**ACCESS TO DATA: THE EUROPEAN SOLUTION FOR TEXT AND DATA MINING AND THE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE**

by

*Luisa Gatti*

*Master of Laws in Intellectual Property - 2019/2020*

*WIPO Academy, University of Turin and International Training Centre of the ILO*

**Abstract**

In recent years, data has been central to the transformation of market and society. Data-driven innovation requires access to data and its analysis in order to extract knowledge and value from the information it conveys. This is particularly crucial for the development of artificial intelligence. However, such use and exploitation of data may interfere with the exclusive rights of intellectual property owners. After a brief introduction, this paper will examine the restrictions deriving from copyright and the exceptions permitted under the legal framework prior to the coming into force of Directive (EU) 2019/790. This will be followed by a description of the new exceptions for text and data mining introduced by the above-mentioned Directive, evaluating the positive and negative aspects, and assessing whether these new provisions will prove to be a balanced solution for access to and analysis of data.

**Table of contents**

1. INTRODUCTION
2. EDMOND DE BELAMY AND OTHERS: CREATIVITY REQUIRES DATA
3. THE CONFLICT BETWEEN IPRs AND THE TRAINING OF AI SYSTEMS
4. THE LEGAL FRAMEWORK BEFORE THE ADVENT OF DIRECTIVE (UE) 2019/790
5. AN UPDATE TO COPYRIGHT: THE NEW MANDATORY EXCEPTIONS
6. Art. 4 and the new dynamism of the express reservation in the light of access to data
7. CONCLUSIONS
8. **Introduction**

The aim of this paper is to investigate the phenomenon of data mining in the context of copyright and database *sui generis* right, analysing the legal framework set by the European Union, especially the provisions contained in Art. 4 of Directive (UE) 2019/790, in the light of the increasing need to access and analyse data.

New technologies and the process of digitization are leading to a data-driven world in which market and society are totally reshaped. Data has become central to this transformation, as it is the “new oil”[[1]](#footnote-1) and the “lifeblood of economic development”.[[2]](#footnote-2) The total volume of data is constantly expanding from an estimated 33 zettabytes in 2018 to a forecast figure of 175 zettabytes by 2025.[[3]](#footnote-3) The importance of data and the necessity of developing a data-driven economy has been stressed many times by the European Union and has been confirmed by recent legislative reforms. One of the best-known examples is Regulation (EU) 2016/679 (General Data Protection Regulation, or GDPR, on the protection of natural persons with regard to the processing of personal data and on the free movement of such data), which states that notwithstanding recognition of the protection of personal data as a fundamental right,[[4]](#footnote-4) the free movement of personal data must not be restricted.[[5]](#footnote-5) This was followed by Regulation (EU) 2018/1807 (on a framework for the free flow of non-personal data), which enhanced the so-called fifth freedom as a pre-requisite for a smoothly functioning and dynamic market.[[6]](#footnote-6) Extended access to data is a key factor in this context. However, the availability of data is fruitless without the contextual possibility of analysing it in order to extract its maximum value.[[7]](#footnote-7)

Oil is only valuable if you can burn it. Similarly, data, as the new oil, needs to be analysed in order to create knowledge and innovation. Nowadays, there are many technologies that can perform automated computational analysis of digital content, whether it be text, sounds, images or, more generally, data. These techniques are generally known as text and data mining and consist in processing large amounts of information in order to gain new knowledge and discover new patterns and trends. Such capacity for processing data is particularly evident in artificial intelligence (AI) systems[[8]](#footnote-8) and especially in machine-learning techniques whereby “algorithms are trained to infer certain patterns based on a set of data in order to determine the actions needed to achieve a given goal”.[[9]](#footnote-9)

These technologies are having a positive impact on our society and can be useful in many fields. Despite the benefits they bring, the activities performed by AI may lead to infringements of the exclusive rights granted to intellectual property right (IPR) owners. The following section will describe a practical case illustrating the interplay between the training of an AI system and existing legal restrictions. The paper will then identify the copyright-related issues that arise, and review the possible solutions provided by the legal framework prior to the advent of Directive 2019/790 on copyright and, subsequently, by the Digital Single Market (DSM) Directive. Finally, it will analyse the new mandatory exceptions for text and data mining (TDM) introduced by the DSM Directive, and attempt to assess whether the solutions provided will support or hinder the need to allow access to data and develop AI.

1. **Edmond de Belamy and others: creativity requires data**

AI and TDM are starting to play a substantial role in many different fields because of the increased efficiency they provide and the new opportunities they create.[[10]](#footnote-10) Art, which is just one of the many fields in which AI is used, gives a well-known example that illustrates the main copyright issues involved in the act of training an AI system.

On October 2018, Christie’s sold a picture entitled “Portrait of Edmond de Belamy” for 432,500 USD. The subject of the work is a member of the fictional Belamy family, who, judging by his dark frockcoat and plain white collar, seems to be a churchman. The work looks unfinished as the features of the figures are indistinct and blurry and there are some blank areas. The portrait became famous because it was advertised as the first ever AI-generated work to be auctioned[[11]](#footnote-11). It was also an object of controversy because it was produced by a collective called Obvious, using, without acknowledgement, the code of another artist and programmer, Robbie Barrat, who shared his algorithms online via an open-source license.[[12]](#footnote-12) The technology used was a generative adversarial network (GAN) method, consisting of two algorithms, the “generator” and the “discriminator”, in competition one with the other. Using this method, images with common features are first collected to create the data set used to train the AI system. Then the generator is employed to generate images which are supposed to “fool” the discriminator. The role of the second algorithm is to compare the images the generator sends to it with pictures in the data set, trying to determine whether they were made by humans or by the algorithm. The generator obtains feedback from the discriminator and uses it to adapt and generate more and more realistic images, until the discriminator recognizes the pictures concerned as human-created.[[13]](#footnote-13)

The data set used for the AI system consisted of 15,000 portraits from the 14th to the 19th centuries, taken from the online encyclopaedia “WikiArt”.[[14]](#footnote-14) The question to be answered is: had the images still been covered by copyright, could the activity of collecting pictures and using them to develop the AI system have constituted a copyright infringement?

**3. The conflict between IPRs and the training of AI systems**

As we have seen, the creation of the portrait involved a huge number of pictures. The use of such technology is peculiar because, even though the purpose was to create a work of art, the exploitation of such images using AI is so different from the “traditional” method. The images are involved in an automated process, in which the knowledge and value extracted depend on the volume and completeness of the dataset, and the correlation of the data.[[15]](#footnote-15) Furthermore, the process is not dependent on human creativity, but on the generation of information.[[16]](#footnote-16)

Considering the implications for copyright, it may be argued that TDM requires the extraction of data that, though embedded within the works of art concerned, amount to objective patterns and facts, and as such are not protected.[[17]](#footnote-17) One of the fundamental and generally recognized principles of copyright law is the idea/expression dichotomy, according to which copyright does not protect ideas and information but rather the way they are expressed. Consequently, we can view TDM and the training of an AI system as a “neutral” activity from the point of view of intellectual property protection: TDM is an act of analysis; the aim is not to benefit from the expressive features of the works of art, but to extract information from their content and about their content that *per se* falls outside the scope of copyright protection. In a recent document published by the WIPO, a “Draft issues paper on intellectual property policy and artificial intelligence”, [[18]](#footnote-18) the conflict between copyright and the development of AI is specifically raised using interesting and ambiguous expressions such as “data subsisting in copyright” or “data protected by copyright”.[[19]](#footnote-19) Indeed, from the point of view of the person training an AI system, copyrighted works are part of a data set and are not used on account of their expressive features. From the point of view of the right holder, on the other hand, it is because of its expressive elements that a work is able to become a repository of data.[[20]](#footnote-20)

The activity of reading and studying a work does not cause particular problems if it is undertaken by a human being, but when it is carried out by a machine in an automated process, it may infringe copyright. This is because, when an automated process acts on copyrighted works in order to extract the non-protected information, in most cases it has to create a reproduction of the works concerned. This is the moment when copyright enters the framework, as Art. 2 of Directive 2001/29 (Infosoc Directive) grants the right holder an exclusive right of reproduction.

TDM may consist in the extraction and reproduction of a database protected by copyright or by *sui generis* rights. Particularly in the case of *sui generis* rights, the interference with data mining is significant and not connected to mere technical reasons. This form of protection is more closely linked to the market use of data because: 1) protection is granted in consideration of the “substantial investment in either the obtaining, verification or presentation of the contents”;[[21]](#footnote-21) 2) the exclusive right reserved to the database owner is the right to prevent extractions and re-utilizations of the whole or of a substantial part of the contents.[[22]](#footnote-22)

It should also be mentioned that TDM may involve data that does not fall within the protection of copyright or of the database *sui generis* right. However, the absence of protection under these forms of IPR[[23]](#footnote-23) does not exclude the possibility of the owner imposing restrictions (also pertaining to TDM) through contractual provisions.[[24]](#footnote-24)

In conclusion, from the point of view of intellectual property, TDM may be legally undertaken if either the authorization of the right holder has been obtained or there is an exception covering such activities. As we shall see, legal access to the work does not imply free access to the data.[[25]](#footnote-25)

**4. The legal framework before the advent of Directive (UE) 2019/790**

Among the exceptions and provisions already present in the European Union system before the coming into force of Directive 2019/79, the most obvious is the exception for temporary acts of reproduction contained in Art. 5.1 of Directive 2001/29 (Infosoc Directive). This is one of the exceptions dealing with the problem of making copies in the context of digital technology. The fact is that the online environment generates recurrent situations where copying is a requirement for engaging in acts such as browsing and caching,[[26]](#footnote-26) which are common and necessary steps in a technological process. The European legislator considered such situations when introducing the limitation and, taking into account the common ground of the digital environment, it may be regarded as the closest to the context of TDM. Furthermore, it is the only mandatory exception in the Infosoc Directive and is implemented in all Member States in mostly identical wording.[[27]](#footnote-27)

Pursuant to this limitation, the reproduction of works is allowed without prior authorization in the presence of certain conditions which, according to the European Court of Justice, are cumulative[[28]](#footnote-28) and must be interpreted restrictively.[[29]](#footnote-29) The act must be i) temporary, whether “transient or incidental”, ii) an “integral and essential part of a technological process”, iii) performed with the sole purpose of enabling “a transmission in a network between third parties by an intermediary or a lawful use” and iv) without any independent economic significance.

Whether this exception can apply to a TDM process depends on the technologies used and how it conforms to the above critical points. Among them, the necessary presence of a temporary act as transient or incidental is one of the most crucial. In its Infopaq I Decision, the Court of Justice stated that an act is transient “only if its duration is limited to what is necessary for the proper completion of the technological process of which it forms an integral and essential part, being understood that that process must be automated so that it deletes that act automatically, without human intervention, once its function of enabling the completion of such a process has come to an end”.[[30]](#footnote-30) On the other hand, the concept of an “incidental” copy seems to include acts that last longer than a transient copy but are a necessary stage in the main act of exploitation of the work, provided that the reproduction is nevertheless temporary.[[31]](#footnote-31) It is unlikely that a copy used to mine data would be transient: the material must be available for a certain period of time to be transformed, loaded and/or analysed.[[32]](#footnote-32) Furthermore, human intervention to delete copies of works is usually required, so the cancellation is not automatic, as stipulated by the Court of Justice.[[33]](#footnote-33)

However, it is possible that a temporary copy made during extraction could be considered incidental to the process, for example if it were made for the purpose of analysis of the work.[[34]](#footnote-34) It would depend a lot on the technology used, but in many cases the copies made would be permanent.[[35]](#footnote-35) It has been noted that, for high-quality data analyses, as in the case of training an AI system, temporary reproductions are usually not sufficient. In most situations, the data set needs to be normalized, annotated or altered. All of these preparatory works usually depend on longer storage periods than the Infosoc Directive permits.[[36]](#footnote-36) It is also possible that the materials would need to be retained for such uses as verification, aggregation with other data sets and further analysis.

 Another requirement is the absence of independent economic advantage. The copy must not have any value distinct or separable from the economic benefit derived from the lawful use of the work concerned, nor may it generate an additional economic advantage going beyond that derived from that use of the protected work.[[37]](#footnote-37) A single act of reproduction may not be of significant economic value, which, however, is not the same as it not having any independent economic value. TDM activity seems to determine an additional economic benefit that can be potentially very huge.[[38]](#footnote-38) This may be one of the most challenging requirements, as it does not depend on the specific technique used or the technological development. In conclusion, the temporary reproduction exception may not apply to TDM activities and, even in cases where it may, there is still not enough legal certainty, partly due to technical uncertainties, to ensure that users of AI will be sufficiently insulated from liability.

In the context of copyright, the private use exception provided by Art. 5.3 b) of the Infosoc Directive is also relevant as it involves “reproductions on any medium made by a natural person for private use and for ends that are neither directly nor indirectly commercial, on condition that the rightholders receive fair compensation …”. The application is restricted to natural persons, but there are some categories that might benefit from this limitation, such as individual researchers[[39]](#footnote-39) or users of AI tools.[[40]](#footnote-40) However, the exception is narrowed down even further by limitation of the purpose: making a private copy is allowed only in the absence of direct or indirect commercial use, and subject to fair remuneration. Moreover, even though this voluntary exception has been implemented in all the Member States,[[41]](#footnote-41) national legislatures have introduced the exception in very different ways, resulting in considerable fragmentation.[[42]](#footnote-42)

For databases protected by copyright, a possible exception for TDM activities is provided by Art. 6 of Directive (EU) 1996/9, which permits the “normal use of the contents by the lawful user”. However, this exception is not clearly defined and is left to the interpretation of the individual Member States.[[43]](#footnote-43)

For databases protected by *sui generis* rights, it is possible to extract and/or re-utilize insubstantial parts of their contents, evaluated qualitatively and/or quantitatively, for any purposes whatsoever, as provided for by Art. 8.1 of Directive 96/9. However, such TDM activity may involve the copying of huge amount of data and be in conflict with such provisions.

Finally, a research exception limited to a specific purpose, disallowing both direct and indirect commercial activities, is provided for copyright and *sui generis* database rights.[[44]](#footnote-44)

In conclusion, in a system like the European Union’s, with a limited list of exemptions, which in any case are interpreted restrictively, there was previously little room for certainty as to the legality of such activities as TDM. Provision of a specific exception was therefore necessary, as existing limitations did not seem to accommodate the specific characteristics of TMD activities.

**5. An update to copyright: the new mandatory exceptions**

Given the legal uncertainty and the increasing relevance of TDM, the European Union legislator introduced, with Directive 2019/790 (on copyright in the Digital Single Market), two mandatory exceptions for TDM, the first for purposes of scientific research (Art. 3), the second for all other purposes (Art. 4).

Art. 3 permits the extraction and reproduction of texts and data from works or other subject matter lawfully accessible, thus providing an exception where the reproduction of copyright-protected works is concerned. This exception also applies to the new right granted to publishers of press publications by Art. 15 of the DSM Directive, and the right to prevent extraction accorded to database owners. Copies of the contents may be retained for the purpose of scientific research, including for the verification of research results. However, the scope of the exception relates to a restricted list of beneficiaries, research organizations and cultural heritage institutions, which either act “on a not-for-profit basis or in the context of a public-interest mission recognised by the State”.[[45]](#footnote-45) According to Art. 7.1, this exception cannot be overridden by contract.

The second exception, contained in Art. 4, is of particular importance for the present dissertation: it does not require that beneficiaries have a specific qualification, and it comprises all the activities of extraction and reproduction of legally accessible works and other subjects that fall outside the scope of the exception for research purposes or other applicable exceptions. Furthermore, the exception is also extended to software authors’ rights of reproduction and adaptation provided by Art. 4.1 a) and b) of Directive 2009/24/EC.[[46]](#footnote-46) However, copies of works and other subject matter may be kept only for the time necessary for the purposes of TDM, and application of the exception is subject to the condition that use for the purposes of TDM has not been expressly reserved by right holders in an appropriate manner.

Thus, Art. 4 allows unqualified beneficiaries to carry out data-mining activities, provided the content is lawfully accessible and the right holder has not expressed a reservation.

An interpretation of the concept of “lawful access” is given by Recital 14 of the DSM Directive and applies to content freely available online, as well as material made available to the user by contractual arrangements (e.g., subscriptions paid by an individual or institution, or open-access licenses).[[47]](#footnote-47)

There is, however, an opt-out mechanism where this exception is concerned:[[48]](#footnote-48) data mining is permitted by default, unless this form of use of the works in question has been expressly reserved in an “appropriate manner”. This is confirmed by the fact that Art. 7.1 does not mention Art. 4 among the provisions that cannot be overridden by contract. A similar mechanism is also stipulated in Articles 8 and 12, which provide for extended collective licences on copyright (ECLs). Article 8 obliges Member States to introduce a legal solution to allow representative Collective Management Organisations (CMOs) to issue licenses for the use of out-of-commerce works by cultural heritage institutions "irrespective of whether all rights holders covered by such licences have mandated the collective management organisation"; Article 12 comprises an optional rule that goes beyond the case of out-of-commerce works and introduces a legal basis for extended collective licensing in "well-defined areas of use, where obtaining authorisations from rights holders on an individual basis is typically onerous and impractical".[[49]](#footnote-49) Both Art. 8.4 and Art. 12.3 c) further require Member States to allow right holders to “easily and effectively” exclude their works and other subject matter from the license mechanism.[[50]](#footnote-50) Because of the similarity of structures, it has actually been suggested that Art. 4 should be seen as a form of extended collective licensing, also considering the emphasis of the DSM Directive on encouraging licensing.[[51]](#footnote-51)

The opt-out system is therefore a key element of Art. 4, and a crucial aspect of it is the determination of what may be considered an “appropriate manner” of retaining the possibility of performing TDM activities on the content concerned.

It will be up to Member States, when transposing the Directive, to further clarify this aspect and, ultimately, if the national transposition does not resolve these uncertainties, the interpretation of the rule will be referred to the courts. However, some indications are already given by Recital 18 of the DSM Directive. The text of the recital is not perfectly clear, but it may be interpreted as saying that, in the case of content made publicly available online, only “machine-readable means”, such as “metadata and terms and conditions”, should be considered, provided they do not prevent uses other than the reservation of rights for the purpose of data mining. In the case of content not publicly available online, “other means”, not machine-readable, may be suitable and may take the form of contractual agreements and unilateral declarations.[[52]](#footnote-52)

Some further indications for understanding the concept of “appropriate manner” may be gained by looking at previous exceptions, already implemented, including the possibility for right holders of reserving use of their works. For example, the exception contained in Art. 5.3 c) of Directive 2001/29, pertaining to the reproduction by the press, communication to the public or making available of published articles on certain topics, allows right holders to reserve such rights and has been implemented by the Italian legislator, requiring to place the words “reproduction reserved” or an analogous expression, even in an abbreviated form, at the beginning or end of an article.[[53]](#footnote-53) From this precedent, we may deduce that, on one hand, reservation should not be particularly burdensome for right holders, allowing them to retain use for TDM in a reasonably easy way. A similar conclusion may also be arrived at by considering the above-mentioned provision on ECLs, which requires that right holders be allowed to “easily and effectively” exclude their works from the license mechanism. On the other hand, the indication should be clear and codified in such a way as to be readily recognizable.

In the context of the provisions that we are discussing, and for the sake of their relevance to the digital environment, it is important to properly identify what constitutes an “appropriate manner” which is also “machine-readable”. It has been suggested that contractual provisions confirmed and/or embedded in the technical instructions of a website should be considered sufficient,[[54]](#footnote-54) for example by using a robots.txt file: an exclusion protocol incorporated in the root directory of a website to instruct robots on the permitted use of the site.[[55]](#footnote-55) Another way may be to use meta-tags giving instructions on each individual web page and describing what uses are not allowed.[[56]](#footnote-56) Confirmation of the possibility of their use can be derived from the jurisprudence on implied licenses. In my opinion, the following decisions pertaining to implied licenses connected with online content and machine-readable measures come closest to the provision we are considering, because implied licenses create an opt-out system whereby content may be used in the absence of a proper reservation on the part of the right holder.[[57]](#footnote-57) In the US case of Field v. Google, the court held that Google was allowed to create a "cached" version of the plaintiff’s website containing his copyrighted works because license could have been implied by the absence of metatags.[[58]](#footnote-58) In Europe, a German court ruled that a website owner who did not adopt widely available measures, such as a robot exclusion protocol (robot.txt), granted an implied licence to Google to crawl and display her works as thumbnails in Google’s image-search function[[59]](#footnote-59). The implied license theory was later rejected by the Court of Appeal[[60]](#footnote-60) and by the German Federal Supreme Court.[[61]](#footnote-61) However, the latter decided in Google’s favour and held that the reproduction of the images as preview pictures was not unlawful because Google could interpret the claimant’s actions of not using the technical possibilities available as consent with regard to the display of her works in Google’s image-search results.[[62]](#footnote-62) From these decisions, we may deduce that warnings given by robot.txt files and metatags may be an appropriate manner of reserving use for TDM.

Another point that should be stressed about the opt-out system is that the subject allowed to reserve the use of works for TDM is the “right holder”. This term seems not to include other subjects, such as licensees. This may have an impact on websites and platforms sharing user content. Will they be able to legally prevent TDM by third parties, and how they will introduce the opt-out mechanism into their frameworks?[[63]](#footnote-63)

In conclusion, the European legislator has opted for a dual solution, distinguishing between specific beneficiaries that come within the scope of Art. 3 and other subjects whose activities will depend in particular on the absence of any reservation of rights on the part of the right holder. The introduction of limitations on TDM was certainly needed, and the European Union has positively introduced the exceptions contained in Arts. 3 and 4 as mandatory.[[64]](#footnote-64) However, there is a need to assess whether certain aspects of the reform are problematic, especially from the point of view of those wanting to perform TDM activities.

**6. Art. 4 and the new dynamism of the express reservation in the light of access to data**

The European Union, in an attempt to find a balance between the different visions and interests of right holders and users, gave its answer in the shape of the reform contained in the DSM Directive. The European system now provides an exception for determined subjects, research organizations and cultural heritage institutions, but it also provides for innovation on the part of the private sector, for which the reform prescribes a different treatment. Under the general exception contained in Art. 4 everything may potentially become mineable by default. However, given the possibility of reserving use, it is in fact left to the right holder to allow TDM activities. Consequently, the Directive legitimizes a derivative market in TDM, enabling right holders to control, license or entirely prohibit such activities.[[65]](#footnote-65) This is explicitly confirmed by Recital 18, according to which right holders “remain able to license the uses of their works or other subject matter” in all cases falling outside the scope of the exceptions provided for by Art. 3 of the DSM Directive or by Directive 2001/29.

However, in order to enjoy such exclusivity in respect of their works and other subject matter, right holders are required to expressly reserve such use. This condition imposed by the legislator suggests that the solution advanced by the European Union where this general exception is concerned is to regulate TDM by imposing a formality on the right holder. In the context of copyright, some academics have expressed opinions in favour of formalizing exclusive rights in the light of digitization, as this will make for more legal certainty, enable a more balanced flow of information,[[66]](#footnote-66) and create a system better tailored for the production and distribution of works in the digital environment.[[67]](#footnote-67)

In the case of Art. 4, a mechanism of this kind is also provided for the database *sui generis* right. The impact of this new provision on regulation of the *sui generis* right may be more relevant than it appears at first sight. As anticipated in Section 3, the interference between a database and TDM activity is stronger than in the case of copyright. On the one hand, according to Art. 7.1 of Directive 96/9, it is the exclusive right of the database owner to prevent extraction of all or substantial parts of the database for any reason. On the other hand, TDM is a broad concept involving different activities in many fields, which typically correspond with the way a database is exploited. Considering that the mere consultation of a database is outside the scope of protection,[[68]](#footnote-68) apart from extractions which consist in the straightforward reproduction and appropriation of the database, almost all the other uses may fall within the exception. The exclusive right of the database owner may therefore be considerably restricted, if not voided, unless the owner expressly reserves its use. This may have the practical effect of totally reshaping the database regime towards protection through the use of formalities. However, the relationship between the database regime and the TDM exception, and the impact of the latter on the former, will need to be further investigated. It should be added that, where the right to extract and/or re-utilize insubstantial parts of a database (already mentioned in Section 4) are concerned, it would seem that the database owner’s option of reserving TDM use cannot be extended to prohibiting the exploitation of insubstantial parts, as this user right, provided for by Art. 8.1 of Directive 96/9, sets a limit to the extension of the *sui generis* rights, which apply only to substantial parts of the database. This is confirmed by the fact that Art. 15 of Directive 96/9 expressly states that arrangements under Art. 8 cannot be contractually overridden.

The above-mentioned use of formalities may create an interesting positive paradigm, also applicable beyond the context of TDM.[[69]](#footnote-69) However, there are other elements of the reform to consider that may be critical.

The European Union has created a complex legal framework with two exceptions that explicitly deal with TDM and all the previous limitations still applicable, especially the exception for temporary acts of reproduction.[[70]](#footnote-70) In many cases, it may not be easy to distinguish between different situations. The first doubt is whether the legal certainty that was promised by the reform has really been achieved.[[71]](#footnote-71) This may have a direct impact on the need to access data: the fragmentation of legal treatments dependent on many variables is conceptually and practically contrary to the smooth functioning of TDM, which requires simple rules for accessing data sets.

Another aspect of the reform is the transaction costs that a subject interested in TDM may have to bear, as this carries the risk of creating an obstacle to data access. Such transaction costs are of two kinds. The first pertains to the searching for the content to be mined, which will be more time-consuming and costly if it is necessary to distinguish between materials that can be mined free of charge and those that cannot, partly because of the technologies used. The second is represented by the need to negotiate a license if the content owner has reserved use for the purposes of TDM. Such reservation is highly predictable, especially in case of huge data sets, and is already established practice.[[72]](#footnote-72) This being the case, the bargaining power of the respective parties should also be taken into consideration. It may be difficult for many enterprises, especially start-ups and SMEs, to negotiate with owners of big data sets.[[73]](#footnote-73) The risk is that these subjects will be excluded from the field of TDM and AI, and that their efforts to innovate will be hindered. Given the importance of the initial data set required for training an AI system, in terms of its quality and quantity, there is a likelihood that the systems developed will be of lower quality if they are based on small data sets that can be more easily and cheaply accessed.

In order to at least reduce the negative effects of the first layer of transaction costs, a solution may be to provide clear, standardized indications concerning the “appropriate manner” of reserving use for TDM, which should be harmonized as much as possible among the Member States. Considering online contents, it may not be feasible for every machine to read and interpret written terms and conditions which vary considerably from website to website. Therefore, to facilitate TDM, it would be useful to use a protocol like robot.txt to create a binary “mine”/“don’t mine” rule. National legal systems need to come up with a formula that establishes a proper balance between the certainty sought by the reform and the flexibility that technological development necessarily requires.[[74]](#footnote-74)

As far as the second layer is concerned, one way of lowering transaction costs might be to issue standard licenses for the use of content covered by IP rights that were appropriate, consistent and adequately tailored for TDM.[[75]](#footnote-75) It is worth mentioning in this context that the Creative Commons (CC) licenses used to regulate the exploitation of works online by setting different combinations of pre-determined conditions are also suitable for TDM,[[76]](#footnote-76) providing those adopting such licenses with metadata[[77]](#footnote-77) that enables a machine to identify all the elements of a CC license.[[78]](#footnote-78) Furthermore, it is possible to add a CC+ protocol indicating the willingness of the owner to grant authorizations to users beyond those provided by the standard CC license, especially for commercial uses, and stating how such further permissions may be obtained.

Another element that may prove to be useful in reducing transaction costs and boosting the possibility of TDM and AI is the promotions of data pools.[[79]](#footnote-79) A potential miner may need to access many works or sets of works that belong to different right holders and are not free to mine. In these situations, the miner will have to request many licenses, an activity that may prove to be too burdensome in terms of time and costs. The creation of a sole set of works, aggregating the content of many right holders, would result in a “one-stop shop” for potential licensees.[[80]](#footnote-80) The possibility to license content for TDM would induce right holders to create high-quality data sets for commercialization and circulation. As a further step, they could be encouraged to create data pools that would be beneficial for themselves and for users.

 **7. Conclusions**

Recently, on 19 February 2020, the European Union published its “European strategy for data”,[[81]](#footnote-81) the same day as a White Paper on Artificial Intelligence was released[[82]](#footnote-82). In these documents the link between access to data and the development of AI is clearly stressed. Although there is no explicit mention of the conflict between materials covered by IP rights, on the one hand, and TDM techniques and the training of AI systems, on the other, this is an important issue when it comes to enhancing a data-driven economy. A problem that is more clearly underlined in other important documents, such as the already mentioned “Draft issues paper on intellectual property policy and artificial intelligence”, published by WIPO.

Arts. 3 and 4 of the DSM Directive introduce two exceptions for text data mining, in an attempt to clarify and solve the issue. This intervention has the positive effect of harmonizing measures among all the Member States by imposing the exceptions as mandatory. However, the complex framework deriving from the reform may not solve the issue of legal certainty.

A key element of the reform is the opt-out system provided by Art. 4, which regulates TDM by imposing formalities on right holders, thus creating an interesting paradigm for access to data but, at the same time, allowing right holders to reserve the right to mine and possibly to license their content for this purpose. Establishing a legitimate market for TDM may have the beneficial effect of encouraging right holders to create high-quality data sets. From this point of view, it may be an incentive to innovation. On the other hand, this system may give rise to excessive transaction costs for miners, especially start-ups and SMEs, acting as an obstacle to their activities and so hindering their access to data and undermining their role in a data-driven economy.

**Bibliography**

**Books**

Borghi M., Karapapa S., *Copyright and Mass Digitization*, Oxford University Press, 2013.

Lopez-Tarruella A., *Google and the Law: Empirical Approaches to Legal Aspects of Knowledge-Economy Business Models,* Springer, 2012.

* Ottolia A., *Big Data e Innovazione Computazionale*, Giappichelli, 2017.

**Essays, white papers and reports**

AI-HTEG, *A definition of AI: Main capabilities and scientific disciplines,* B-1049 Brussels, 8 April 2019.

Allgrove B., *The search engine’s dilemma: implied licence to crawl and cache?,* in *Journal of Intellectual Property Law & Practice,* 2007, Vol. 2, No. 7, p. 437 ff.

Banterle F., *Data Ownership in the Data Economy: A European Dilemma*, in Synodinou T., Jougleux P., Markou C., Prastitou T. (eds.), *EU Internet Law in the Digital Era*, Springer, 2019, p. 199 ff.

Bottis M., Papadopoulos M., Zampakolas C., Ganatsiou P., *Text and Data Mining in Directive 2019/790/EU Enhancing Web-Harvesting and Web-Archiving in Libraries and Archives,* in *Open Journal of Philosophy*, 2019, Vol. 9, p. 369 ff.

Carroll M., *Creative Commons as Conversational Copyright,* in P. K. Yu (ed*.), Intellectual Property and Information Wealth: Issues and Practises in the Digital Age*, Praeger, 2007, p. 445 ff.

* Caso R., *Il conflitto tra diritto d’autore e ricerca scientifica nella disciplina del text and data mining della direttiva sul mercato unico digitale*, Trento Law and Technology Research Group Research Paper no. 38, 2020, available at <https://zenodo.org/communities/trentolawtechgroup/>, last accessed on 25 July 2020.

* Caspers M., Guibault L. et Al., *Baseline report of policies and barriers of TDM in Europe*, in *FutureTDM,* 2016, http://www.futuretdm.eu/knowledge-library/?b5-file=4588&b5-folder=2227, last accessed on 25 July 2020.
* Clark B., *Google Image Search does not infringe copyright, says Bundesgerichtshof,* in *Journal of Intellectual Property Law & Practice*, 2010, Vol. 5, No. 8, p. 553 ff.
* Communia, *DSM Directive Implementation Guidelines*, 2019, <https://www.notion.so/DSM-Directive-Implementation-Guidelines-45233be9c0e143338860ae5a03118bf3>, last accessed 25 July 2020;
* Ducato R., Strowel R., *Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to ‘‘Machine Legibility*’’, in *IIC,* 2019, p. 649 ff.
* European Commission, *Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases*, 2018.
* European Commission, *White paper on Artificial Intelligence - A European approach to excellence and trust,* Brussels, 19.2.2020, COM(2020) 65 final.
* European Commission, *A European strategy for data,* Brussels, 19.2.2020 COM(2020) 66 final.
* Frosio G., *Sharing or Platform Urban Mobility? Propertization from Mass to MaaS*, in M. Finck, M. Lamping, V. Moscon, H. Richter (eds.), *Smart Urban Mobility,* Springer, 2020, forthcoming.

Geiger C, Frosio G., Bulayenko O., *Facilitating Wider Access to Europe’s Cultural Heritage in the Digital Environment: Opinion of the CEIPI on the European Commission’s Copyright Reform Proposal, with a Focus on Access to Out-of-Commerce Works*, Centre for International Intellectual Property Studies (CEIPI) Research Paper No. 2018-11.

* Geiger C., Frosio G, Bulayenko O., *Text and Data Mining: articles 3 and 4 of the Directive 2019/790/EU*, Centre for International Intellectual Property Studies (CEIPI) Research Paper No. 2019-08.
* Goodfellow I.J. et al., *Generative Adversarial Netwoks*, 2014, available at <https://arxiv.org/abs/1406.2661>, last accessed on 25 July 2020.
* Jasiewicz M.I., *Copyright Protection in an Opt-Out World: Implied License Doctrine and News Aggregators*, in *The Yale Law Journal*, 2012, Vol. 122, No. 3, p. 837 ff.
* Martens B., *The impact of data access regimes on artificial intelligence and machine learning*, Digital Economy Working Paper, 2018-09, JRC Technical Reports, <http://hdl.handle.net/10419/202237>, last accessed on 25 July 2020.
* Mezzanotte F., *Le «eccezioni e limitazioni» al diritto d’autore UE (parte II: Le libere utilizzazioni nell’ambiente digitale*, in *AIDA* 2017, p. 300 ff.
* Montagnani M.L., Aime G., *Il text and data mining e il diritto d’autore*, in *AIDA* 2017, p. 376 ff.
* Ottolia A., *L'utilizzo computazionale dell'opera dell'ingegno in internet*, in *AIDA* 2014, p. 386 ff.
* Pihlajarinne T., *Setting the Limits for the Implied License in Copyright and Linking Discourse – the European Perspective*, in *EIPR* 2012, p. 707 ff.
* Raue B., *Free Flow of Data? The Friction Between the Commission’s European Data Economy Initiative and the proposed Directive on Copyright in the Digital Single Market,* in *IIC,* 2018, p. 379 ff.
* Reinsel D., Gantz J., Rydning J., *The Digitization of the World - From Edge to Core*, International Data Group (IDC) White Paper, November 2018.

Ricolfi M., *Making Copyright Fit for the Digital Agenda*, in *Il Diritto d’autore*, 2011, p. 359 ff.;

* Riis T., Rognstad O. et Al., *Comment of the European Copyright Society on the Implementation of the Extended Collective Licensing Rules (Arts. 8 and 12) of the Directive (EU) 2019/790 on Copyright in the Digital Single Marke*t, in *SSRN Electronic Journal*, 11 June 2020.

Ritter J., Mayer A., *Regulating Data as Property: A New Construct for Moving Forward*, in *Duke L & Tech Rev.*, 2018, Vol. 16, p. 226 ff.

* Rosati E., *Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and Its Role in the Development of AI Creativity*, in *Asia Pac. Law Rev.*, 2019, p. 198 ff.
* Sag M., *The New Legal Landscape for Text Mining and Machine Learning*, in *Journal of the Copyright Society of the USA*, 2019, Vol. 66, p. 291 ff.
* Scalzini S., *L’estrazione di dati e di testo per finalità commerciali dai contenuti degli utenti. Algoritmi, proprietà intellettuale e autonomia negoziale*, in *Analisi Giuridica dell'Economia*, 2019, p. 395 ff.
* Sieman J.S., *Using the Implied License to Inject Common Sense into Digital Copyright,* in *North Carolina Law Review*, 2007, Vol. 85, p. 885 ff.

Sprigman C., *Reform(aliz)ing Copyright,* in *Stan. L. Rev.*, 2004, Vol. 57, p. 485 ff.

Strowel A., *Reconstructing the Reproduction and Communication to the Public Rights: How to Align Copyright with Its Fundamentals*, in Hugenholtz B. (ed.), *Copyright Reconstructed Rethinking Copyright’s Economic Rights in a Time of Highly Dynamic Technological and Economic Change*, Wolters Kluwer, 2018, p. 203 ff.

* Triaille J.P. , de Meeûs d’Argenteuil J., Francquen A., *Study on the legal framework of text and data mining (TDM*), 2014, prepared for the European Commission, <http://ec.europa.eu/internal_market/copyright/docs/studies/1403_study2_en.>, last accessed on 25 July 2020.
* Van Gompel S., *Formalities in the Digital Era: an Obstacle or Opportunity?*, in Bently L., Suthersanen U., Torremans P. (eds), *Global Copyright: Three Hundred Years Since the Statute of Anne, from 1709 to Cyberspace,* Cheltenham: Edward Elgar, 2010, p. 395 ff.
* Van Gompel S., *Copyright Formalities in the Internet Age: Filters of Protection or Facilitators of Licensing*, in *Berkeley Tech. L.J.,* 2013, Vol. 28, p. 1425 ff.
* WIPO Secretariat, *Draft Issues Paper on Intellectual Property Policy and Artificial Intelligence,* 13 December 2019.
* Zech H., *Data as a Tradeable Commodity*, in De Franceschi A. (ed.), *European Contract Law and the Digital Single Market*, Cambridge, 2016, p. 49 ff.

**Blogs, online newspapers and other webpages**

Bhageshpur K., *Data Is The New Oil -- And That's A Good Thing,* 15 November 2019, <https://www.forbes.com/sites/forbestechcouncil/2019/11/15/data-is-the-new-oil-and-thats-a-good-thing/>, last accessed on 25 July 2020.

Christie’s, *Is artificial intelligence set to become art’s next medium?,* <https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx>, last accessed on 25 July 2020.

European Parliament press release, *Free flow of non-personal data: Parliament approves EU’s fifth freedom,* 4 October 2018, https://www.europarl.europa.eu/news/it/press-room/20180926IPR14403/free-flow-of-non-personal-data-parliament-approves-eu-s-fifth-freedom, last accessed on 25 July 2020.

* Hugenholtz B., *The New Copyright Directive: Text and Data Mining (Articles 3 and 4)*, in *Kluwer Copyright Blog*, 24 July 2019, <http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/>, last accessed on 25 July 2020.

Nugent C., *The Painter Behind These Artworks Is an AI Program. Do They Still Count as Art?,* 20 August 2018, <https://time.com/5357221/obvious-artificial-intelligence-art/>, last accessed on 25 July 2020.

Stamatoudi I., *An EU exception for Text and Data Mining in the pipeline*, in *Kluwer Copyright Blog*, 11 May 2016, <http://copyrightblog.kluweriplaw.com/2016/05/11/eu-exception-text-data-mining-pipeline/>, last accessed on 25 July 2020.

The Economist, *The world’s most valuable resource is no longer oil, but data*, 6 May 2017, <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>, last accessed on 25 July 2020.

* Turkewitz N., *Sustainable Text and Data Mining: A Look at the Recent EU Copyright Directive*, 2019, [https://medium.com/@nturkewitz\_56674/sustainabletext-and-data-mining-an-look-at-the-recent-eu-copyright-directive-9ea13ba05f60](https://medium.com/%40nturkewitz_56674/sustainabletext-and-data-mining-an-look-at-the-recent-eu-copyright-directive-9ea13ba05f60), last accessed on 25 July 2020.
* Vincent J., *How Three French Students Used Borrowed Code to Put the First AI Portrait in Christie’s*, <https://www.theverge.com/2018/10/23/18013190/ai-art-portrait-auction-christies-belamy-obvious-robbie-barrat-gans>, last accessed on 25 July 2020.

**Court decisions**

A. Germany

Painter v Google 3 O 1108/05, Landgericht Erfurt, 15 March 2007.

Higher Regional Court of Jena, 27 February 2008, 2 U 319/07.

German Federal Supreme Court (Bundesgerichtshof), 29 April 2010, I ZR 69/08 (‘Vorschaubilder’).

B. European Union

CJEU, case C-203/02, British Horseracing Board Ltd v. William Hill Organization Ltd., 9 November 2004.

* CJEU, case C-5/08, Infopaq International A/S v. Danske Dagblades Forening, 16 July 2009.

CJEU, case C‑403/08 and C‑429/08, Football Association Premier League e a, 4 October 2011.

* CJEU, case C-302/10, Infopaq International A/S v. Danske Dagblades Forening, 17 January 2012.

CJEU, case C‑360/13, Public Relations Consultants Association v. Newspaper Licensing Agency, 5 June 2014.

* CJEU, case C‑30/14, Ryanair Ltd v. PR Aviation BV, 15 January 2015.

C. United States of America

Field v. Google Inc., 412 F.Supp. 2d 1106 (D. Nev. 2006).

1. See, for example, the article in The Economist, *The world’s most valuable resource is no longer oil, but data*, 6 May 2017, <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>, last accessed on 25 July 2020; or K. BHAGESHPUR, *Data Is The New Oil -- And That's A Good Thing,* 15 November 2019, <https://www.forbes.com/sites/forbestechcouncil/2019/11/15/data-is-the-new-oil-and-thats-a-good-thing/>, last accessed on 25 July 2020. [↑](#footnote-ref-1)
2. European Commission, *A European strategy for data*, Brussels, 19.2.2020, COM(2020) 66 final, p. 2. [↑](#footnote-ref-2)
3. D. Reinsel, J. Gantz, J. Rydning, *The Digitization of the World - From Edge to Core,* International Data Group (IDC) White Paper,November 2018, p. 5. [↑](#footnote-ref-3)
4. As stated in Recital 1 of Regulation (UE) 2016/679. [↑](#footnote-ref-4)
5. See Recital 13: “The proper functioning of the internal market requires that the free movement of personal data within the Union is not restricted or prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data”, or Art. 1.3: “The free movement of personal data within the Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data”. [↑](#footnote-ref-5)
6. See European Parliament press release, *Free flow of non-personal data: Parliament approves EU’s fifth freedom,* 4 October2018, <https://www.europarl.europa.eu/news/it/press-room/20180926IPR14403/free-flow-of-non-personal-data-parliament-approves-eu-s-fifth-freedom>, last accessed on 25 July 2020.  [↑](#footnote-ref-6)
7. B. Raue, *Free Flow of Data? The Friction between the Commission’s European Data Economy Initiative and the proposed Directive on Copyright in the Digital Single Market,* in *IIC* 2018, p. 380: “… without the help of automated search algorithms this data mountain shrinks to a state of electricity on a hard drive”. [↑](#footnote-ref-7)
8. The following definition of artificial intelligence, given by the High-Level Expert Group of Artificial Intelligence set up by the European Commission, explains the correlation between AI and data: “Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions”. (AI-HTEG, *A definition of AI: Main capabilities and scientific disciplines,* B-1049 Brussels, 8 April 2019). [↑](#footnote-ref-8)
9. European Commission, *White paper on Artificial Intelligence - A European approach to excellence and trust*, Brussels, 19.2.2020, COM(2020) 65, p. 16. [↑](#footnote-ref-9)
10. For examples of the application of such technologies, see E. 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*, in *Asia Pac. Law Rev.* 2019, p. 198 ff. [↑](#footnote-ref-10)
11. See, from Christie’s website, *Is artificial intelligence set to become art’s next medium?,*<https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx>, last accessed on 25 July 2020. [↑](#footnote-ref-11)
12. See the article by J. Vincent, *How Three French Students Used Borrowed Code to Put the First AI Portrait in Christie’s*, <https://www.theverge.com/2018/10/23/18013190/ai-art-portrait-auction-christies-belamy-obvious-robbie-barrat-gans>, last accessed on 25 July 2020. [↑](#footnote-ref-12)
13. I. J. Goodfellow et al., *Generative Adversarial Networks*, 2014, <https://arxiv.org/abs/1406.2661>, last accessed on 25 July 2020. [↑](#footnote-ref-13)
14. See the article by C. Nugent, *The Painter Behind These Artworks Is an AI Program. Do They Still Count as Art?,* 20 August 2018, <https://time.com/5357221/obvious-artificial-intelligence-art/>, last accessed on 25 July 2020. [↑](#footnote-ref-14)
15. B. Martens, *The impact of data access regimes on artificial intelligence and machine learning*, Digital Economy Working Paper 2018-09, JRC Technical Reports, p. 5. [↑](#footnote-ref-15)
16. M. Sag, *The New Legal Landscape for Text Mining and Machine Learning*, in *Journal of the Copyright Society of the USA* 2019, Vol. 66, p. 306. [↑](#footnote-ref-16)
17. C. Geiger, G. Frosio, O. Bulayenko, *Text and Data Mining: Articles 3 and 4 of Directive 2019/790/EU*, Centre for International Intellectual Property Studies (CEIPI) Research Paper No. 2019-08, p. 6; R. Caso*, Il conflitto tra diritto d’autore e ricerca scientifica nella disciplina del text and data mining della direttiva sul mercato unico digitale*, Trento Law and Technology Research Group Research Paper n. 38, 2020, p. 8; A. Strowel, *Reconstructing the Reproduction and Communication to the Public Rights: How to Align Copyright with Its Fundamentals*, in B. Hugenholtz (ed.), *Copyright Reconstructed: Rethinking Copyright’s Economic Rights in a Time of Highly Dynamic Technological and Economic Change*, Wolters Kluwer, 2018, p. 203; R. Ducato, A. Strowel*, Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to “Machine Legibility*”, in *IIC* 2019, p. 649; M. Borghi, S. Karapapa, *Copyright and Mass Digitization*, Oxford University Press, 2013, p. 51; M. Sag, *The New Legal Landscape For Text Mining and Machine Learning*, supra at note 16, p. 307. [↑](#footnote-ref-17)
18. WIPO Secretariat, *Draft Issues Paper on Intellectual Property Policy and Artificial Intelligence,* 13 December 2019, pp. 5-6. [↑](#footnote-ref-18)
19. These expressions recall the debate around “data propertization”; on this subject, see G. Frosio, *Sharing or Platform Urban Mobility? Propertization from Mass to MaaS*, in M. Finck, M. Lamping, V. Moscon, H. Richter (eds.), *Smart Urban Mobility, Springer*, 2020, forthcoming; J. Ritter, A. Mayer, *Regulating Data as Property: A New Construct for Moving Forward,* in *Duke L & Tech Rev.* 2018, Vol. 16, p. 226 ff.; F. Banterle, *Data Ownership in the Data Economy: A European Dilemma,* in T. Synodinou, P. Jougleux, C. Markou, T. Prastitou (eds.), *EU Internet Law in the Digital Era*, Springer, 2019, p. 199 ff.. [↑](#footnote-ref-19)
20. A. Ottolia, *Big data e innovazione computazionale*, Giappichelli, 2017, p. 33. [↑](#footnote-ref-20)
21. As stated in Art. 7.1 of Directive 96/9. [↑](#footnote-ref-21)
22. A. Ottolia, *L'utilizzo computazionale dell'opera dell'ingegno in internet*, in *AIDA* 2014, p. 386 ff. [↑](#footnote-ref-22)
23. The DSM Directive, under Recital 9, expressly states that TDM carried out “in relation to mere facts or data that are not protected by copyright” does not require authorization “under copyright law”. [↑](#footnote-ref-23)
24. CJEU, case C‑30/14, Ryanair Ltd v. PR Aviation BV, 15 January 2015. For comments on the impact of this decision see European Commission, *Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases*, 2018, p. 23 ff. [↑](#footnote-ref-24)
25. A. Ottolia, *L'utilizzo computazionale dell'opera dell'ingegno in internet*, supra at note 22, p. 388. [↑](#footnote-ref-25)
26. See Recital 33 of the Infosoc Directive. [↑](#footnote-ref-26)
27. M. Caspers, L. Guibault et Al., *Baseline report of policies and barriers of TDM in Europe*, 2016, p. 23. [↑](#footnote-ref-27)
28. CJEU, case C-302/10, Infopaq International A/S v. Danske Dagblades Forening, 17 January 2012, § 26. [↑](#footnote-ref-28)
29. CJEU, case C-5/08, Infopaq International A/S v. Danske Dagblades Forening, 16 July 2009, § 56, 57; case C‑360/13, Public Relations Consultants Association v. Newspaper Licensing Agency, 5 June 2014, § 23; and cases C‑403/08 and C‑429/08, Football Association Premier League e a, 4 October 2011, § 162. [↑](#footnote-ref-29)
30. CJEU, case C-5/08, supra at note 29, § 64. [↑](#footnote-ref-30)
31. J.P. Triaille, J. de Meeûs d’Argenteuil, A. Francquen, *Study on the legal framework of text and data mining (TDM)*, De Wolf & Partners, 2014, p. 43. [↑](#footnote-ref-31)
32. J.P. Triaille, J. de Meeûs d’Argenteuil, A. Francquen, *Study on the legal framework of text and data mining (TDM)*, supra at note 31, p. 46. [↑](#footnote-ref-32)
33. M. Caspers, L. Guibault et Al., *Baseline report of policies and barriers of TDM in Europe*, supra at note 27, p. 24. [↑](#footnote-ref-33)
34. J.P. Triaille, J. de Meeûs d’Argenteuil, A. Francquen, *Study on the legal framework of text and data mining (TDM)*, supra at note 31, p. 46. [↑](#footnote-ref-34)
35. M.L. Montagnani, G. Aime, *Il text and data mining e il diritto d’autore*, in *AIDA* 2017, p. 384; I. Stamatoudi, *An EU exception for Text and Data Mining in the pipeline*, in *Kluwer Copyright Blog*, 11 May 2016, <http://copyrightblog.kluweriplaw.com/2016/05/11/eu-exception-text-data-mining-pipeline/>, last accessed on 25 July 2020. [↑](#footnote-ref-35)
36. B. Raue, *Free Flow of Data? The Friction Between the Commission’s European Data Economy Initiative and the proposed Directive on Copyright in the Digital Single Market,* in *IIC* 2018, p. 381. [↑](#footnote-ref-36)
37. CJEU, case C-302/10, supra at note 28, §50. [↑](#footnote-ref-37)
38. I. Stamatoudi, *An EU exception for Text and Data Mining in the pipeline*, supra at note 35. [↑](#footnote-ref-38)
39. C. Geiger, G. Frosio, O. Bulayenko, *Text and Data Mining: articles 3 and 4 of the Directive 2019/790/EU*, supra at note 17, p. 14. [↑](#footnote-ref-39)
40. See for example AI tools for artists at this webpage: <https://aiartists.org/ai-generated-art-tools>. [↑](#footnote-ref-40)
41. F. Mezzanotte, *Le «eccezioni e limitazioni» al diritto d’autore UE (parte II: Le libere utilizzazioni nell’ambiente digitale*, in *AIDA* 2017, p. 309. [↑](#footnote-ref-41)
42. M. Caspers, L. Guibault et Al., *Baseline report of policies and barriers of TDM in Europe*, supra at note 27, p. 35. For a report on the different systems of private copying levies provided by Member States, see H. Wijminga, W. Klomp, M. van der Jagt, J. Poort, *International Survey on Private Copying,* WIPO Report, 2016. [↑](#footnote-ref-42)
43. Μ. Bottis, M. Papadopoulos, C. Zampakolas, P. Ganatsiou, *Text and Data Mining in Directive 2019/790/EU Enhancing Web-Harvesting and Web-Archiving in Libraries and Archives,* in *Open Journal of Philosophy* 2019, Vol. 9, p. 388. [↑](#footnote-ref-43)
44. See Art. 5.3 a) of the Infosoc Directive for copyright; Art. 6.2 b) and Art. 9 b) of Directive 96/9 respectively for databases protected by copyright and by *sui generis* rights. [↑](#footnote-ref-44)
45. Recital 12 of Directive 2019/790. Recital 11 also extends the exception to research activities carried out in the context of public-private partnerships. [↑](#footnote-ref-45)
46. For a comment on the asymmetry between Art. 3 and Art. 4, see R. Caso, *Il conflitto tra diritto d’autore e ricerca scientifica nella disciplina del text and data mining della direttiva sul mercato unico digitale*, Trento LawTech Research Paper no. 38, 2020, p. 9. [↑](#footnote-ref-46)
47. See the definitions given in Recitals 14 and 18 of the DSM Directive and the comments of B. Hugenholtz, *The New Copyright Directive: Text and Data Mining (Articles 3 and 4)*, in *Kluwer Copyright Blog*, July 24th 2019, <http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/>, last accessed on 25 July 2020; Μ. Bottis, M. Papadopoulos, C. Zampakolas, P. Ganatsiou, *Text and Data Mining in Directive 2019/790/EU Enhancing Web-Harvesting and Web-Archiving in Libraries and Archives*, in *Open Journal of Philosophy* 2019, p. 388. [↑](#footnote-ref-47)
48. C. Geiger, G. Frosio, O. Bulayenko, *Text and Data Mining: articles 3 and 4 of the Directive 2019/790/EU*, supra at note 17, p, 27; S. Scalzini, *L’estrazione di dati e di testo per finalità commerciali dai contenuti degli utenti. Algoritmi, proprietà intellettuale e autonomia negoziale*, in *Analisi Giuridica dell'Economia* 2019, p. 416. [↑](#footnote-ref-48)
49. For a comment on these provisions, see C. Geiger, G. Frosio, O. Bulayenko, *Facilitating Wider Access to Europe’s Cultural Heritage in the Digital Environment: Opinion of the CEIPI on the European Commission’s Copyright Reform Proposal, with a Focus on Access to Out-of-Commerce Works*, Centre for International Intellectual Property Studies (CEIPI) Research Paper No. 2018-11; Riis T., Rognstad O. et Al., *Comment of the European Copyright Society on the Implementation of the Extended Collective Licensing Rules (Arts. 8 and 12) of Directive (EU) 2019/790 on Copyright in the Digital Single Marke*t, in *SSRN Electronic Journal*, 11 June 2020. [↑](#footnote-ref-49)
50. Art. 8.4 extends such provision to the exceptions contained in Art. 8.2 pertaining to cases in which there is no representative CMO that can issue the ECL. [↑](#footnote-ref-50)
51. N. Turkewitz, *Sustainable Text and Data Mining: A Look at the Recent EU Copyright Directive*, 16 May 2019, [https://medium.com/@nturkewitz\_56674/sustainabletext-and-data-mining-an-look-at-the-recent-eu-copyright-directiveea13ba05f60](https://medium.com/%40nturkewitz_56674/sustainabletext-and-data-mining-an-look-at-the-recent-eu-copyright-directiveea13ba05f60), last accessed on 25 July 2020. [↑](#footnote-ref-51)
52. “*[…]* *In the case of content that has been made publicly available online, it should only be considered appropriate to reserve those rights by the use of machine-readable means, including metadata and terms and conditions of a website or a service. Other uses should not be affected by the reservation of rights for the purposes of text and data mining. In other cases, it can be appropriate to reserve the rights by other means, such as contractual agreements or a unilateral declaration […]*”. [↑](#footnote-ref-52)
53. The exception is regulated in the Italian legal system by Art. 65.1 Law 633/1941 and Art. 7 of the “Regolamento per l’esecuzione della legge 22 aprile 1941, n. 633”. [↑](#footnote-ref-53)
54. B. Hugenholtz, *The New Copyright Directive: Text and Data Mining (Articles 3 and 4),* supra at note 47. [↑](#footnote-ref-54)
55. For more information about the functioning of these files see <http://www.robotstxt.org/robotstxt.html>. [↑](#footnote-ref-55)
56. Unlike a robots.txt file, which must be stored in the root folder of the server on which the pages reside, the use of meta-tags makes it possible to specify instructions for every single web page; for more information see J. S. Sieman, *Using the Implied License to Inject Common Sense into Digital Copyright,* in *North Carolina Law Review* 2007, Vol. 85, p. 908. [↑](#footnote-ref-56)
57. For a view of implied licenses as opt-out systems, see M. I. Jasiewicz, *Copyright Protection in an Opt-Out World: Implied License Doctrine and News Aggregators*, in *The Yale Law Journal* 2012, Vol. 122, p. 837 ff. and J. S. Sieman, *Using the Implied License to Inject Common Sense into Digital Copyright*, supra at note 56, p. 889. [↑](#footnote-ref-57)
58. Field v. Google, Inc., 412 F.Supp. 2d 1106 (D. Nev. 2006). For a comment on this decision, see J. S. Sieman, *Using the Implied License to Inject Common Sense into Digital Copyright*, supra at note 56, p. 906 ff. [↑](#footnote-ref-58)
59. Painter v Google 3 O 1108/05, Landgericht Erfurt, 15 March 2007. For comment, see B. Allgrove, *The search engine’s dilemma: implied licence to crawl and cache?* in *Journal of Intellectual Property Law & Practice* 2007, Vol. 2, p. 437 ff. [↑](#footnote-ref-59)
60. Higher Regional Court of Jena, 27 February 2008, 2 U 319/07. [↑](#footnote-ref-60)
61. German Federal Supreme Court (Bundesgerichtshof), 29 April 2010, I ZR 69/08 (‘Vorschaubilder’). For comment on the case, see B. Clark, *Google Image Search does not infringe copyright, says Bundesgerichtshof,* in *Journal of Intellectual Property Law & Practice* 2010, vol. 5, p. 553 ff.; A. Lopez-Tarruella, *Google and the Law: Empirical Approaches to Legal Aspects of Knowledge-Economy Business Models,* Springer, 2012, p. 185 ff. [↑](#footnote-ref-61)
62. On the limits and the lack of harmonization of the doctrine of implied licenses in the context of linking, see T. Pihlajarinne, *Setting the Limits for the Implied License in Copyright and Linking Discourse – the European Perspective*, in *EIPR* 2012, p. 707 ff. [↑](#footnote-ref-62)
63. For an analysis of the relationship between platforms and user in the light of the new provisions, see S. Scalzini, *L’estrazione di dati e di testo per finalità commerciali dai contenuti degli utenti. Algoritmi, proprietà intellettuale e autonomia negoziale*, supra at note 48, 395 ff. [↑](#footnote-ref-63)
64. C. Geiger, G. Frosio, O. Bulayenko, *Text and Data Mining: articles 3 and 4 of Directive 2019/790/EU*, supra at note 17, p. 36. [↑](#footnote-ref-64)
65. B. Hugenholtz, *The New Copyright Directive: Text and Data Mining (Articles 3 and 4),* supra at note 47. [↑](#footnote-ref-65)
66. S. Van Gompel, *Formalities in the Digital Era: an Obstacle or Opportunity?,* in L. Bently, U. Suthersanen, P. Torremans (eds), *Global Copyright: Three Hundred Years Since the Statute of Anne, from 1709 to Cyberspace*, Edward Elgar, 2010, p. 395 ff. [↑](#footnote-ref-66)
67. M. Ricolfi, *Making Copyright Fit For the Digital Agenda*, in *Il Diritto d’autore* 2011, p. 359 ff. [↑](#footnote-ref-67)
68. CJEU, case C-203/02, British Horseracing Board Ltd v. William Hill Organization Ltd., 9 November 2004, § 54. [↑](#footnote-ref-68)
69. However, in the context of copyright, the use of formalities may be in violation of Art. 5 of the Berne Convention, which does not allow this type of intellectual property to be subject to formalities. See S. Van Gompel, *Copyright Formalities in the Internet Age: Filters of Protection or Facilitators of Licensing*, in *Berkeley Tech. L.J*. 2013, Vol. 28, p. 1429. [↑](#footnote-ref-69)
70. According to Recital 9 of the DSM Directive, this exception should continue to apply to TDM techniques. [↑](#footnote-ref-70)
71. See Recitals 11 and 18 of the DSM Directive. [↑](#footnote-ref-71)
72. For an overview on the general trend towards contractually banning TDM, see R. Ducato, A. Strowel*, Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to “Machine Legibility*”, supra at note 17, p. 674. [↑](#footnote-ref-72)
73. Regarding the exclusion of start-ups under the reform, see C. Geiger, G. Frosio, O. Bulayenko, *Text and Data Mining: articles 3 and 4 of Directive 2019/790/EU,* supra at note 17, p. 36. [↑](#footnote-ref-73)
74. Regarding the need to indicate the use of robot.txt files, see Communia, *DSM Directive Implementation Guidelines*, <https://www.notion.so/DSM-Directive-Implementation-Guidelines>, last accessed on 25 July 2020. [↑](#footnote-ref-74)
75. Regarding standard licenses for the re-use of data, see H. Zech, *Data as a Tradeable Commodity*, in A. De Franceschi (ed.), *European Contract Law and the Digital Single Market*, Intersentia, 2016, p. 49 ff. [↑](#footnote-ref-75)
76. M. Caspers, L. Guibault et AL., *Baseline report of policies and barriers of TDM in Europe*, supra at note 27, p. 71 ff. [↑](#footnote-ref-76)
77. For the use of Creative Commons licenses as a model for new formalities in the context of copyright, see S. van Gompel, *Copyright Formalities in the Internet Age: Filters of Protection or Facilitators of Licensing*, supra at note 69, p. 1436; C. Sprigman, *Reform(aliz)ing Copyright,* in *Stan. L. Rev.* 2004, Vol. 57, p. 564. [↑](#footnote-ref-77)
78. See M. Carroll, *Creative Commons as Conversational Copyright,* in P. K. Yu (ed*.), Intellectual Property and Information Wealth: Issues and Practises in the Digital Age*, Praeger, 2007, p. 445 ff. [↑](#footnote-ref-78)
79. European Commission, *White paper on Artificial Intelligence - A European approach to excellence and trust*, Brussels, 19.2.2020, COM(2020) 65 final, p. 3. [↑](#footnote-ref-79)
80. For a detailed analysis of data pools, see A. Ottolia*, Big Data e innovazione computazionale*, supra at note 20, p. 272 ff. [↑](#footnote-ref-80)
81. European Commission, *A European strategy for data*, Brussels, 19.2.2020 COM(2020) 66. [↑](#footnote-ref-81)
82. European Commission, *White Paper On Artificial Intelligence - A European approach to excellence and trust,* Brussels, 19.2.2020 COM(2020) 65. [↑](#footnote-ref-82)