to the goodwill of rightholders, there is the risk of making this exception impracticable since its effectiveness will depend on the implementation (or not) of the opt-out<sup>31</sup>. Admittedly, the objective of the legislator was to take into account the legitimate interests of rightholders, and in particular that of the producers of databases also covered by the exception, who have an interest in preventing the over-exploitation of their databases which would be impacted by the provision<sup>32</sup>. However, if the intention was to promote research, and in particular to create a legal framework stimulating innovation, it would have been possible to subject text and data mining activities exploited for commercial purposes to a right to remuneration<sup>33</sup>. The uncertainties created by the directive on this point, which will be found again during the transposition phase at the national level<sup>34</sup>, may well lead to a major delay for Europe in the development of new technologies and in the field of artificial intelligence<sup>35</sup>, while elsewhere these activities having high innovation potential are possibly

---

<sup>30</sup> Art. 4 (3). However, it is not clear which are these “appropriate” tools to exercise the opt-out: technological, contractual, or both?

<sup>31</sup> Some works will be available for research, others not. It is in fact very likely that the beneficiaries, generally hostile to this new exception, will systematically use the default opt-out to defeat its implementation.

<sup>32</sup> This is underlined by Binctin, *supra*, p. 7.

<sup>33</sup> See the proposal in this sense Geiger, Frosio & Bulayenko, “Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?”, *supra*, p. 838. The issue of remuneration is particularly discussed in the field of machine learning, which requires the machine to reproduce and store large set of potentially copyrighted work to produce valuable output. As right clearance would bring enormous legal uncertainty due to the important amount of works at stake and thus too high transaction costs, a statutory remuneration right could facilitate the development of machine learning and allow to remunerate creators when appropriate (for example when the AI is producing a derivative work in a commercial context). See in this spirit also Kop, *supra*, at p.7, proposing “the creation of an online one-stop-shop clearinghouse with mandatory or statutory licensing for machine learning training datasets alike a pan-European, multi-territorial collective rights agency”. More generally on statutory remunerations rights as a workable compromise solution in the digital environment, see Christophe Geiger & Oleksandr Bulayenko, “Creating Statutory Remuneration Rights in Copyright law: What Policy Options under the International Framework?”, *CEIPI Research Paper* No. 2020-05, available at SSRN: <https://ssrn.com/abstract=3722108>

<sup>34</sup> See also Séverine Dusollier, “The 2019 Directive on Copyright in the Digital Single Market : Some progress, a few bad choices, and overall a failed ambition”, *Common Market Law Review* 2020, Vol. 57, No. 4, p. 987: “The exception for TDM for non-research purposes is thus rather precarious and is subservient to its prohibition by rightholders”.

<sup>35</sup> See in this sense also Ducato & Strowel, *supra*, p. 649 et sq.: “Neither the initial proposal by the European Commission focusing on the research context, nor the final provisions of the CDSM Directive appear sufficient to facilitate the use of TDM for improved smart disclosure and, more broadly, for AI applications”.

already covered by open norms such as fair use<sup>36</sup> or more flexible exceptions<sup>37</sup>. This goes without mentioning the fact that the European legislator also saw fit to specify that the famous three-step test is applicable to the two newly created exceptions<sup>38</sup>, as well as the cryptic article 6(4) of the 2001 InfoSoc-directive<sup>39</sup>, which - it must be recalled - aimed to resolve conflicts between users and rightholders in the event of the implementation of technical protection measures, yet to this day remains uncertain in its understanding and application<sup>40</sup>. Lastly, even if the scope of exceptions has been widened, many questions pending in the context of the directive proposal remain unsolved in the final text, giving rise to significant legal uncertainty over text and data mining activities, and thus running the risk of damaging the competitiveness of the European Union as a space for research and creativity on the world stage<sup>41</sup>.

---

<sup>36</sup> See in this sense Matthew Sag, “The New Legal Landscape for Text Mining and Machine Learning”, *Journal of the Copyright Society of the USA* 2019, Vol. 66 p. 291; Michael Carroll, “Copyright and the Progress of Science: Why Text and Data Mining Is Lawful”, 53 *U.C. Davis L. Rev.* 893 (2019); Geiger, Frosio & Bulayenko, “Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?”, *supra*, p. 821. As Bernt Hugenholtz rightly puts it (in: “The New Copyright Directive: Text and Data Mining (Articles 3 and 4)”, *supra*), “the TDM provisions of the DSM Directive secure considerably less freedom to text and data mine than they initially appear to do. The opt-out clause of Art. 4, in particular, leaves for-profit miners in the EU at the mercy of the content owners. This puts AI developers, journalists, commercial research labs, and other innovators at a competitive disadvantage in comparison with the United States, where text and data mining is deemed fair use, even if it is done for profit”.

<sup>37</sup> See in this sense Martin Senftleben et al., “Ensuring the Visibility and Accessibility of European Creative Content on the World Market: The Need for Copyright Data Improvement in the Light of New Technologies” (12 Feb. 2021), available at: <https://ssrn.com/abstract=3785272>; Josef Drexl, Reto Hilty et al., “Artificial Intelligence and Intellectual Property Law, Position Statement of the MPI for Innovation and Competition of 9 April 2021 on the Current Debate”, Max Planck Institute for Innovation and Competition Research Paper No. 21-10, at p. 3. Japan for example has introduced in 2019 a new exception comprehensively allowing the use of a work that is aimed at neither enjoying nor causing another person to enjoy the work (art. 30-4 of the Japanese Copyright Act), which includes the use of copyright protected works for text and data mining activities, also for commercial purposes. On this provision, see Tatsuhiro Ueno, “The Flexible Copyright Exception for ‘Non-Enjoyment’ Purposes – Recent Amendment in Japan and Its Implication”, *GRUR Int.* 2021, Vol. 70, p. 145.

<sup>38</sup> On the (still) uncertain interpretation of the three-step test, see i.e. Christophe Geiger, “From Berne to National Law, via the Copyright Directive: The Dangerous Mutations of the Three-Step Test”, *EIPR* 2007, Vol. 29, p. 486.

<sup>39</sup> See Art. 7, paragraph 2 of the directive, which states: “Article 5(5) of Directive 2001/29/EC shall apply to the exceptions and limitations provided for under this Title. The first, third and fifth subparagraphs of Article 6(4) of Directive 2001/29/EC shall apply to Articles 3 to 6 of this Directive”.

<sup>40</sup> On this article, see e.g. Christophe Geiger, “The Answer to the Machine should not be the Machine, Safeguarding the Private Copy Exception in the Digital Environment”, *EIPR* 2008, Vol. 30, p. 121, replying to Charles Clark, “The Answer to the Machine is in the Machine”, in: Hugenholtz (ed.), *The Future of Copyright in a Digital Environment*, *supra*, at p. 139; P. Bernt Hugenholtz, “Copyright, Contract and Code: What Will Remain of the Public Domain”, 26 *Brook. J. Int'l L.* 77 (2000), at p. 83.

<sup>41</sup> See in particular the condition of the lawful source, which has been maintained (Art. 3(1) and Art. 4 (1), as well as the uncertainties concerning the activities of TDM in public/private partnerships, however strongly encouraged by the European Union in its

In sum, if the stated ambition of the Commission is to propose proactive policy which positions the European Union at the forefront of digital innovation and artificial intelligence, it is highly feared that such a task will simply prove impossible if the debate on the scope of the text and data mining exceptions is not also urgently reopened in the very near future<sup>42</sup>. The impulse could come from the World Intellectual Property Organization, which recently opened a major consultation on intellectual property and artificial intelligence, including the topic of text and data mining<sup>43</sup>. In this context, scholars have invited WIPO to take the lead in this area, for example through a legislative proposal aimed at creating a new exception to copyright on the model of the Marrakesh Treaty<sup>44</sup>. More fundamentally, this will only be possible through the rethinking of the underlying and still widely accepted assumption that only strong exclusive rights can serve innovation and creativity, and by acknowledging that limitations and exceptions can equally perform that task.<sup>45</sup>

---

research policy. In this regard, recital 11 of the directive remains rather vague: “While research organisations and cultural heritage institutions should continue to be the beneficiaries of that exception, they should also be able to rely on their private partners for carrying out text and data mining, including by using their technological tools”. Does this mean that there will be an extension of the exception to the private operator in the event of a research contract with a public body? (See in this sense Binctin, *supra*, p. 23). This remains unclear; it is very likely that we will have to wait for a decision by the CJEU to specify the exact contours of the exception. In the meantime, uncertainty will persist, risking further widening the gap between the EU and other regions of the world on digital innovation and research. Criticizing this uncertainty see also Rosanna Ducato & Alain Strowel, “Ensuring Text and Data Mining: Remaining Issues with the EU Copyright Exceptions and Possible Ways Out”, *EIPR* 2021, Vol. 43, p. 322 sq.

<sup>42</sup> It is to be hoped that in the meantime, national legislators will make maximum use of the room for maneuver left by the directive when transposing it. They could for example implement the “opt out” in a manner that safeguards the effectiveness of the exception. For this purpose, Member states could interpret “expressively reserved in appropriate manner” very restrictively (imposing e.g. some formalities to rightholders on the reservation, such as a registration requirement, etc.); making the mere implementation of Technical Protection Measure (TPM) to reserve the work insufficient. Alternatively, the “appropriateness” of the opt-out could include an element of “reasonableness” of the reservation. Finally, a restrictive interpretation of the possibility to opt-out could be mandated by the “effet utile” of the provision and the need to interpret it in the light of the fundamental right to information protected by Art. 10 ECHR.

<sup>43</sup> “WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI)”, May 21, 2020, 2<sup>nd</sup> session, WIPO/IP/AI/2/GE/20/1 REV, paragraph 24, asking this fundamental question: “Should the use of the data subsisting in copyright works without authorization for machine learning constitute an infringement of copyright?” (§24, i).

<sup>44</sup> See Flynn, Geiger, Quintais et al., “Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action”, *supra*.

<sup>45</sup> See e.g. Christophe Geiger, “Copyright as an Access Right, Securing Cultural Participation through the Protection of Creators’ Interests”, in: Rebecca Giblin & Kim G. Weatherall (eds.), *What if we could reimagine copyright?*, Acton: Australian National University (ANU) Press, 2016, at p. 73 et sq.; “Taking the Right to Culture Seriously: Time to Rethink Copyright Law”, in: Ch. Geiger (ed.), *Intellectual Property and Access to Science and Culture: Convergence or Conflict?*, CEIPI/ ICTSD publication series on “Global Perspectives and Challenges for the Intellectual Property System”, No. 3, Geneva/ Strasbourg, 2016, p. 84 (at 90): “Promoting Creativity through Copyright

