2020

Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action

Sean Flynn
American University Washington College of Law, sflynn@wcl.american.edu

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

João Pedro Quintais
University of Amsterdam, j.p.quintais@uva.nl

Thomas Margoni
University of Glasgow, Thomas.Margoni@glasgow.ac.uk

Matthew Sag
Loyola University Chicago School of Law, msag@luc.edu

See next page for additional authors

Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action

Sean Flynn, Christophe Geiger and João Pedro Quintais

with

Thomas Margoni, Matthew Sag, Lucie Guibault, and Michael Carroll[1]

European Intellectual Property Review 2020, Issue 7 (forthcoming)

[1] Sean Flynn is Professorial Lecturer and Director, Program on Information Justice and Intellectual Property, American University Washington College of Law; Christophe Geiger is Professor of law and Director of the Research Department, CEIPI, University of Strasbourg (France); João Pedro Quintais is Postdoctoral researcher and lecturer , Institute for Information Law, University of Amsterdam; Thomas Margoni is Senior Lecturer in Intellectual Property and Internet Law and co-director of CREATe at the School of Law, University of Glasgow; Matthew Sag is Georgia Reithal Professor of Law and Associate Dean for Faculty Research and Development, Loyola University Chicago School of Law; Lucie Guibault is Associate Professor of Law, Associate Dean of Graduate Studies, and Associate Director, Law & Technology Institute, Schulich School of Law, Dalhousie University; Michael Carroll is Professor of Law and Faculty Director, Program on Information Justice and Intellectual Property, American University Washington College of Law. The authors are members of the Global Expert Network on Copyright User Rights, coordinated by the Program on Information Justice and Intellectual Property, Washington College of Law. We thank the signatories of the *Joint Comment to WIPO on Copyright and Artificial Intelligence*, InfoJustice (Feb. 17, 2020), http://infojustice.org/archives/42009, on which this note is based. We thank Margarita Gorospé and Andrés Izquierdo for their excellent research and editorial assistance.

Abstract

Last year, before the onset of a global pandemic highlighted the critical and urgent need for technology-enabled scientific research, the World Intellectual Property Organization (WIPO) launched an inquiry into issues at the intersection of intellectual property (IP) and artificial intelligence (AI). We contributed comments to that inquiry, with a focus on the application of copyright to the use of text and data mining (TDM) technology. This article describes some of the most salient points of our submission and concludes by stressing the need for international leadership on this important topic. WIPO could help fill the current gap on international leadership, including by providing guidance on the diverse mechanisms that countries may use to authorize TDM research and serving as a forum for the adoption of rules permitting cross-border TDM projects.

Copyright law provides protection of the material interests of authors through rights to exclude certain uses of their works, including of their reproduction.[2] At the same time, one of the universally accepted axioms of copyright law is that exclusivity should apply only to original expression, not to facts, ideas, procedures, or methods of operation.[3] It is also universally accepted that copyright contains free spaces to ensure follow-on creativity and to secure important fundamental rights and the public interest, in particular allowing research to be undertaken using protected material.[4] We are far from an international consensus about how to give effect to the boundary between copyright and research rights in the context of text and data mining, machine learning, and artificial intelligence (AI). WIPO could play a vital constructive role in establishing a consistent international baseline that resolves potential tensions between copyright and text and data mining practices. It could also facilitate cross-border text and data mining research and collaboration. We explain these points in further detail below.

---

[2] As we describe below, TDM can involve activities that may implicate other exclusive rights.

[3] See Trade-Related Aspects of Intellectual Property Rights (TRIPS) Agreement, 1994, Pt. II: Standards concerning the availability, scope and use of Intellectual Property Rights, Art. 9(2) *accessed in* https://www.wto.org/english/docs_e/legal_e/27-trips_01_e.htm; see generally Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66 J. of the Copyright Soc'y of the USA (2019). In the EU, the current standard of originality resulting from interpretation of the Court of Justice of the EU in key judgments is that of the "author's own intellectual creation", which expresses his or her "free and creative choices". See, CJEU, decision of 16 July 2009, Case C-5/08, Infopaq, at paras 50-51, and CJEU, decision of 7 March 2013, Case C-145/10, Painer, at paras 94; M.M.M. van Eechoud, Along the Road to Uniformity Diverse Readings of the Court of Justice Judgments on Copyright Work, 3 (2012) JIPITEC, 1, para 60.

[4] See e.g., Ruth Okediji (ed.), Copyright Law in an Age of Exceptions and Limitations, CUP, 2017; 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, 515. In the European context see e.g. with further references Christophe Geiger and Franciska Schönherr, *Defining the Scope of Protection of Copyright in the EU: The Need to Reconsider the Acquis regarding Limitations and Exceptions*, in: T.-E. Synodinou (ed.), Codification of European Copyright Law, Challenges and Perspectives, 133 (Kluwer Law International, 2012); *Limitations to Copyright in the Digital Age*, in: A. Savin and J. Trzaskowski (eds.), Research Handbook on EU Internet Law, 110 (Edward Elgar, 2014).

I. COPYRIGHT AND TEXT AND DATA MINING RESEARCH

This comment focuses on "text and data mining" research, which is often a necessary step to train machine learning systems or processes. We use the term "text and data mining" ("TDM") to describe any application of a computational process to materials to derive data from or about those works.[5] TDM can be used to discover new facts, such as correlations, patterns and links between information points in the database. "Machine learning" applies additional analysis and processes to information often gleaned from TDM to enable machines to dynamically "learn" new tasks for which they were not specifically programmed. The term "artificial intelligence" or "AI" is often used as an umbrella term to describe a number of technologies or systems, including what we define as "machine learning" or an advanced application of it (e.g. deep learning), as well as evolutionary algorithms and rules-based systems.[6]

Many of the most useful TDM and AI projects involve the use of copyright-protected works. The BlueDot project that discovered the novel coronavirus outbreak, for example, analyzed "a variety of information sources, including chomping through 100,000 news reports in 65 languages a day" to recognize patterns between health outbreaks and travel.[7] Other TDM projects are mining scientific publications about the coronavirus family to aid vaccine research.[8] More generally, TDM has the capacity to create new knowledge from new and old data regardless of the field of application. Examples of its varied application include combating disinformation, developing predictive health-care modelling, and constructing multilingual translation tools.[9]

---

[5] The recent *Copyright in the Digital Single Market Directive* defines "text and data mining" as "any automated analytical technique aimed at analyzing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations. See Directive 2019/790 of the European Parliament and of the Council of 17 April 2019, Art. 2(2), on Copyright and Related Rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC [CDSM Directive].

[6] The terminology is not subject to a consensus in the literature. Some refer to "text data mining" to indicate that the data being "mined" is from "text." Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66 J. OF THE COPYRIGHT SOC'Y OF THE USA (2019) (defining "text" as including images, audio visual content, etc. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3331606). Others prefer "data analysis" Jean-Paul Triaille, et al., *Study on the Legal Framing of Text and Data Mining*, 8-9 (2014) ("embracing the word "data" as the broader term that includes text, images, etc., and preferring "analysis" to "mining"). This terminology covers seven distinct forms of computational research. Michael W. Carroll, *Copyright and the Progress of Science: Why Text and Data Mining is Lawful*, 53 U.C. DAVIS L. REV. 893, 899 n.19 (2019). The distinction between "machine learning" and "artificial intelligence" is very indeterminate in the academic literature. See Matthew Sag (2019) (discussing definitions of machine learning and artificial intelligence). See generally, Thomas Margoni, *Text and Data Mining in Intellectual Property Law: Towards an Autonomous classification of Computational Legal Methods*, CREATe working paper [TBC]/2020, forthcoming in Calboli I. & Montagnani L., Handbook on Intellectual Property Research, OUP, 2020; Josef Drexl et al., *Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective*, Max Planck Institute for Innovation & Competition Research Paper No. 19-13 (2019), https://ssrn.com/abstract=3465577.

[7] Mark Prosser, *How AI Helped Predict the Coronavirus Outbreak Before it Happened,* Singularity Hub (2020); Corey Stieg, *How this Canadian Start-Up Spotted Coronavirus Before Everyone Else Knew About it,* Make It: CNBC (2020), (describing how BlueDot discovered the path of a spreading virus by combining various datasets into a machine learning program).

[8] See Will Knight, *Researchers Will Deploy AI to Better Understand Coronavirus*, Wired, (2020), https://www.wired.com/story/researchers-deploy-ai-better-understand-coronavirus/.

[9] The EU H2020 project, OpenMinTeD collects various examples in this sense at http://openminted.eu/blog/ under "TDM Stories". For specific examples see Zalando for an application

Engaging in TDM often entails making both temporary and permanent reproductions of copyrighted works. Temporary reproductions are made any time a researcher makes a query of a database. These copies may be fleeting and, as such, could fall within limitations and exceptions (L&Es) for the making of transitory copies of works for the purpose of facilitating a technological process, which are provided for in many laws.[10] TDM research also requires the making of more permanent copies to construct the database of works to be mined. Additional reproductions and communications of the database itself may be necessary to permit other researchers to use it and to test for accuracy, replicability and transparency. The question for global copyright rules is therefore whether any of these uses of works fall within the exclusive rights of copyright holders, for which a license must be obtained.

Although the enormous scientific and cultural progress that TDM can enable may require merely technical reproductions of copyright-protected works, TDM need not come at the expense of rights holders. These reproductions do not compromise the core interests of exclusive rights, which is to prohibit unauthorized reproductions that can substitute for the work of the author. It could even be argued that these incidental reproductions are outside of the scope of exclusive rights[11]. Also, as has been underlined by several scholars, mere reading does (and should) not involve a copyright relevant action[12], and neither should "the act of reading a work into a computer's random access memory.[13]" More fundamentally, at its heart,

---

of TDM to linguistics: http://openminted.eu/tdm-stories-zalando-links-languages-tdm/. Similarly, see Nanni et al., *Building entity centric event collections*, in, PROC. Of JCDL, 2017 for an example of the application of TDM techniques to web archives in order to assess the impact and diffusion of current events.

[10] See e.g., 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 O.J. (L 167), 5(1); Copyright Act of 1968 (as consolidated in 2015) Section 43A, 43B (Australia) (copyright "is not infringed by making a temporary reproduction of the work or adaptation as part of the technical process of making or receiving a communication"; Copyright Act 2007 (as amended in 2011) (Israel) ("the transient copying, including incidental copying, of a work, is permitted if such is an integral part of a technological process whose only purpose is to enable transmission of a work as between two parties, through a communications network, by an intermediary entity, or to enable any other lawful use of the work, provided the said copy does not have significant economic value in itself"); Copyright Act 1994 (as reprinted in 2016) (New Zealand), Section 43A ("A reproduction of a work does not infringe copyright in the work if the reproduction— (a) is transient or incidental; and (b) is an integral and essential part of a technological process for— (i) making or receiving a communication that does not infringe copyright; or (ii) enabling the lawful use of, or lawful dealing in, the work; and (c) has no independent economic significance.").

[11] Christophe Geiger, Giancarlo Frosio and Oleksandr Bulayenko, *Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?*, 49(7) Int'l Rev. INTELLECTUAL PROP. & COMPETITION L. 814, 817 (2018): ("It could even be argued that this activity is outside the scope of exclusive rights and that any restriction would amount to undermine the underlying rationales of copyright protection and result in an inadmissible restriction of freedom of expression and information as protected by e.g., the European Court of Human Rights (ECHR) and the Charter of Fundamental Rights of the European Union").

[12] See P. Bernt Hugenholtz, *Auteursrecht op informatie*, Kluwer, Deventer, 1989, p. 167, and from the same author: *Fierce Creatures, Copyright Exemptions: Towards Extinction ?*, in IFLA/IMPRIMATUR, *Rights, Limitations and Exceptions : Striking a Proper Balance*, Consensus Forum, Amsterdam Oct. 1997, Forum Report, p. 14 ("Copyright does not (or should not) impede the right of individual to be informed or to receive copyright protected information (freedoms protected, e.g., by article 10. 1 of the European Convention on Human rights). Under existing copyright law, mere acts of information reception or consumption (e.g., reading a book, listening to a concert, watching television) are not restricted acts").

[13] Jessica Litman, *The Exclusive Right to Read,* 13 Cardozo Arts & Ent. L.J. 29, 34-43 (1994). Similarly, in the context of TDM, the "right to read should be the right to mine". See with further

copyright law is concerned with the communication of an author's original expression to the public. TDM and other so-called "non-expressive" uses do not communicate original expression to the public[14]. On the other hand, denying the ability to make reproductions of works needed to undertake TDM would deny access to the very ideas, facts and data at the root of these works, thereby limiting the enjoyment of what we refer to as a "right to research."

II. THE RIGHT TO RESEARCH

Rights to conduct and receive or access research have a strong fundamental rights justification, in particular with regard to freedom of information and the public's right to information.[15] In part to serve these fundamental interests, privileged uses to conduct research with materials protected by IP law are quite common. These protections of research activities can be found in restrictions on the scope of exclusive rights or through the provision of L&Es.[16] In addition to promoting

---

references, Christophe Geiger, Giancarlo Frosio, and Oleksandr Bulayenko, *Crafting a Text and Data Mining Exception for Machine Learning and Big Data in the Digital Single Market*, in: X. Seuba, C. Geiger and J. Pénin (eds.), Intellectual Property and Digital Trade in the Age of Artificial Intelligence and Big Data, CEIPI/ ICTSD Series on "Global Perspectives and Challenges for the Intellectual Property System", Volume 5, Geneva/ Strasbourg, 2018, 95"

[14] For more extensive discussion of the fundamental purpose of copyright to protect against "expressive" uses of works, see Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning*, 66J. of the Copyright Soc'y of the USA, 3, 9-19 (2019) (explaining the doctrinal and normative foundations indicating that "at its heart, copyright law is concerned with the communication of an author's original expression to the public"); https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3331606 (hereinafter, Sag 2019); Matthew Sag, *Copyright and Copy-Reliant Technology*, 103 Northwestern L. Rev. 1067, 1624-1645 (2009) (discussing copyright doctrine and policy that indicate that "that nonexpressive uses of copyrighted works – i.e. acts of copying that do not communicate the author's original expressing to the public – should not generally be regarded as infringing"); Michael W. Carroll, *Copyright and the Progress of Science: Why Text and Data Mining is Lawful*, 53 U.C. DAVIS L. REV. 893, 903-04 (2019) (describing acts of reproduction necessary to typical text and data mining research).

[15] 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 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 Article 5(1) of the German basic law; Article 16(3) of the Federal Swiss Constitution or Art. 11 of the French Declaration of Human rights. The right to information consists in a passive freedom of receiving existing information, and in 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 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 'New Directions in Copyright Law, F. Macmillan (ed.), Vol. 5, 24 (Cheltenham, UK: Edward Elgar, 2007). Making the explicit link between TDM and Freedom of information in the the European context, see Christophe Geiger, *Making Europe fit for the Digital Age? Big Data, Artificial Intelligence, Freedom of Information and the failed Text and Data Mining provisions in EU Copyright law*, Paper presented at the Institute for Information Law, Amsterdam, 10 March 2020 https://www.ivir.nl/ivir-lecture-by-christophe-geiger/. In the EU, a "right to research" could find support also in Article 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."

[16] While we focus in this comment on rights to research with copyright-protected works, there are also research rights in other strands of IP law. For example, rights to "experimental use" of patented subject matter occur in most patent laws. See Richard Gold & Yann Joly, *The Patent System and Research Freedom: A Comparative Study*, The World Intellectual Property Organization (2010) (surveying research exclusions and exceptions in patent law); Whittemore v. Cutter, 29 Fed. Cas. 1120 (C.C.D. Mass.

research through the exclusion of facts from the scope of protection, copyright laws frequently contain limitations and exceptions for uses of protected works for "research" or "private study."[17] Some may refer to such provisions of the law as only creating legal "privileges." We choose to refer to them as "rights" to track the evolving discourse of academics and courts and to signal their strong connection to fundamental rights.[18]

1813) (opinion of Justice Story) ("it could never have been the intention of the legislature to punish a man, who constructed such a machine merely for philosophical experiments, or for the purpose of ascertaining the sufficiency of the machine to produce its described effects"). Similarly, trade secret laws commonly include the right to use "the same technical or commercial information, if they acquired or developed such information independently by themselves through their own R&D, reverse engineering or marketing analysis." a right to reverse engineer products in ways that may reveal protected information. World Intellectual Property Organization, *Frequently Asked Questions: Trade Secrets*, https://www.wipo.int/tradesecrets/en/tradesecrets_faqs.html; see generally Pam Samuelson, *Reverse Engineering Under Siege*, 1 (2002) ("Reverse engineering has always been a lawful way to acquire trade secrets"). In Europe, see also of Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure, arts 3.1(b) and 5(a).

[17] See WIPO Copyright Treaty (preamble) ("Recognizing the need to maintain a balance between the rights of authors and the larger public interest, particularly education, research and access to information"); see e.g., EU Infosoc Directive (2001), Art. 5(3) (permitting members to provide exceptions or limitations for "scientific research"); s 29, Canadian Copyright Act, c. C-42 (authorizing fair dealing for the purpose of "research, private study"); Art. 53(2) first sentence, item 1, Act on Copyright and Related Rights (Copyright Act, as amended up to Law of October 1, 2013) (Germany) (permitting the making copy of a work for "scientific use" ("wissenschaftlichen Gebrauch")); Juan Carlos Monroy Rodríguez, *WIPO Study on the Limitations and Exceptions to Copyright and Related Rights for the Purposes of Educational and Research Activities in Latin America and the Caribbean* (SCCR/19/4), Nov. 30, 2009 (surveying examples of research exceptions in Latin America).

[18] "User rights" is a broader and more useful term than "limitations and exceptions" because such rights to use may be provided as well by exemptions or limitations on the scope of protection e.g. through application of the exclusion of facts and ideas from the scope of protection. See Sean Flynn & Michael Palmedo, *The User Rights Database: Measuring the Impact of Opening Copyright Exceptions*, paper presented at the Fifth Global Congress on IP and the Public Interest (Sept. 2018) (defining "user rights" as "rights to use copyrighted material without the permission of owners"). Framing exceptions and limitations as "rights" also implies their enforceability, in particular with regard to rule-outs in contracts or technical blockages. The CJEU has explicitly recognized that exceptions and limitations in copyright law "do themselves confer rights on the users of works or of other subject matter". Funke Medien [2019], C-469/17, at para. 70 and Spiegel Online [2019], C-516/17, at para. 54. For a comment, see Christophe Geiger and Elena Izyumenko, *The Constitutionalization of Intellectual Property law in the EU and the Funke Medien, Pelham and Spiegel Online decisions of the CJEU: Progress, but still some way to go!*, 51(3), INT'L REV. INTELLECTUAL PROP. & COMPETITION L. 282 (2020). See also CJEU, C-117/13, Ulmer [2014], at [43]; CJEU, C-314/12, UPC Telekabel [2014], at [57]; CJEU, C-201/13, Deckmyn [2014], at [26] ("[T]he exceptions to the rights set out in Articles 2 and 3 of [InfoSoc] directive, which are provided for under Article 5 thereof, seek to achieve a 'fair balance' between, in particular, the rights and interests of authors on the one hand, and the rights of users of protected subject-matter on the other"). User rights in copyright have been recognized by other national courts, perhaps most prominently by the Canadian Supreme Court in *CCH Canadian Ltd v Law Society of Upper Canada* [2004] 1 SCR 339. ("The fair dealing exception is perhaps more properly understood as an integral part of the Copyright Act than simply a defence . . . . The fair dealing exception, like other exceptions in the Copyright Act, is a user's right"). Academic analysis is increasingly adopting the "user rights" vocabulary to describe authorizations to use copyrighted works that may arise from exclusions, exemptions, limitations and exceptions and other aspects of copyright law. See Carys J. Craig, *Globalizing User Rights-Talk: On Copyright Limits and Rhetorical Risks*, Osgoode Legal Studies Research Paper, Am. U. Int'l L. Rev. Vol 33, No. 1 (2017),; Niva Elkin-Koren, *Copyright in a Digital Ecosystem: A User-Rights Approach*, in Copyright Law in an Age of Exceptions and Limitations, 132 (Ruth Okediji ed., 2017); Abraham Drassinower, *Taking User Rights Seriously*, in In the Public Interest: The Future of Canadian Copyright Law, 462 (Michael Geist ed. 2005); Christophe Geiger, *Copyright as an Access Right, Securing Cultural Participation through the Protection of Creators' Interests*, in: R. Giblin and K. G. Weatherall (eds.), *What if we could reimagine copyright?*" Acton, Australian National University (ANU) Press, 2016, 73 (at 94); Christophe Geiger, *Promoting Creativity through Copyright Limitations: Reflections on Concept of Exclusivity in Copyright Law*, 12(3) Vand. J. Ent. & Tech. L. 515, 525 (2010): ("someone who uses a work in a way that a copyright limitation

For some TDM processes -- such as making a query of an existing database – the exclusion of facts and ideas from copyright protection may be sufficient to authorize the activity. But other actions, such as the creation of a database of reproductions for the mining process, appear to require explicit authorization. To provide such authorization, the laws of a growing number of jurisdictions around the world are recognizing L&Es to exclusive rights for "text and data mining,"[19] "information analysis,"[20] "computational analysis,"[21] or similar activities or purposes. In other countries, it may be sufficient to adopt an interpretation or best practices with regard to existing exceptions for "research" or other purposes. As we discuss below, no consensus model has yet emerged.

III. THE CURRENT PATCHWORK OF L&ES FOR TDM

The few countries that have put in place clarifications of copyright to permit TDM have not painted with the same brush[22]. WIPO should develop guidance on the potential ramifications of various restrictions to research that appear in some laws, and highlight best practices in this regard. We describe below some of the most important variations in user rights or L&Es to conduct TDM. Specifically, we compare authorization of TDM through open general exceptions and through specific L&Es in their treatment of the subject matter covered, rights covered, commercial use restrictions, transfer and sharing of data (including cross border), lawful access requirements, and contractual and technical restrictions.

Open general exceptions. Common law countries frequently define "research" or "study" as one of the purposes for which general "fair use" or "fair dealing" exceptions apply.[23] The fairness requirement can be interpreted to be met for any use that does not express works to the public, and therefore does not harm markets for copyright-protected subject matter.[24] One of the key benefits of general fair use and fair dealing exceptions is that they are normally "open" in the sense of being

---

legitimates relies not on a limitation-protected interest, but on the copyright in its negative aspect. He thus relies also on a right, namely copyright as a whole, which materializes as a result of a balancing between exclusivity and the need to keep a creation free of a monopoly"); T. Riis and J. Schovsbo, *User's Rights, Reconstructing Copyright Policy on Utilitarian Grounds*, European Intellectual Property Review (2007) 1; G. Frosio, *Reforming the C-DSM Reform: A User-Based Copyright Theory for Commonplace Creativity*, INT'L REV. INTELLECTUAL PROP. & COMPETITION L. (2020). For connections to fundamental rights, see FN 17, above.

[19] Directive on Copyright in the Digital Single Market, 2019, Arts. 3-4 (European Union).

[20] Japan Copyright Act, 2006, Art. 47-7 (Japan).

[21] Copyright, Designs and Patents Act, 1988, Arts. 29A (U.K.).

[22] For a comparative approach of EU legislations that have implemented specific exceptions, see Christophe Geiger et al., *The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects*, In-Depth Analysis for the Directorate-General for Internal Policies of the Union, Policy Department Citizens Rights and Constitutional Affairs, European Parliament, February 2018 http://www.europarl.europa.eu/RegData/etudes/IDAN/2018/604941/IPOL_IDA(2018)604941_EN.pdf; Centre for International Intellectual Property Studies (CEIPI) Research Paper No. 2018-02, 23 (2018), https://ssrn.com/abstract=3160586.

[23] See e.g., Copyright Act, 2019, Sec. 29 (Canada), *reprinted in* https://laws-lois.justice.gc.ca/eng/acts/c-42/page-9.html#h-103270; see also Sec. 52(1)(a) of Copyright Act, Sec. 52(1)(a) (India) (fair dealing for "private or personal use, including research") http://copyright.gov.in/Documents/CopyrightRules1957.pdf.

[24] See *Authors Guild v. Google, Inc.*, 804 F.3d 202, 215 (2d Cir. 2015); *Authors Guild, Inc. v. HathiTrust* 755 F.3d 87, 105 (2d Cir. 2014).

applicable to the full range of works, exclusive rights, and users regulated by copyright.[25] These should be the models that the WIPO encourages countries to adopt. L&Es that are crafted with conditions other than the requirement that the use be fair to the author -- as mandated by international, regional or national variants of the three-step test[26] -- risk restricting their effectiveness in responding to research needs and priorities.[27]

Subject matter covered. It is essential that research activities can take place with any type of protected work or other subject matter. Literary works may be mined for a wide variety of projects, from science to the humanities. Photographs are often mined to teach computers to recognize different real-world objects. Audiovisual works and broadcasts can be mined to create speech translation tools. Most of the TDM L&Es we have reviewed are open to application to any kind of work/subject matter. An exception is France's current law that only applies to scientific articles, and may need to be changed to comply with the EU Directive on Copyright in the Digital Single Market (CDSM).[28]

Rights covered. Some specific TDM L&Es restrict their applicability to a subset of exclusive rights. The CDSM Directive, for example, requires exceptions only for the right of reproduction in its application to different types of subject matter.[29] An L&E to the right of reproduction is essential to enable the creation of databases for TDM projects, but it is not sufficient to enable the full range of uses of works required.[30] TDM researchers may need to undertake activities that implicate rights to communicate or make available works to the public, for example to share the database with other researchers for collaboration or validation (e.g. through

---

[25] For a description of "openness" of copyright exceptions, see Sean Flynn and Mike Palmedo, *The User Rights Database: Measuring the Impact of Copyright Balance,* InfoJustice working paper (2018), Several scholars have emphasized that such an open clause has several advantages with regard to flexibility to address uses such as TDM and would not be incompatible with an author's right approach, and thus have advocated for its introduction in EU law: see e.g. Bernt Hugenholtz and Martin Senftleben, *Fair Use in Europe. In Search of Flexibilities,* Institute for Information Law Research Paper No. 2012-33 (Amsterdam, November 2011); C. Geiger, *Flexibilising Copyright – Remedies to the Privatisation of Information by Copyright Law*, 39(2) IIC 178 (2008). Christophe Geiger, Giancarlo Frosio and Oleksandr Bulayenko, *The EU Commission's Proposal to Reform Copyright Limitations: A Good but Far too Timid Step in the Right Direction*, 40(1) EIPR 4 (2018); Christophe Geiger and Elena Izyumenko, *Towards a European "Fair Use" Grounded in Freedom of Expression*, Am. U. Int'l L. Rev. 2019, Vol. 35, No. 1, 1.

[26] See WIPO Copyright Treaty, art. 10; TRIPS art. 13; Berne Convention art. 9(2).

[27] See, e.g., Art. 7(2) of the Copyright in the DSM Directive, which subject the text and data mining exceptions in Arts 3 and 4 to the three-step in Article 5(5) of Directive 2001/29/EC.

[28] Id. Digital Republic Act, 2016 (France), *accessed in* (restricting TDM rights to use of "scientific writings").

[29] See e.g., Arts 3 and 4 of the Copyright in the DSM Directive.

[30] For an overview of the exclusive rights triggered by the TDM L&E in EU law, see João Pedro Quintais, *The New Copyright in the Digital Single Market Directive: A Critical Look*, European Intellectual Property Review 2020(1). *See generally,* Christophe Geiger, Giancarlo Frosio, and Oleksandr Bulayenko, *Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?,* International Review of Intellectual Property and Competition Law (IIC) 2018, Vol. 49, No. 7, 814; Christophe Geiger, Giancarlo Frosio, and Oleksandr Bulayenko, *"Crafting a Text and Data Mining Exception for Machine Learning and Big Data in the Digital Single Market"*, in: X. Seuba, C. Geiger and J. Pénin (eds.), Intellectual Property and Digital Trade in the Age of Artificial Intelligence and Big Data, CEIPI/ ICTSD Series on "Global Perspectives and Challenges for the Intellectual Property System", Volume 5, Geneva/ Strasbourg, 2018, 95; Christophe Geiger, Giancarlo Frosio, and Oleksandr Bulayenko, "*Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU*", in: C. Saiz Garcia and R. Evangelio Llorca (eds.), Propiedad intelectual y mercado único digital europe, 27 (Tirant lo blanch, 2019); Rossana Ducato and Alain Strowel, *Limitations to text and Data Mining and Consumer Empowerment: Making the Case for a Right to 'Machine Legibility*', 50(6) IIC Int'l Rev. Intell. Prop. & Competition L. 649 sq (2019).

replicability).[31] Researchers may need to undertake activities that implicate adaptation rights, for example to ensure that materials are in machine readable formats.[32] If "storage" is a separately protected right, then an L&E to enable TDM must include storage rights.[33] As noted in the references above, many of these uses are authorized in some -- but not all -- of the specific copyright L&Es for TDM. A better approach, that should be put forward by WIPO, is to define a fairness test that all research uses -- openly defined -- must comply with.

Commercial use restrictions. It can be tempting, but harmful, to restrict TDM rights to non-commercial uses. Consideration of whether the use in question is connected to a for-profit activity can be justly considered in a fairness analysis. But a blanket restriction on using a L&E by projects with a commercial component can be harmful.[34] The data mining project that discovered the novel coronavirus, for example, was organized by a private company. The copies that produced the Hathi Trust -- "a partnership of academic and research institutions, offering a collection of millions of titles digitized from libraries around the world" for TDM research[35] -- was created through a public-private partnership.[36] Many public interest activities, such as journalism, are conducted almost entirely by commercial actors.[37] Many university and other non-profit research projects -- including the discovery of

---

[31] Some TDM rights anticipate such needs. See Digital Republic Act, Loi Pour Une République Numérique, 2016, Art. 38 (France), *reprinted in* https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000033202746&categorieLien=id (including rights to reproduce a database); Act on Copyright and Related Rights, Urheberrechtsgesetz, 2017, Sec. 60d (Germany), *reprinted in* https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html#p0431 (authorizing "making available" of the database "to a limited circle").

[32] For a law that anticipates such needs, see Copyright Law of Japan, Article 47, *reprinted in* https://www.cric.or.jp/english/clj/doc/20161018_October,2016_Copyright_Law_of_Japan.pdf (including "adaptation" among authorized uses); Thomas Margoni, (2019) Artificial intelligence, machine learning and EU copyright law: who owns AI?, AIDA: Annali italiani del diritto d'autore, della cultura e dello spettacolo, XXVII, pp. 281-304 (explaining need for adaptation right in TDM).

[33] See e.g. Art. 3(2) DSM Directive (allowing secure storage and retention of copies of mined works "for the purposes of scientific research, including for the verification of research results"); See also Art. 60d (3) UhrG (German Copyright Act), including very limited "storage" possibilities since the results of the TDM "corpus" can be sent to some institutions designated by law for long-term storage. Any other copy made should be deleted), Article L.122-5, 10 of the French IP code, which includes storage rights for TDM activities, the modalities of which should be clarified by a decree; 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/?print=print (explaining that storage rights are "important because empirical scientific research generally requires research data to remain available for corroboration purposes.")..

[34] See e.g., Article 3 of DSM Directive, applying to TDM done by "research organizations" and "cultural heritage institutions". N.B. Recital 11 regulates the case of public-private partnerships. Importantly, the question of whether the right to conduct TDM research (or the respective L&E) should be limited to particular types of institutions (i.e. non-commercial actors) is analytically distinct from the question of whether to impose special obligations on large commercial Internet platforms and technology companies.

[35] See HathiTrust Digital Library, https://www.hathitrust.org/ (last visited Mar. 26, 2020).

[36] Aaron Elkiss, *Beyond Google Books: Getting Locally-Digitized Material into HathiTrust*, Perspectives From HathiTrust (2015), https://www.hathitrust.org/blogs/perspectives-from-hathitrust/beyond-google-books-getting-locally-digitized-material-hathitrust.

[37] This is true as well with many public interest AI applications. See, e.g., *AI For Good with Microsoft*, MICROSOFT, https://www.microsoft.com/en-us/ai/ai-for-good (last visited Feb. 6, 2020).

medical treatments and cures -- are later commercialized. WIPO should be guiding countries on the harms associated with blanket exclusions of commercial uses from TDM L&Es.

Transfer and sharing - including across borders. The right to reproduce and transfer a database of materials to another researcher -- including across borders -- is important to ensure the efficiency and efficacy of collaborative research, including for validation purposes. A minority of laws address this issue explicitly. German law authorizes making available a corpus to a "specifically limited circle of persons for their joint scientific research, as well as to individual third persons" for quality assurance.38 No law that we have reviewed explicitly authorizes communicating or making available a database across borders. Imagine a researcher in the EU, where making a TDM database would be lawful under the CDSM Directive, collaborating with a researcher in the U.S., where TDM is also lawful. Can the EU researcher transfer a database lawfully made in the EU to the partner researcher in the US? The answer is unclear (at best), because the respective EU L&E only applies to "reproduction," not to communication or making available rights.

This may be the area where WIPO norm setting is most justified and urgent. The Marrakesh Treaty to Facilitate Access to Published Works for Persons Who Are Blind, Visually Impaired, or Otherwise Print Disabled solved a similar problem in reference to the cross border exchange of accessible materials for the blind and visually impaired.39 WIPO's Standing Committee on Copyright and Related Rights should consider a similar norm for the cross border sharing of lawfully produced research materials, including for TDM.

Lawful access requirements. From the legislative texts we have examined, three of the specific L&Es for TDM research require that the materials used to create a database be "lawfully accessed."40 Other provisions are silent on this matter41. From the perspective of fairness to the rights of the author, the source of the copy should not matter as long as the use does not harm a market. For example, it should not matter if the copies in a closely held research corpus technically violated a

---

38 Act on Copyrights and Related Rights, 2018, Germany *accessed in* https://www.gesetze-im-internet.
de/englisch_urhg/englisch_urhg.html.; For a comparative presentation of several national examples of TDM provisions enacted prior to the EU CDSM directive, see Christophe Geiger et al., *The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects*, Centre For International Intellectual Property Studies (CEIPI) Research Paper No. 2018-02, 23 (2018).

39 Art. 5(1) ("Contracting Parties shall provide that if an accessible format copy is made under a limitation or exception or pursuant to operation of law, that accessible format copy may be distributed or made available by an authorized entity to a beneficiary person or an authorized entity in another Contracting Party.").

40 See e.g., Arts. 3, 4 of the CDSM Directive; UK Copyright, Designs and Patents Act 1988 (as updated), Art 29A. See also in the same spirit Article L122-5, 10 of the CPI, French where the exception for TDM cover acts of reproduction from "lawful sources".

41 For example, in the EU, Germany and Estonia have not included such a condition in their specific TDM exception. For a comparative approach of all TDM limitations and exceptions implemented so far in the EU, see Christophe Geiger, Giancarlo Frosio, and Oleksandr Bulayenko, "*Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU*", in: C. Saiz Garcia and R. Evangelio Llorca (eds.), Propiedad intelectual y mercado único digital europe, 27 (Tirant lo blanch, 2019); CEIPI Research Paper No. 2019-08; available at SSRN: https://ssrn.com/abstract=3470653

publisher license of the library it was drawn from.42 WIPO should study this issue closely and advise countries on whether lawful access requirements are needed to protect rights holders' interests and the effect they could have on needed research activities. Where these requirements are in place and have undesirable effects, WIPO should recommend best practices and interpretations to curtail such effects and enable TDM for research purposes.

Contractual and technical restrictions. Even where copyright permits a use, contractual and technical protection measures can prevent it. Libraries and publishers are currently scrambling to change contracts because they often prevent online learning and digital research that are acutely needed to combat the coronavirus pandemic.43 Some TDM L&Es can be overridden by contract.44 And TDM projects may also be blocked by digital locks that prevent otherwise lawful activity. WIPO should develop guidance on the options available to countries to prevent TDM L&Es from being overridden by contractual and technical means.

IV. THE NEED FOR WIPO ACTION

The coronavirus pandemic has highlighted the need for coordinated global action, including action to liberate digital research technologies to meet the challenge. As the primary international institution for the creation of guidelines and binding norms in the area of copyright, WIPO has a key role in clarifying the path forward toward a world in which research is not hampered by outdated and overly restrictive copyright laws. WIPO can, and should, provide guidance to its members on the diverse mechanisms that countries may use to authorize research, including for TDM research needed in machine learning and AI applications.

---

42 See Michael Carroll, *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful,* 53 U.C. DAVIS L. REV. 893, 951-8 (2019) ("copying from an infringing source necessary for TDM research is still fair use").

43 See University Information Policy Officers, *Vendor Love in the Time of COVID-19* (2020), https://tinyurl.com/vendorsupportedaccess. *Cf.* Matthew Sag, *The New Legal Landscape for Text Mining and Machine Learning,* 66 J. OF THE COPYRIGHT SOC'Y OF THE USA, 3 (2019) (discussing contract, TPM, and cross-border issues); Thomas Margoni & Giulia Dore, *Why We Need a Text and Data Mining Exception (But it is Not Enough)*, Zenodo (2016) (stating that "a TDM exception, not limited to non-commercial purposes . . . should be implemented as soon as possible"); Ian Hargreaves, *Digital Opportunity: A Review of Intellectual Property and Growth,* (2011) (recommending TPM exception for data mining).

44 Article 7(1) of the CDSM Directive, for example, states that "[a]ny contractual provision contrary to the exceptions provided for in Articles 3, . . . shall be unenforceable." This means that the exception for TDM by research and cultural institutions in Article 3 cannot be overriden by contract. Other TDM projects, subject to Art. 4 of the Directive, are subject to reservation by rights holders, including through "machine-readable means in the case of content made publicly available online." *Cf* Hugenholtz (explaining that the "lawful access" requirement in EU law "does not . . . imply that rightholders may contractually rule out text and data mining in their terms of agreement. Article 7 expressly provides that any contractual provision contrary to Article 5 is unenforceable. Note as well that the option to 'opt out' out of the TDM exemption is provided only in respect of the non-research uses governed by Art. 4.").