



Copyright and Text and Data Mining: Is the Current Legislation Sufficient and Adequate?

**Juan-Carlos Fernández-Molina and
Fernando Esteban de la Rosa**

abstract: Text and data mining activities—that is, the automated processing of digital materials to uncover new knowledge—have become more frequent in all areas of scientific research. Because they require a massive use of copyrighted work, there are evident conflicts with copyright legislation. Countries at the forefront of research and development have begun to address this issue. This paper presents the basic aspects of legislation applicable to text and data mining activities. It offers a detailed comparative analysis of the norms of the main jurisdictions that have regulated them to date, highlighting in each case the positive and negative aspects. An adequate knowledge of these laws is not only important for researchers but also important for the academic librarians who provide advice and support in these matters.

Introduction

Technological advances come hand in hand with exciting new research tools for researchers and scholars in all scientific fields. One is software that enables researchers to copy works into digital databases, allowing for computational analyses across aggregated sets of texts, data, sounds, or images. In these cases, researchers do not actually read, display, or share vast amounts of copyrighted material. Instead, they use data analysis techniques to extract factual information about the works and incorporate the findings into their studies or reports. Although this nonconsumptive research has been more common in scientific and technical fields, it has also spread to a wide variety of disciplines, including social sciences and digital humanities. These fields of study take advantage of new technologies to use copyrighted material in all sorts of nonconsumptive research works.

portal: Libraries and the Academy, Vol. 24, No. 3 (2024), pp. 653–672.

Copyright © 2024 by Johns Hopkins University Press, Baltimore, MD 21218.

In any scientific field, research depends increasingly upon data analysis techniques. As the quantity of information grows, so does the need to trust computers and software when analyzing great amounts of informational material that human beings, for reasons of time and resources, are not capable of reading, visualizing, and examining. Data analysis furthermore offers the possibility of discovering new relationships between information and data that were overlooked until now. Text and data mining (hereafter TDM) is the term that most commonly designates the automated processing of digital materials to uncover new knowledge or insights.¹ The materials may include texts, images, sounds, data, or other elements.

Text and data mining (TDM) entails a massive use of works that, for the most part, are copyrighted, producing an evident conflict with copyright legislation. To understand the nature of the conflict between TDM activities and copyright law, it is beneficial to

Text and data mining (TDM) entails a massive use of works that, for the most part, are copyrighted, producing an evident conflict with copyright legislation.

describe in some detail what exactly these activities are, and in what order they take place. We shall follow the scheme of four stages described by Matthew Sag.² In the first place, the researcher must access the works or materials of whatever type to be mined. Secondly, the researcher must copy the materials into a “corpus,” a collection of works prepared for the research. This corpus is often shared with collaborating researchers. Then the mining process begins: analytical processing of the information by means of computers and algorithms that make temporary and technical reproductions of works.

Finally, the researcher disseminates the results, often communicating all or part of the information sources to peers, whether for illustrative purposes or for validation and verification. The copy and mining of works and materials affects the right of reproduction, in that their dissemination and distribution to third parties means communication to the public. In other words, any researcher who carries out TDM activities should bear in mind the copyright laws to avoid problems of legal liability.

The norms regulating TDM activities also have great relevance for information professionals, as made manifest in the statement by the International Federation of Library Associations and Institutions (IFLA).³ The norms are of special interest for academic librarians, who have been in charge of negotiating and licensing TDM access for their users,⁴ and who in recent years have taken on a new role as advisers or supporters in the development of TDM projects,⁵ either individually or by means of the copyright or scholarly communication offices established in many university libraries.⁶ Furthermore, TDM activities are core components of the so-called smart libraries, which employ digital technologies such as artificial intelligence and data analytics to improve services, resources, and operations.⁷ In sum, knowledge of the basic aspects of legislation governing TDM is of interest for academic librarians, both when advising or counseling their users and for the establishment of smart libraries.

This study is intended to underline the basic aspects of legislation applicable to TDM and to serve as a guide or reference for persons involved in research projects or similar initiatives that use TDM techniques. It moreover attempts to shed light on the positive



and negative aspects of the different legislative solutions adopted by the countries that have regulated such matters to date.

Methodological Issues

To attain the objectives of this study, we carried out a comparative analysis of the fundamental legal elements of a sample of jurisdictions, both countries and regions, regulating TDM activities. In deriving the sample, we considered geographic diversity, type of legal system, and which countries pioneered in this type of legislation. Japan was the first country to approach this matter, followed by the United Kingdom and the United States. Then we analyzed the European Union (EU) regulation, for which Estonia, France, and Germany were pioneers, although their legislation now takes a secondary place since the approval of a new copyright directive that affects the EU's 27 member countries.⁸ Separate mention is due for Latin America, as a recent study cites Ecuador as a country whose legislation has a clause of fair use permitting TDM activities.⁹ We agree, however, with Sean Flynn, Luca Schirru, Michael Palmedo, and Andrés Izquierdo that this reading of Ecuadorean law is not quite accurate, and interpretative errors stem from the confusing wording of a new provision.¹⁰ It introduces a fair use clause that acts as a "safe harbor" for libraries, freeing them from responsibility for the action of their users.¹¹ Along these lines, most experts come to the conclusion that, unfortunately, no Latin American country allows TDM activities.¹²

For our comparative analysis, a series of aspects were defined and used as the framework to examine each norm. As the starting point, we adopted the three stages (access, copying, mining) defined by Eleonora Rosati to explain TDM activities.¹³ These stages were complemented by a fourth stage, disseminating, added by Matthew Sag.¹⁴ To make the framework as complete as possible, we also used aspects and questions that, in the opinions of Daniel Gervais and of Sean Flynn and his coauthors, must be taken into account.¹⁵ The first aspect is the subject matter covered, that is, the type of work and material affected. Afterward, the prerequisites for access to this material are analyzed, along with the types of institutions and individuals who may benefit from the norms in favor of TDM activities. Then, because a number of rights are conceded to authors through copyright legislation, this study analyzes which are affected by the norms. It is likewise relevant to know if some specific end purpose is required for the activities to be permitted, especially if there are limitations for commercial aims. Once the materials have been copied for use in TDM activities, one must consider which type of operations would be allowed for the materials, and if it is possible to conserve them for other uses or if the copies made must be destroyed. Research activities tend to be carried out by teams of researchers who may belong to different institutions and countries, making it key to examine whether transfer of the corpus copied or the results of the process can be shared with other colleagues. Lastly, it is necessary to analyze the impact of private ordering measures (contractual and technical restrictions) on the possibilities of legally developing TDM activities.



Japan

Japan stands out as the first country that modified its copyright law (in 2009) to include an open exception that would facilitate TDM activities . . .

Japan stands out as the first country that modified its copyright law (in 2009) to include an open exception that would facilitate TDM activities (Article 47-7 before the enactment of the 2018 Amendment). Shortly thereafter, it became clear that the law was insufficient

to cover the advance of technology. Moreover, it had certain ambiguous aspects that could imply legal uncertainty.¹⁶ As a result, in 2018 the law was reformed to introduce a number of flexible copyright exceptions, among them Article 30-4.¹⁷ This article (“Exploitation without the purpose of enjoying the thoughts or sentiments expressed in a work”) introduces a general rule that permits making use of any work as long as it is done for “non-enjoyment purposes” of the expressive elements of the work, and its use does not unreasonably prejudice the

interests of the copyright owner.¹⁸ To clarify the difference between “enjoyment” and “non-enjoyment,” the law lists three examples: (1) for use in testing to develop or put into use technology; (2) for use in data analysis; and (3) being exploited in a way that does not involve what is expressed in the work as perceived by the human senses.

This norm is open and flexible, hardly imposing restrictions. Access to the work need not fulfill any condition. The law applies to all types of materials, and no restrictive conditions are imposed on who may become the beneficiary—any person or institution, lucrative or not. Nor are there limitations regarding the rights of public communication and reproduction, so that the materials can be copied, stored for posterior use, or even communicated to third parties. According to Tatsuhiko Ueno, if a work is acquired by means of a contract—even if the act of copying it is not authorized—it does not infringe copyright.¹⁹

This broad rule favoring TDM activities is justified by the notion that no infringement of copyright occurs if using the work is not intended for enjoyment. A comparable conception underlies copyright, implying that infringement occurs only with “the use of the work as a work.”²⁰

The United Kingdom

Another leader in the regulation of TDM activities was the United Kingdom (UK). In 2014, it was still part of the European Union, a key consideration. The EU directive of 2019—specifically regulating TDM activities—had not yet been ratified. In addition, EU membership called for following the 2001 directive, which did not allow for new copyright exceptions.²¹ Hence, the law was necessarily based on existing exceptions. The new Article 29A of the UK’s Copyright, Designs and Patents Act therefore adopts as a starting point Article 5.3(a) of the directive of 2001. This article is dedicated to the exceptions for the rights of reproduction and of communication to the public when the sole purpose is illustration for teaching or scientific research.²² The new article, “Copies for text and data analysis for noncommercial research,” permits any person to make a



copy of a given work to carry out “a computational analysis of anything recorded in the work.” This authorization of TDM activities is subject to many restrictions. It is not clear whether the restrictions were included because the British lawmakers believed them beneficial or because the limitations defined at that time by the copyright directive required them. As we shall soon see, both reasons appear to have had an influence.

Since no clear limitations governed the type of work affected, any informational material could enjoy the exception for computational analysis. There are, moreover, no restrictions as to the type of person who might benefit from the exception, so that apparently anyone could. In practice, however, the exception is limited to persons tied to the academic or research realm in a noncommercial context. There is an explicit restriction regarding access, establishing that only persons with lawful access can use the works or materials. The UK Intellectual Property Office, on their list of frequently asked questions (FAQs), explains lawful access as “where researchers have the legal right to access a copyright work to read it,” including “paying for a subscription” to a journal or database, or material published under “open licenses.”²³ It therefore includes both access through payment by means of licensing agreements and free access to material with Creative Commons (CC)-type licenses. One obvious consequence of this restriction is that the budgeting capacity of the institution or researcher in question might limit the TDM activities, to the point of making TDM impossible if the user could not afford to pay for lawful access to the copyrighted material.

Likewise significant is the restriction about the purposes of copying, which is only permitted when the sole aim of research is noncommercial. In essence, this means just in a research context, never for profit. As mentioned earlier, this exception in favor of TDM activities is based on Article 5.3(a) of the 2001 directive, the scope being limited to teaching and research. This restriction stands as a strong obstacle, impeding numerous TDM activities involving other realms.

Once the copies of the works or materials have been made, can they be stored for future use? Given that Article 5.3(a) does not address this point, one might surmise that no restrictions govern storage and ulterior conservation. But there are restrictions as to communication to third parties. Article 29A(2)(a) expressly prohibits transfer of a copy to any other person, except when authorized by the copyright owner. Equally clear is the explanation given by the UK Intellectual Property Office: copies made for TDM activities “can’t be shared, sold, or made publically available in any way.” While Article 5.3(a) of the directive permits exceptions both for reproduction and for communication to the public, the British lawmakers decided to expressly exclude the option of public communication—an important limitation for the practical application of the article. It is common practice for researchers to work in teams with related institutions or even other countries, and sharing information is essential for their endeavors.

. . . the budgeting capacity of the institution or researcher in question might limit the TDM activities, to the point of making TDM impossible if the user could not afford to pay for lawful access to the copyrighted material.

Finally, it is necessary to analyze the impact of private ordering measures on the possibilities of developing TDM activities legally. As with other copyright exceptions, the reform of the 2014 law affirms that such restrictions may not be annulled by contract. Article 29A(5) states that “to the extent that a term of a contract purports to prevent or restrict the making of a copy which, by virtue of this section, would not infringe copyright, that term is unenforceable.” This wording, protecting copyright exceptions from override by contract, deserves very positive appraisal. Technological protection may, however, interfere with the enjoyment of this exception. Again, the FAQs of the UK Intellectual Property Office indicate that publishers can apply technological measures on networks for purposes such as security or stability. Although it is then acknowledged that such measures “should not stop or unreasonably restrict any researcher’s ability to benefit from the exception,” the law does not permit circumvention of these measures to enjoy copyright exceptions. As a result, it could prevent researchers from taking advantage of the exception in favor of TDM activities.

Even though the pioneer British law allowed TDM activities under specific circumstances, it is hampered by insufficiencies and restrictions. Some result from their

Even though the pioneer British law allowed TDM activities under specific circumstances, it is hampered by insufficiencies and restrictions.

foundation on Article 5.3(a) of the directive of 2001. Others, and in particular the prohibition of communicating to other persons the copied materials, are exclusively the responsibility of the British lawmakers.²⁴ Efforts are underway to modify and adapt the law. Alina Trapova and João Pedro Quintais stress that the UK Intellectual Property Office has initiated public consultation about the intersection between artificial intelligence and intellectual property laws, which should serve as a basis for legislative reform.²⁵

The United States

Unlike the European Union and other countries, the United States has never established a specific exception for TDM activities. Instead, their inclusion in copyright law is analyzed

Unlike the European Union and other countries, the United States has never established a specific exception for TDM activities.

from the perspective of the doctrine of fair use (Section 107 of the Copyright Act of 1976). The flexibility of this framework has made it possible for this doctrine to be applied (often) to adapt new technologies to copyright, especially when certain uses of protected works were not originally foreseen—the case of text and data mining. The doctrine in question is based on a joint application of four factors: (1) the purpose and character of the use, including whether use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work.



One fundamental aspect underlying interpretation of the first factor, and probably the most important element in the recent application of fair use, is the notion of transformative use. A new work is understood to be transformative (therefore implying fair use) when the original is used as “raw material, transformed in the creation of new information, new aesthetics, new insights and understandings.”²⁶ While considerations regarding fair use take further factors into account, such as the impact of its use on the potential market for the protected work, nowadays the transformative character weighs the most. Thus, the more transformative a new work is, the less the importance of the rest of the factors—including use of a commercial nature—that could be a counterweight to fair use.²⁷ Closely related to this idea, and relevant for our study, is the doctrine of “non-expressive use,” arising from the case *Sega Enterprises Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992). Accordingly, any act of reproduction that is not destined to produce human enjoyment, appreciation, or comprehension of the copied expression would be an act of reproduction that does not communicate the original expression of the author to the public and so would not conflict with the author’s copyright.²⁸

A landmark case concerned with the legality of TDM activities is *Authors Guild, Inc. v. Google Inc.*, 954 F. Supp. 2d 282 (S.D.N.Y. 2013), renowned in the context of the Google Books project. The case entails both transformative and non-expressive uses. The project revolved around a massive digitalization, not authorized by the rights holders, of books protected by copyright. The aim was to create a function for the location and visualization of fragments of the works in Google’s search engines. The court concluded that the copying of books by Google constituted fair use, given that it implied a highly transformative use of the original works by means of the search engine created. Google’s intention was to make available to its users significant information about the books copied, facilitating the search for works containing a specific word of interest. What is more, the function “ngrams” permitted readers to know the frequency of use in the aggregate corpus of books published for different historic periods. These two functions proved sufficiently transformative to claim fair use of those works protected by copyright. As for the fourth fair use factor, the court determined that Google Books does not replace the books because it is not a tool for reading books. The court applied the reasoning of non-expressive use to delimit the effect of use in the potential marketplace of the protected work. The ruling came to acknowledge that Google Books could be detrimental for the market value of a work if, for instance, a user was interested in obtaining information about a historic event but, upon viewing a fragment through Google Books where this information appeared, the user lost interest in buying the book in question. Yet since intellectual property does not protect facts or ideas, the court concluded that the eventual effect on sales would be irrelevant for copyright. This case stands as a strong support for TDM techniques, essentially underlining their potential for innovation and progress.²⁹ For this reason, Sag holds that in the wake of this case, there remains no doubt that TDM is a use permitted by U.S. legislation.³⁰

Closely connected to this court decision is another precedent for TDM activities: the *Authors Guild, Inc. v. HathiTrust* case, 755 F.3d 87 (2d Cir. 2014). The court allowed the libraries involved in the HathiTrust Project to digitalize works protected by copyright for the purpose of allowing full-text searches. It also authorized the participating libraries to provide disabled users access to the full-text works and allowed the libraries

to preserve the books protected by copyright in a digital format. The ruling concluded that the creation of a full-text database is a transformative use: when a word search is undertaken, the result is different in the intention, character, expression, meaning, and message from the page (and from the book) from which it is extracted. Therefore, there is scant resemblance between the original work and the full-text results of the HathiTrust Digital Library.³¹

Both cases were decided in appeal by the Second Circuit Federal Court and make manifest that the copy of protected works may be an intermediate step to access facts or ideas, elements of a work not protected by copyright, or to achieve a transformative use of the original works. The doctrine of fair use appears to cover such uses. As James Grimmelmann noted, "Copyright has concluded that reading by robots doesn't count."³² We should also point out that both cases show how noncommercial and commercial uses are permitted: HathiTrust is not a lucrative entity, whereas Google is a commercial, for-profit corporation.

We might conclude that the doctrine of fair use broadly permits TDM activities. There are no limits for the works or materials involved, and lawful access to the works is not required. The beneficiary may be an individual or institution, dedicated to research or not. The doctrine affects all the rights, including reproduction, communication to

... the doctrine of fair use broadly permits TDM activities. There are no limits for the works or materials involved, and lawful access to the works is not required.

the public, and distribution. No not-for-profit end purpose is demanded, and the conservation of the corpus and its transmission or sharing with third parties are allowed.

This doctrine may be overridden, nonetheless, by contractual and technical restrictions. Since the 1996 international treaty of the World Intellectual Property Organization (WIPO) aimed to update copyright norms in the digital setting,³³ the laws of most countries prohibit circumventing technological protection measures that would control the use of copyrighted works.³⁴ As this prohibition

could endanger the enjoyment of copyright exceptions, national laws establish certain conditions to it. Thus, under particular circumstances users are allowed to circumvent the technological measures to take advantage of the copyright exceptions to which they are entitled. In the United States, a periodic procedure was set (every three years) for determining whether the circumvention prohibition was exerting or would likely have an adverse effect on users' ability to make non-infringing uses of certain classes of copyrighted works. The exceptions approved become valid for the following three years. In the latest proceeding of this nature, in 2021, TDM activities were included.³⁵

While the contents can be positively appraised, as they allow one to elude technological protection when engaging in TDM activities, the measure imposes some restrictions. The first is that it can only be by a "researcher affiliated with a nonprofit institution of higher education, or by a student or information technology staff member of the institution at the direction of such researcher," and solely "for the purpose of scholarly research and teaching." We thus run into restrictions about who can carry out this activity (for one, not a corporation) and for what purposes. Another restriction imposed is that a copy of



the work must be “lawfully acquired and owned by the institution, or licensed to the institution without a time limitation on access.” Once again, we encounter the requisite of lawful access to the work. Finally, the rule establishes that the institution should apply effective security measures to prevent further dissemination or downloading of works and to limit access to only the authorized persons (the affiliated researcher) or to researchers affiliated with other institutions of higher education solely for purposes of collaboration or replication of the research.

The main challenge for the development of TDM activities in the United States lies in contractual protection. Indeed, the impact of the principle of “freedom of contract” implies that if the work is made available to the public by means of a contract, whatever is stipulated in that contract must be fulfilled, even if this means impeding enjoyment of the rights granted by law.³⁶ Studies show that the licenses for digital resources such as journals and ebooks rarely permit TDM activities.³⁷ In short, while overall appraisal of the U.S. situation is positive, there are important restrictions if the work is protected technologically, and above all if it is protected by contract, given the possibilities for the contract to annul the users’ rights.

The European Union

In the European Union, this problem also emerged early on, given its impact for research and innovation. A vague solution was sought in the legislation already in place, particularly through the concept of exception for transitory reproduction (Article 5.1) of the copyright directive of 2001. It was not correctly adjusted to what was required at that time, however, leaving questionable the fulfillment of the two main requisites: just facilitating transmission in a network and the absence of independent economic significance. Furthermore, this exception covers only the copies that are automatically erased, whereas TDM calls for conservation of the copy to verify in the future the validity of the conclusions reached.³⁸ An attempt to face the problem with the use of licenses was also made, with the explicit backing of the European Commission,³⁹ but in effect the conditions were diverse and the costs of transaction were multiplied. It became evident that a specific exception was needed to allow for these activities. Their inclusion among the economic rights of the authors was an undesired effect of regulation (which, when approved, could not foresee today’s level of technological development, including TDM). A number of European countries dealt with this issue individually (the UK, Estonia, France, and Germany) through legislative solutions that varied considerably. Given the situation, it was deemed that exceptions to this regard be included in the directive of 2019.

After looking over several options, the proposal of the directive presented by the European Commission foresaw only the regulation of text and data mining by research organizations for the purposes of scientific investigation.⁴⁰ The proposal received considerable criticism for leaving out TDM activities carried out by other subjects and for other purposes.⁴¹ Consequently, the final version established a double regulation: Article 3, dedicated to text and data mining for the purposes of scientific research, and Article 4, dedicated to the cases not included in the previous precept.



Article 3

Article 3 is full of conditions and restrictions. But they are not established for the type of works affected, taking in “works or other subject matter.” The article does restrict the type of works in terms of access, which should be lawful. Lawful access includes (1) public open access licenses (CC or similar); (2) licensing agreements (subscriptions to journals, ebooks, or databases); (3) access to free contents available online, for example, website contents or social network platforms; and (4) access produced without consent of the rights holder but on the basis of a copyright exception (for example, public lending).⁴²

Strict conditions are also set out regarding who might be the beneficiaries of this exception. It applies only to research organizations and cultural heritage institutions. Among other implications, this means that only legal entities are beneficiaries, leaving out individual researchers not integrated into an institution. As criticized by Roberto Caso, this restriction acts against the promotion of “citizen science” by the European Union.⁴³

The conditions do not end there. Research organizations must have as their primary goal to conduct scientific research or to carry out educational activities also involving the conduct of scientific research. This leaves out other entities whose research activities are not strictly scientific, for example, investigations by journalists.⁴⁴ It is also essential that the entity fit into one of

... this restriction acts against the promotion of “citizen science” by the European Union.

the three premises identified by the directive: (1) that the research be not-for-profit; (2) that even in the case of for-profit endeavors, all the profits be invested in scientific research; and (3) that entities develop their research activity “pursuant to a public interest mission recognized by a Member State,” which includes private entities with commercial interests that carry out research activities in collaboration with public organizations. Article 2.3 defines a cultural heritage institution as “a publicly accessible library or museum, an archive or a film or audio heritage institution.” There is no requirement that the institution be not-for-profit or that businesses participate. TDM activities need not be carried out directly by the benefiting entities, as third parties might also be subcontracted to participate, even for-profit firms, as made explicit in Recital 11 of the directive.

The rights harbored are only those of reproduction (allowing the reproductions of works or other subject matter), omitting the right of communication to the public. This has important consequences. It may be argued that the information obtained with mining is different from that included in the materials used, hence communicating the information to the public does not entail communication of the initial works or materials. Nevertheless, there are scenarios in which a text used for mining is shared with third parties—for instance, to determine a paper’s suitability for publication by peer review, to verify the results obtained, or simply to share the corpus with collaborating researchers. German law previous to the 2019 directive foresees this possibility and expressly establishes that the corpus obtained may be communicated to a limited circle of persons for joint scientific research, or to permit third persons to evaluate the quality of the scientific research involved.⁴⁵



The directive expressly permits storage of the copies of works and materials under the condition that the setting has an adequate level of security. This possibility of conservation is intended to allow repeated access to the copies for posterior verification, in addition to facilitating new use of the copies in later research efforts. According to Recital 15 of the directive, the member states can approve specific dispositions for the conservation of copies, for example, the designation of trusted bodies to do so. This stipulation was meant to address publishers' fears that parallel or "shadow" libraries might be created. Such national measures should be adopted after discussion with

Article 3 itself establishes that the member states shall encourage rights holders and beneficiaries of the exception to cooperate so as to conform the best practices with regard to the storage and conservation of copies.

relevant stakeholders. Further, to not unduly restrict the application of the exception, the arrangements should be proportionate and limited to what is needed for retaining the copies in a safe manner and preventing their unauthorized use. Article 3 itself establishes that the member states shall encourage rights holders and beneficiaries of the exception to cooperate so as to conform the best practices with regard to the storage and conservation of copies.

As mentioned in our analysis of the beneficiaries of this exception, the purposes of TDM activities are fully relevant. It does not suffice that the beneficiary be a research organization or cultural heritage institution. Additionally, the TDM activities must be carried out specifically with that finality. The directive does not demand that the research have no commercial purpose, but all profits derived should be reinvested by the benefiting institution into the research activity per se.

No analysis of Article 3 would be complete without examining the measures of protection based on private ordering: contractual and technological protection. Aware that contracts may eliminate de facto the enjoyment of copyright exceptions, the European lawmakers establish in Article 7.1 of the directive that "any contractual provision contrary to the exceptions provided for in Articles 3, 5 and 6 shall be unenforceable." Without such a clause, the rights holders would be rendered helpless in any attempt to legally facilitate TDM activities for research purposes.

For technological protection, however, the case is otherwise. The directive permits the adoption of measures to guarantee the security and integrity of networks and databases. These measures are intended to impede unwarranted text and data mining, the emptying of databases, or undue use of works or loans. The problem is that these technological measures can operate against the best interests of researchers by complicating the performance of TDM activities. They can mean a greater investment in time, or even a full-force obstacle. For this reason, the article states that "such measures shall not go beyond what is necessary to achieve that objective." The regulation of the technological protection is complemented by Article 7.2 of the directive, which states that certain provisions can be applied in view of Article 6.4 of the 2001 directive, intended to impede an interference of technological measures in the enjoyment of exceptions.

Hence, although the technological measures are protected against circumvention, the member states were expected to take steps so that the rights holders might facilitate the exception for beneficiaries' sake.

Article 6.4 proved ineffective, however. It obliges rights holders to provide the means to exert an exception, without granting users any authority to carry out on their own acts to elude technological measures.⁴⁶ The fact that technological protection impedes the enjoyment of copyright exceptions constitutes no justification for trespassing technological measures. It merely permits the request that they be removed. In no case can access achieved by circumventing technological protection be considered lawful. The directive could have made clear that exceptions to mining are protected against their annulment by technological protection. Such a statement would definitely affirm that this type of protective measure cannot impede enjoyment of the new obligatory exceptions, and that legitimate and effective measures to guarantee their suppression should be available to users. The declarations in Recital 16 can be seen as a weak attempt along these lines, not effective in practice.⁴⁷

Article 4

As a complement to TDM for research purposes, Article 4 of the directive addresses TDM for any other reasons, including commercial ones. Unlike the previous article, this one sets few limitations, while maintaining the requisite that access be lawfully achieved. At any rate, the fundamental characteristic of this article is that the extent of mining is limited by its application only where the use of works or other subject matter has not been expressly reserved by rights holders. This reservation of rights drastically reduces

The advantage of being able to carry out TDM activities even for commercial purposes is practically invalidated by the possibility that the rights holder can refuse to provide legal access to a work or can concede access only under certain conditions.

the number of potential beneficiaries. It leaves journalists, artificial intelligence (AI) developers, and commercial research labs at a clear competitive disadvantage in comparison with similar institutions in the United States, Japan, Israel, or Singapore, where TDM activities are allowed even when for profit.⁴⁸

The opt-out mechanism should be practiced "in an appropriate manner," for example, by machine-readable methods, in the case of content publicly available online. It would be helpful for the member states to give some guidance on

the means considered inappropriate for publicly available online contents. Moreover, the reservations of rights made through contractual agreements or unilateral declarations should be explicit, transparent, precise, easily identifiable, and expressed in unambiguous terms.⁴⁹ Finally, we should note that Article 4 allows for storage, as long as needed, of the information obtained for the purposes of text and data mining.

Global assessment of this article can hardly be positive. The advantage of being able to carry out TDM activities even for commercial purposes is practically invalidated by



the possibility that the rights holder can refuse to provide legal access to a work or can concede access only under certain conditions. In effect, a norm conceived to protect the general interest is left in the hands of private interests.

Cross-Border TDM Activities

As we have seen, worldwide legal systems presently adopt different approaches to TDM activities. In many cases, the issues traverse geographic borders. For example, the researchers of one country might collaborate with those of another country (or countries) and share the initial corpus or end results of TDM. The data could be extracted from databases rooted in another country. Physical books might be bought abroad or ebooks acquired from another country. The result is a sense of insecurity among researchers. They are uncertain about the legality of their activities, because the answers put forth by the different national laws do not always coincide.

Therefore, until a uniform international regulation prevails, the only way to find out if a TDM activity is legal is to carefully consult the applicable law.

Although there is no unified international regulation to determine the law applicable to intellectual property rights, the trend in national systems dictates that the law applicable to infractions of copyright is the *lex loci protectionis*, or the law of the country where the call for protection is invoked. It is true that some systems point toward the law of the state of origin.⁵⁰ However, this law is more pertinent in cases of copyright ownership than in regard to the content of copyright and the determination of alleged infringement. For instance, according to Article 8 Regulation 864/2007 (Rome II) on the law applicable to noncontractual obligations, all authorities in Europe are required to apply the law of the country for which protection is claimed (*lex loci protectionis*) to determine the applicable regime to noncontractual obligation arising from an infringement of an intellectual property right.⁵¹ In attention to the public interest at stake, it is not allowed for parties to modify this regime by a choice of law agreement.

The application of this rule leads in most cases to the application of the law of the country in which the infringement takes place. For instance, the criteria used by the conflict of laws rule leads to the law of the country where a physical book is scanned. To determine the applicable law, it does not matter in which country the book was originally published. TDM is deemed legal if the country where copies are made permits such TDM activities, regardless of a text's national origin. Any doubts about the legality of cross-border collaboration must be resolved by verifying whether the activities carried out (access to content, extraction or copying of content, preprocessing of texts and data, extraction of structured data, or publication of results, including copy of original texts) enjoy a legal

. . . until a uniform international regulation prevails, the only way to find out if a TDM activity is legal is to carefully consult the applicable law.

TDM is deemed legal if the country where copies are made permits such TDM activities, regardless of a text's national origin.

exception—even if such tasks take place through cloud services. The legality of sharing and communicating results of TDM depends on the law of the country where the data are utilized and where data mining takes place. The activity may also be prosecuted in a different country where the results are published if an infringement of copyright occurs. The information obtained with TDM is not identical to that included in the materials used: therefore, its communication or provision does not imply communication of the original works. Legally sharing research materials calls for attention to the provisions of the law of the country where the communication is destined.

When data access takes place under the terms of a licensing agreement, one must first consult the national law applicable to the contract. It is also necessary to check the terms of the licensing agreement to ascertain whether the TDM activity is permitted. Thus, for e-books and texts being accessed under a licensing agreement, it is essential to know the terms, no matter the origin of the text at hand. If the license does not authorize TDM activities, it is illegal in the United States (in principle) to download the files and use them for TDM purposes. Yet in such cases, it is necessary to pay attention to legal provisions regarding the inapplicability of those terms. For instance, when TDM activities are undertaken in a country belonging to the European Union, Article 7.1 of Directive 2019/790 envisages that any contractual provision contrary to the legal exceptions provided for in Articles 3, 5, and 6 shall be unenforceable. Mandatory rules within the copyright act have to be considered as overriding provisions in the sense of Article 9 Regulation 593/2008 (Rome I) on contractual obligations.⁵² The respect of overriding mandatory rules is regarded as crucial to safeguard a country's public interests—for example, political, social, and economic organizations—to the extent that they may govern any situation falling within their scope, irrespective of laws otherwise applicable to the contract under the regulation.

Conclusions

An adequate regulation of TDM activities should grant that researchers carry out their projects using open access materials as well as subscription-based contents. It should allow for commercial purposes in addition to noncommercial ones, not exclusively for scientific research intentions. Likewise, it should allow for the storage of a corpus or copies generated by TDM, along with the option of sharing and comparing the results of a project with third parties, not just a private circle of contacts. It should also facilitate the performance of TDM activities by all types of users, not just researchers or professors, and whether for profit or not. Lastly, this exception should be protected against its annulment by contract or technological measures.

If we compare this “ideal” model with the laws enumerated and analyzed in this paper, we conclude that the Japanese legislation is the most open and thus closest to the ideal. Also deserving positive appraisal is the U.S. doctrine of fair use, complemented with the possibility of circumventing technological measures of protection under determined circumstances. It loses effectiveness, however, because of the prevalence of contracts involving copyright exceptions. As Michael Carroll warns us, the United States is losing its competitive edge because researchers cannot access the full text of scientific journals without undersigning license agreements that limit the benefits of



copyright law's flexible aspects.⁵³ In turn, while the law of the UK stands as a pioneer attempt to settle this issue, the end result is insufficient. Its effectiveness is hindered by limitations set by the EU legislation at that time (no longer applicable) and restrictions imposed by its own initiatives. We will have to wait to see if an upcoming reform makes proper adjustments to the current needs. The European directive has some positive aspects as well, for instance, being an obligatory exception for all the member states and the impossibility of its annulment

by contract. Still, it imposes too many restrictions, which significantly diminish its utility in the TDM setting. Its global appraisal does not increase with the contributions of its Article 4; while it allows mining even in the context of work-for-profit, the fact that it can be annulled by a reservation of rights on the part of the rights holder vastly diminishes its efficacy.

It is clear that this problem cannot be approached exclusively from a national or regional perspective. An international focus is necessary to facilitate that TDM activities be developed properly across borders. In this sense, WIPO has launched a public survey about AI in which TDM activities are expressly addressed.⁵⁴ It would be helpful if this organization took on a leading role in the matter. Ideally, an international agreement would establish a baseline for regulating TDM activities at the worldwide level, facilitating the development of cross-border activities. If not, the harmful disparities in our global research system will be aggravated. As underlined by Sean Fiil-Flynn and his coauthors, it is no coincidence that the most open regimes for TDM research correspond to some of the richest countries worldwide, whereas the poorer ones have more restrictive systems.⁵⁵

Amid the search for an international solution, one must not underestimate the notion of "flexible copyright" to keep pace with technological evolution, avoiding market inefficiencies owing to a lack of consonance between legislation and technological realities.⁵⁶ The call for flexibility is rooted in the very justification of copyright in countries having common law and those founded on civil law.⁵⁷ In the former, the justification lies in "encouragement of learning" (under the Statute of Anne) or to "promote the progress of science" (under the U.S. Constitution)—no doubt better achieved by allowing TDM activities than by restricting them. In civil law countries, which lend great importance to moral rights, TDM activities do not affect the essence of the work as created by the author, nor do they undermine the author's reputation.

... the United States is losing its competitive edge because researchers cannot access the full text of scientific journals without undersigning license agreements that limit the benefits of copyright law's flexible aspects.

... the most open regimes for TDM research correspond to some of the richest countries worldwide, whereas the poorer ones have more restrictive systems.



Juan-Carlos Fernández-Molina is a professor in the Department of Information and Communication Studies at the University of Granada in Granada, Spain; he may be reached by email at: jcfernand@ugr.es

Fernando Esteban de la Rosa is a professor in the Department of Private International Law at the University of Granada in Granada, Spain; he may be reached by email at: festeban@ugr.es

Notes

1. Jean-Paul Triaille, Jérôme de Meeûs d'Argenteuil, and Amélie de Francquen, "Study on the Legal Framework of Text and Data Mining (TDM)," European Commission, 2014, <https://data.europa.eu/doi/10.2780/1475>.
2. Matthew Sag, "The New Legal Landscape for Text Mining and Machine Learning," *Journal of the Copyright Society of the USA* 66, 2 (2019): 291–367, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3331606.
3. International Federation of Library Associations and Institutions (IFLA), "IFLA Statement on Text and Data Mining," 2013, <https://www.ifla.org/es/publications/ifla-statement-on-text-and-data-mining-2013/>.
4. Leslie A. Williams, Lynne M. Fox, Christophe Roeder, and Lawrence Hunter, "Negotiating a Text Mining License for Faculty Researchers," *Information Technology and Libraries* 33, 3 (2014): 5–21, <https://doi.org/10.6017/ital.v33i3.5485>; Philip Young, Inga Haugen, Edward F. Lener, Virginia Pannabecker, and Collin Brittle, "Library Support for Text and Data Mining: A Report for the University Libraries at Virginia Tech," 2017, <http://hdl.handle.net/10919/78466>.
5. Darby Orcutt, "Liaisonship, the Law, & Libraries: Supporting Content Mining Research," *Against the Grain* 29, 4 (2017): 26, <https://doi.org/10.7771/2380-176X.7821>; Mary Ellen Bates, "Mad Men and Research Librarians," *Online Searcher* 43, 2 (2019): 64; Rachael G. Samberg and Cody Hennesy, "Law and Literacy in Non-Consumptive Text Mining: Guiding Researchers through the Landscape of Computational Text Analysis," chap. 17 in *Copyright Conversations: Rights Literacy in a Digital World*, Sara R. Benson, ed. (Chicago: Association of College and Research Libraries, 2019), 289–315; University of California, Berkeley, "Building Legal Literacies for Text Data Mining," 2021, <https://berkeley.pressbooks.pub/buildingltdm/back-matter/white-paper/>; Jessica Harris and Kristin E. Martin, "Let's Get Technical—Text and Data Mining Support at the University of Chicago Library," *Against the Grain* 34, 2 (2022): 57–58.
6. Donna L. Ferullo, "Major Copyright Issues in Academic Libraries," *Journal of Library Administration* 40, 1–2 (2004): 23–40, https://doi.org/10.1300/J111v40n01_03; Rebecca S. Albitz, "Copyright Information Management and the University Library: Staffing, Organizational Placement and Authority," *Journal of Academic Librarianship* 39, 5 (2013): 429–35, <https://doi.org/10.1016/j.acalib.2013.04.002>; LeEtta Schmidt, "Copyright Educational Services and Information in Academic Libraries," *Public Services Quarterly* 15, 4 (2019): 283–99, <https://doi.org/10.1080/15228959.2019.1644267>; Juan Carlos Fernández-Molina, Daniel Martínez-Avila, and Eduardo Graziosi Silva, "University Copyright/Scholarly Communication Offices: Analysis of their Services and Staff Profile," *Journal of Academic Librarianship* 46, 2 (2020), <https://doi.org/10.1016/j.acalib.2020.102133>; Theresa L. Adu and Thomas B. van der Walt, "Copyright Law and Academic Libraries: From Theory to Practice," *portal: Libraries and the Academy* 22, 3 (2022): 725–44, <https://doi.org/10.1353/pla.2022.0029>.
7. Lorena Siguenza-Guzman, Victor Saquicela, Elina Avila-Ordóñez, Joos Vandewalle, and Dirk Catrysse, "Literature Review of Data Mining Applications in Academic Libraries," *Journal of Academic Librarianship* 41, 4 (2015): 499–510, <https://doi.org/10.1016/j.acalib.2015.06.007>; Sumeer Gul and Shohar Bano, "Smart Libraries: An Emerging and

- Innovative Technological Habitat of 21st Century," *Electronic Library* 37, 5 (2019): 764–83, <https://doi.org/10.1108/EL-02-2019-0052>; Alex H. Poole and Deborah A. Garwood, "Natural Allies': Librarians, Archivists, and Big Data in International Digital Humanities Project Work," *Journal of Documentation* 74, 4 (2018): 804–26, <https://doi.org/10.1108/JD-10-2017-0137>; Gaohui Cao, Mengli Liang, and Xuguang Li, "How to Make the Library Smart? The Conceptualization of the Smart Library," *Electronic Library* 36, 5 (2019): 811–25, <https://doi.org/10.1108/EL-11-2017-0248>.
8. European Union, "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. L 130, 17 May 2019," <https://eur-lex.europa.eu/eli/dir/2019/790/oj>.
 9. Niva Elkin-Koren and Neil Weinstock Netanel, "Transplanting Fair Use across the Globe: A Case Study Testing the Credibility of U.S. Opposition," *Hastings Law Journal* 72, 4 (2021): 1121–81, https://repository.uclawsf.edu/hastings_law_journal/vol72/iss4/4/.
 10. Sean Flynn, Luca Schirru, Michael Palmedo, and Andrés Izquierdo, "Research Exceptions in Comparative Copyright," PIJIP [Program on Information Justice and Intellectual Property]/TLS [Program on Technology, Law and Security] Research Paper Series no. 75, 2022, <https://digitalcom-mons.wcl.american.edu/research/75>.
 11. Republic of Ecuador, Ministerio "Código Orgánico de la Economía Social de los Conocimientos, Creatividad e Innovación [Organic code of the social economy of knowledge, creativity, and innovation]," 2016, <https://www.gob.ec/regulaciones/codigo-organico-economia-social-conocimientos-creatividad-innovacion>.
 12. Matías Jackson Bertón, "Text and Data Mining Exception in South America: A Way to Foster AI Development in the Region," *GRUR* (Gewerblicher Rechtsschutz und Urheberrecht [intellectual property and copyright]) *International* 70, 12 (2021): 1145–57, <https://doi.org/10.1093/grurint/ikab081>; Patricia Díaz Charquero and Viviana Rangel, "Políticas de inteligencia artificial y derechos de autor en América Latina (Artificial intelligence and copyright policies in Latin America)," Fundación Karisma, 2023, https://web.karisma.org.co/wp-content/uploads/2023/03/Informe_Políticas_Inteligencia_Artificial_DDHH_LATAM.pdf.
 13. Eleonora Rosati, "Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and Its Role in the Development of AI Creativity," *Asian Pacific Law Review* 27, 2 (2020): 198–217, <https://ssrn.com/abstract=3452376>.
 14. Sag, "The New Legal Landscape for Text Mining and Machine Learning."
 15. Daniel Gervais, "Exploring the Interfaces between Big Data and Intellectual Property Law," *Journal of Intellectual Property, Information Technology and E-Commerce Law* 10, 1 (2019): 3–19, <https://www.jipitec.eu/issues/jipitec-10-1-2019/4875>; Sean Flynn, Christophe Geiger, João Pedro Quintais, Thomas Margoni, Matthew Sag, Lucie Guibault, and Michael W. Carroll, "Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action," *European Intellectual Property Review* 42, 7 (2020): 393–98, <http://dx.doi.org/10.2139/ssrn.3578819>.
 16. Marco Caspers, Lucie Guibault, Kiera McNeice, Stelios Piperidis, Kanella Pouli, Maria Eskevich, and Maria Gavrillidou, "Baseline Report of Policies and Barriers of TDM [text and data mining] in Europe," 2016, https://project.futuretdm.eu/wp-content/uploads/2017/05/FutureTDM_D3.3-Baseline-Report-of-Policies-and-Barriers-of-TDM-in-Europe.pdf.
 17. Japan, "Copyright Act (Act No. 48 of May 6, 1970, as Amended up to January 1, 2022)," <https://www.wipo.int/wipolex/en/text/584874>.
 18. Tatsuhiro Ueno, "The Flexible Copyright Exception for 'Non-Enjoyment' Purposes. Recent Amendment in Japan and Its Implication," *GRUR International* 70, 2 (2021): 145–52, <https://doi.org/10.1093/grurint/ikaa184>.
 19. Ueno, "The Flexible Copyright Exception for 'Non-Enjoyment' Purposes."



20. Rossana Ducato and Alain Strowel, "Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to 'Machine Legibility,'" *IIC—International Review of Intellectual Property and Competition Law* 50, 6 (2019): 649–84, <https://doi.org/10.1007/s40319-019-00833-w>; Alain Strowel, "Reconstructing the Reproduction and Communication to the Public Rights: How to Align Copyright with Its Fundamentals," chap. 7 in *Copyright Reconstructed: Rethinking Copyright's Economic Rights in a Time of Highly Dynamic Technological and Economic Change*, P. Bernt Hugenholtz, ed. (The Hague, Netherlands: Kluwer Law International, 2018), 203–40.
21. European Union, "Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the Harmonisation of Certain Aspects of Copyright and Related Rights in the Information Society," <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32001L0029>.
22. United Kingdom (UK), "Copyright, Designs and Patents Act," 1988, <https://www.legislation.gov.uk/ukpga/1988/48/contents>.
23. UK Intellectual Property Office, "Exceptions to Copyright: Research," 2014, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/375954/Research.pdf.
24. H. Andrés Izquierdo, "Minería de textos y datos e inteligencia artificial: Nuevas excepciones al derecho de autor [Text and data mining and artificial intelligence: new copyright exceptions]," *THÉMIS: Revista de Derecho [THÉMIS: law magazine]* 79 (2021): 323–43, <https://doi.org/10.18800/themis.202101.018>.
25. Alina Trapova and João Pedro Quintais, "The UK Government Moves Forward with a Text and Data Mining Exception for All Purposes," *Kluwer Copyright Blog*, 2022, <https://copyrightblog.kluweriplaw.com/2022/08/24/the-uk-government-moves-forward-with-a-text-and-data-mining-exception-for-all-purposes/>.
26. Pierre N. Leval, "Toward a Fair Use Standard," *Harvard Law Review* 103, 5 (1990): 1105–36, <https://doi.org/10.2307/1341457>.
27. Benjamin L. W. Sobel, "Artificial Intelligence's Fair Use Crisis," *Columbia Journal of Law & the Arts* 41, 1 (2018): 45–97, <https://doi.org/10.7916/jla.v41i1.2036>.
28. Sag, "The New Legal Landscape for Text Mining and Machine Learning."
29. Marta Nadal Cebrián, "¿Pueden los sistemas de inteligencia artificial «aprender» utilizando obras protegidas por derechos de propiedad intelectual sin la autorización de sus titulares? [Can artificial intelligence systems "learn" using works protected by intellectual property rights without the permission of their owners?]" *Revista de Propiedad Intelectual [Journal of intellectual property]* 68 (2021): 15–74.
30. Sag, "The New Legal Landscape for Text Mining and Machine Learning."
31. Izquierdo, "Minería de textos y datos e inteligencia artificial."
32. James Grimmelmann, "Copyright for Literate Robots," *Iowa Law Review* 101, 2 (2016): 657–81, <https://ilr.law.uiowa.edu/print/volume-101-issue-2/copyright-for-literate-robots>.
33. World Intellectual Property Organization (WIPO), "WIPO Copyright Treaty," 1996, <https://www.wipo.int/wipolex/en/text/295166>
34. J. Carlos Fernández-Molina, "Laws against the Circumvention of Copyright Technological Protection," *Journal of Documentation* 59, 1 (2003): 41–68, <https://doi.org/10.1108/00220410310458000>.
35. United States Copyright Office, "Section 1201 Exemptions to Prohibition against Circumvention of Technological Measures Protecting Copyrighted Works," 2021, <https://www.copyright.gov/1201/2021/>.
36. Michael W. Carroll, "Copyright and the Progress of Science: Why Text and Data Mining Is Lawful," *UC Davis Law Review* 53, 2 (2019): 893–964, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3531231.
37. Maarten Truyens and Patrick Van Eecke, "Legal Aspects of Text Mining," *Computer Law & Security Review* 30, 2 (2014): 153–70, <https://doi.org/10.1016/j.clsr.2014.01.009>; Juan-Carlos Fernández-Molina, Kristin R. Eschenfelder, and Alan Rubel, "Comparing Use Terms

- in a Spanish and US Research University E-Journal Licenses: Recent Trends, " *College and Research Libraries* 82, 2 (2021): 158–81, <https://crl.acrl.org/index.php/crl/article/view/24830/3266>.
38. Christophe Geiger, Giancarlo Frosio, and Oleksandr Bulayenko, "Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU," in *Propiedad intelectual y mercado único digital europeo* [Intellectual property and the European digital single market], Concepción Saiz García and Raquel Evangelio Llorca, eds. (Valencia, Spain: Tirant lo Blanch, 2019), 27–71, <http://dx.doi.org/10.2139/ssrn.3470653>.
39. European Commission, "Licences for Europe' Stakeholder Dialogue," 2017, <https://digital-strategy.ec.europa.eu/en/library/licences-europe-stakeholder-dialogue>.
40. European Commission, "Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market," 2016, <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52016PC0593>.
41. Reto M. Hilty and Heiko Richter, "Text and Data Mining," Part B, chap. 1 in *Modernisation of the EU Copyright Rules: Position Statement of the Max Planck Institute for Innovation and Competition*, Reto M. Hilty and Valentina Moscon, eds. (Munich: Max Planck Institute for Innovation and Competition, 2017), 25–34, https://pure.mpg.de/rest/items/item_2470998_12/component/file_2479390/content.
42. Ángel García Vidal, "Propiedad intelectual y minería de textos y datos: Estudio de los artículos 3 y 4 de la Directiva (UE) 2019-790 [Copyright, related rights and text and data mining: Study of Articles 3 and 4 of the Directive (EU) 2019-790]," *Actas de Derecho Industrial y Derecho de Autor* [Industrial law and copyright proceedings] 40 (2020): 99–124, <http://hdl.handle.net/10347/25639>.
43. Roberto Caso, "Il conflitto tra diritto d'autore e ricerca scientifica nella disciplina del text and data mining della direttiva sul mercato unico digitale [The conflict between copyright and scientific research in the text and data mining exception under the digital single market directive]," Trento LawTech Research Paper 38, 2020, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3533401.
44. Jonathan Griffiths, Tatiana Synodinou, and Raquel Xalabarder, "Comment of the European Copyright Society Addressing Selected Aspects of the Implementation of Articles 3 to 7 of Directive (EU) 2019/790 on Copyright in the Digital Single Market," *GRUR International* 72, 1 (2023): 22–36, <https://doi.org/10.1093/grurint/ikac137>.
45. Gervais, "Exploring the Interfaces Between Big Data and Intellectual Property Law."
46. Institute for Information Law, University of Amsterdam, Netherlands, "Study on the Implementation and Effect in Member States' Laws of Directive 2001/29/EC on the Harmonisation of Certain Aspects of Copyright and Related Rights in the Information Society," 2007, https://www.ivir.nl/publicaties/download/Infosoc_report_2007.pdf.
47. Begoña Gonzalez Otero, "Las excepciones de minería de textos y datos más allá de los derechos de autor: La ordenación privada contraataca [The text and data mining exceptions over and above copyright: Private ordering strikes back]," in Saiz García and Evangelio Llorca, *Propiedad intelectual y mercado único digital europeo*, 73–105, <http://dx.doi.org/10.2139/ssrn.3477197>.
48. Flynn, Geiger, Quintais, Margoni, Sag, Guibault, and Carroll, "Implementing User Rights for Research in the Field of Artificial Intelligence."
49. Griffiths, Synodinou, and Xalabarder, "Comment of the European Copyright Society."
50. Pedro Alberto de Miguel Asensio, "La lex loci protectionis tras el Reglamento Roma II [The lex loci protectionis (law of the place of protection) after the Rome II Regulation]," *Anuario español de derecho internacional privado* (Spanish Yearbook of Private International Law), 2007, 375–406.
51. Regulation (EC) No 864/2007 of the European Parliament and of the Council of 11 July 2007 on the Law Applicable to Non-Contractual Obligations (Rome II), <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:199:0040:0049:EN:PDF>.



52. Regulation (EC) No 593/2008 of the European Parliament and of the Council of 17 June 2008 on the Law Applicable to Contractual Obligations (Rome I), <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:177:0006:0016:en:PDF>.
53. Carroll, "Copyright and the Progress of Science."
54. WIPO, "WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI): Third Session," 2020, https://www.wipo.int/meetings/en/details.jsp?meeting_id=59168.
55. Sean M. Fiil-Flynn et al., "Legal Reform to Enhance Global Text and Data Mining Research. Outdated Copyright Laws around the World Hinder Research," *Science* 378, 6623 (2022): 951–53, <https://www.science.org/doi/10.1126/science.add6124>.
56. P. Bernt Hugenholtz, "Flexible Copyright: Can the EU Author's Rights Accommodate Fair Use?" in *Copyright Law in an Age of Limitations and Exceptions*, Ruth L. Okediji, ed. (Cambridge, UK: Cambridge University Press, 2017), 275–91.
57. Mark A. Lemley and Bryan Casey, "Fair Learning," *Texas Law Review* 99, 4 (2021): 743–85, <https://texaslawreview.org/fair-learning/>.

This mss. is peer reviewed, copy edited, and accepted for publication, portal 24.3.