



CATÓLICA
FACULDADE DE DIREITO

ESCOLA DE LISBOA



CATÓLICA
Global
School of
Law

MINING DIGITAL TREASURES

**Text and Data Mining interfering with Copyrights in the EU and in the US: which
legal system is more favourable to TDM users?**

**Master Thesis carried out in the context of the
Master of Transnational Law
under the supervision of Dr. Tito Rendas**

Denise da Cruz Fernandes

Date of submission: 30-06-2020

ABBREVIATIONS

Database Directive = Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases

DCDSM = Directive 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

E&L = Exceptions and limitations

EC = European Commission

EC's Proposal for a Copyright Directive = Proposal for a Directive Of The European Parliament And Of The Council on copyright in the Digital Single Market COM/2016/0593 final - 2016/0280 (COD)

European Courts (CJEU, ECJ and National Courts)

InfoSoc Directive = Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society, Official Journal L 167 22.6.2001 p.10-19

MS = Member State / MSs = Member States

OJEU = Official Journal of the European Union

PPPs = public-private partnerships

SDS = smart disclosure systems

SME = small and medium enterprise

SWD = Staff Working Document

TDM = Text and Data Mining

TPMs = technological protection measures

TRIPS = The World Trade Organization's agreement on Trade Related Aspects of Intellectual Property

USC = United States Code

PL = Public Law

KEY-WORDS: text and data mining, intellectual property, copyrights, fair use, database
suis generis rights

1 - INTRODUCTION

The “fourth industrial revolution” is in the making because people create big amounts of data every hour.¹ The IDC predicts that the Global Datasphere will grow from 33 zettabytes in 2018 to 175 zettabytes by 2025, i.e. an increase of 156 trillion gigabytes of global data volume.²

In the era of Big Data, collecting, processing and transforming those large data sets has increasingly become easier, faster, and cheaper. Those tasks have become a fundamental basis of competition³ and a source of richness, in the form of both the discovery of knowledge and the creation of capital, mainly due to the proliferation of devices with mechanisms that help their human counterparts with it.

Text and Data Mining (or TDM) has been used as a broad term to describe various mechanisms that are able to interpret, analyse and organize any type of content, for the most diverse purposes. “Only a tiny fraction of the digital universe has been explored for analytic value. IDC estimates that by 2020, as much as 33% of the digital universe will contain information that might be valuable if analyzed.”⁴

However, the use of TDM algorithms can interfere with the protection of intellectual property rights, especially copyrights and database rights, depending on the jurisdiction.

¹ Ménière (2017) p. 3

² Reinsel (2018) p. 3

³ Manyika (2011) p. 97

⁴ Gantz (2012) p. 2

Legislators often create different legal provisions to protect copyrighted works from TDM infringement, as well as TDM use protection against claims of copyright infringement. Accordingly, certain legal regimes are considered more restrictive than others where it concerns authorizations for TDM use versus the protection of copyrights. In this case, the EU legal framework is constantly highlighted in literature as too prohibitive for TDM users in several instances, especially in comparison with the US legal system, but is it really?

At first sight, one could be tempted to point out “obvious” differences between the two legal systems. Differently from the EU’s closed system of broad copyrights and strict enumerated TDM exceptions and limitations, now on articles 3 and 4 of the DCDSM, the US legal system appears more flexible in times of technological change and thus more favourable to those using TDM technologies due to the application of the “fair use” open norm present in §107 US Copyright Act. Moreover, the constant development of a broad scope of the term “fair use” through judicial application and relevant case law regarding this matter where the acceptance of “public interest in the access to information” is a significant factor weighing in favour of the fair use defence, reinforces the idea of a more flexible (yet somewhat uncertain) US Copyright regime. On the other side, the EU has a past of inexistent TDM exceptions until recently, with Article 3 of the DCDSM containing a TDM exception to the exclusive rights of reproduction and extraction where the beneficiaries of the exception must belong to certain categories and mine for a certain purpose; and the exception present in article 4 of the DCDSM that, though not as effective in protecting TDM users due to how easy it is to override by rightsholders, was seemingly created in order to protect the entities and help the development of economic sectors that were not covered by the scope of the TDM exception in Article 3, such as Artificial Intelligence and commercial data analytics.

In order to delimit the object of this thesis, this document will be focused on the limitations to TDM created by copyrights and e&l in the EU, as well as the US fair use clause and its interpretation in the American case law.

2 - RESEARCH QUESTIONS

1. How can the formulation of articles 3 and 4 of the DCDSM contribute to achieve the objectives of that Directive?
2. Is the US Copyright legal framework more favourable to TDM users compared with the EU's?

The choice to analyse these research questions in this thesis, in light of these two particular and very different legal systems, arose after coming across various documents concluding that the US legal system had been, at least until the reform of the DCDSM, more favourable to TDM users and researchers, when compared to the EU. In 2018, among one of the official studies requested by the European Parliament's JURI Committee, one phrase stood out: "the US copyright regime is considered more favorable to TDM practices than what appears to be the case under European laws"⁵. According to this study, legal uncertainty and market fragmentation resulting from the different territorial scope of the copyright exceptions and limitations, implemented differently in each Member State, reportedly encouraged researchers that were based in the European Union to outsource⁶ their content mining to the United States, therefore depriving the EU of every advantage that came with their research projects being based in one of the Member States. Thus, the DCDSM was envisioned by the EP and the EC towards problem-solving objectives such as: a higher degree of horizontal harmonization; a reduction of the need for a substantial body of case-law to be settled before the CJEU; augmented growth, sustainable job creation and the competitiveness of the EU economy; not stifling innovation coming from different sectors; reducing rights clearance costs for research organisations without affecting the right holders' subscription market, among others.

Laws and case law connected with "technological uses" of copyrighted content will be considered in this thesis to evaluate how the new TDM exceptions can contribute (or not)

⁵ Rosati (2018) p. 7

⁶ Filippov (2016) p. 4

towards those objectives in the EU and favour TDM users, and compare those results with the ones obtained through the application of the fair use doctrine in the US, to understand which body of Copyright law and case law (legal framework) can present more advantages for TDM users.

3 – TDM: WHAT AND HOW

3.1. THE DEFINITION OF TDM

Text and Data Mining (TDM) is a topic that is becoming more relevant nowadays due to the large amounts of data increasing exponentially by the day, and the advantages that the efficient discovery, analysis and application of the information that can be obtained from that big data can bring for individuals, companies of various sectors and even governments of the world.

Specialists have mentioned that TDM is realistically an umbrella term to describe various, related but different, forms of technologies, with no evident dominance between them, used to process text and data. Therefore, making it difficult to describe it precisely. At least one approach suggests that TDM can be divided in seven practice areas⁷ and others emphasizes that TDM should not be confused with AI or machine learning.⁸

In the midst of the EU acquis, the definition of TDM can be found in article 2 (2) of the DCDSM as “any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations”.

There are several references to TDM in the U.S. Code, for instance, considering it “a program involving pattern-based queries, searches, or other analyses of 1 or more electronic databases [used to] discover or locate a predictive pattern or anomaly indicative

⁷ See Miner (2012) p. 31-32

⁸ Flynn (2020), p. 2

of terrorist or criminal activity on the part of any individual or individuals”⁹, but also to “access, receive, and analyze data and information (...) and to disseminate information”¹⁰ and as a part of the “policies, procedures, and plans to improve program accountability and integrity through targeted and coordinated activities, (...) to identify and reduce errors, waste, fraud, and abuse in programs”¹¹, among others¹².

For purposes of this document, text and data mining will be considered as any automated software mechanisms used to process digital content (including structured and unstructured data¹³) to extract information from it.

In order to analyse Big Data, different TDM techniques can be used so as to obtain the information patterns and the new trends and other correlations between the data. This way, through TDM techniques, users can to scour millions of articles in one go and increase the amount of articles they scour as well¹⁴. Studies show that TDM facilitates the identification of relevant content with just 25% of the manual endeavor necessary to identify them otherwise, and reduces the researcher’s manual systematical review of that content by 75%.¹⁵¹⁶

⁹ 42 U.S. C. § 2000ee-3 (b)(1)

¹⁰ 6 USC § 121(d)(11)

¹¹ 7 USC § 6932 note (2018)

¹² See also 6 USC § 652(e)(1)(L); 6 USC 542 plan elements (2)(B)(1)(b)(2); 7 USC § 1515 (j) (2); 7 USC § 2036b (b)(2)

¹³ Structured data consists of data types that are precisely determined, with patterns that allow them to be readily searched; all the data that is not structured can be regarded as unstructured data, this data is usually more difficult to search and exists in different formats such as audio, video, and social media postings.

¹⁴ McDonald (2012) p.27

¹⁵ Thomas (2011)

¹⁶ Ibid note 16

TDM has been proven essential for various uses. For example, BlueDot¹⁷ was first to publish a scientific paper predicting the global spread of COVID-19 after using their outbreak risk software to analyze patterns between travels and health outbreaks along with “a variety of information sources, including chomping through 100,000 news reports in 65 languages a day”¹⁸. Among other uses, TDM can be used to help: vaccine researchers to mine scientific journals about the coronavirus to find a cure¹⁹; healthcare providers to extract data from patient records to make treatment decisions²⁰; medical professionals access information about prescription medicine²¹; researchers to describe or map new cases as well as to test and validate hypothesis between existing theories²²; public administrations to reuse information about their citizens, instead of requesting them again²³; use AI to filter “real news” from “fake news”²⁴; give consumers better access to information they need to make informed choices though smart disclosure systems (SDSs) that examine problems of transparency, bias, and even discrimination²⁵.

Nevertheless, as it was previously mentioned, depending on the legal system in question, the application of TDM technologies can interfere with the protection of intellectual property, in various stages of its process.

3.2. TDM PROCESS

¹⁷ <https://bluedot.global/> Last visited 29/06/2020

¹⁸ Prosser (2020) and Stieg (2020)

¹⁹ Knight (2020)

²⁰ Shu (2018)

²¹ Kelly (2016) p. 131–147

²² Singh N (2007) pp. 131–147 and Yarkoni (2010) p. 371

²³ ISA² (2019) p. 6

²⁴ Lunden I (2019)

²⁵ Sunstein (2012)

First, the data collector(s) or data analyst(s) access the content, compile it and store it initially, to analyse it afterwards. This can be called the data collection phase. As a preparatory act for TDM, those who collect the data must select the initial materials to work with after obtaining lawful access to that content (e.g. structured/unstructured data, complete/portions of works, obtained from individual sources/databases) which may be freely accessible or only accessible through a license.²⁶ There are different ways to collect this data: using specialized hardware like a sensor network, software tools like Web document crawling to collect documents, and even manual labour e.g. to collect information from user surveys. Normally, the collected data is stored in a database or a data warehouse for further processing²⁷.

Secondly, comes the data pre-processing phase. The collected data may be proper for immediate processing or not. In case it is not, the data mining analyst may need to extract, select and transform features from the gathered data and to clean that data. Disperse data mixed in free-form documents or data that is encoded in complex logs is difficult or rather impossible to process directly after collection. The data coming from raw documents, commercial transactions or system logs, must first be transformed into a format that is compatible with data mining algorithms. A mining analyst skilled in her art will extract several features from raw data that are meaningful, and only those that are meaningful, for the application of the mining algorithm(s) she chose for a specific case. If there are missing or erroneous entries in the extracted data, the analyst will also have to estimate that data or to correct it.²⁸

Extraction, selection, transformation and cleaning of data may occur in this phase. Consequently, certain information may be necessarily copied (temporarily or not) from those materials, for which a specific rightsholder authorization may be required, according to EU law. Exclusive reproduction/extraction rights may be infringed where the data is transformed into a machine-readable file or the content is uploaded to a certain platform to become compatible with the chosen TDM technology.

²⁶ Idem note 6 pp. 4-7

²⁷ Aggarwal (2015) p. 3

²⁸ Ibid note 28

Thirdly, comes the analytical phase. In order to reach a certain objective or obtain a certain output with the data they collected and extracted, the data analysts can either create from scratch or apply an already existing analytical method (also called a data mining problem)²⁹ in its entirety, or as a component of a new algorithm they design on their own³⁰. Thus, the analysts will choose the TDM method they find most appropriate to answer the research question or reach the objective. Broad interpretations of exclusive rights and strict interpretations of E&L with conditions about the quality of the actor or the purpose of the mining can vastly reduce the choices available to the EU data analysts and, thus, their chances of efficiently and effectively reaching their objective, when competing with their American counterparts.

Finally, depending on the algorithm(s), the data analysts might need to do some additional feature selection/removal, for the algorithm to work efficiently, or transform a group of data with a particular set of attributes into a data group with another set of attributes of the same or of a different category.

At last, the output is completed and, in principle, revealed to the world.

²⁹ E.g. if an analyst has the objective to identify leadership themes within the statements presented by a company, she can apply a TDM technique previously used to transform news into news topics. See Radev D pp. 95-98

³⁰ Ibid note 28 p. 4

4. COPYRIGHT AND TDM USE IN THE EU

4.1. EU LEGAL FRAMEWORK BEFORE THE DCDSM (LAW & CASE LAW)

Departing from the existent case law and application of EU copyright/database norms by the EU courts³¹, several acts in the process of TDM use could infringe different exclusive rights given to copyright rightholders and database makers. The Infopac cases³² seem to be the closest the CJEU has gotten to decide over a matter of “TDM use vs. EU exclusive rights”.

The exclusive rights of rightholders considered for the purpose of this document will be: the reproduction right³³, right of communication and making available to the public³⁴, distribution right³⁵ and the extraction right for database makers³⁶.

Before the new DCDSM exceptions for TDM were created, a list of twenty optional limitations and one mandatory exception³⁷ to the different exclusive economic rights of the author’s existed in Article 5 of the EU’s InfoSoc Directive.

For acts of reproductions to fall under the mandatory exception to the reproduction right, they had to fulfil five cumulative conditions, namely: being temporary and

³¹ For the purpose of this thesis, the “EU courts” to be considered are: the CJEU, the ECJ and the national courts of the MSs

³² C-5/08 (Infopaq I) and C-302/10 (Infopaq II)

³³ Article 2 Infosoc Directive

³⁴ Article 3 Infosoc Directive

³⁵ Article 4 Infosoc Directive

³⁶ Article 7(1) Database Directive

³⁷ i.e. the temporary acts of reproduction exception from reproduction rights. Article 5(1) Infosoc Directive

transient/incidental; being an integral and essential part of a technological process; the sole purpose of that process being to enable a transmission in a network between third parties by an intermediary of a lawful use of a work or protected subject-matter; and the act having no independent economic significance.³⁸

The list of E&L was exhaustive and, according to Recital 32 of that Directive, the MSs could not enact other exceptions besides those. However, since the limitations were not mandatory, and the Directive allowed for a considerable amount of leeway regarding the transposal of the limitations, the legislators of the MSs took very different approaches implementing them.

In the national sphere, some MSs like the UK created a specific TDM exception, allowing people to lawfully use text and data analysis for non-commercial research, whilst other MSs decided to adopt no exception nor limitation regarding this matter.

Nevertheless, these exceptions did not offer effective rights for EU TDM users because no TDM use exception nor limitation to copyright was explicitly mandatory for the EU MSs to implement in their legal systems.³⁹ Copyright law did not protect explicitly against the use of TDM in itself. Mining conditions were not expressly listed, everything could be mined so as long as it did not infringe the Intellectual Property rights protected in the EU or if it fell within the scope of the article 5's E&L.

Depending on the protection of the source content and the different methods applied in the midst of the TDM phases, infringement could occur or not. Herein lied the importance of the TDM process for the infringement of IP rights.

According to the EU legal framework before the DCDSM, copyright violations by TDM use could potentially be found in every phase of the TDM process.

³⁸ Article 5(1) and Recital 33 of the InfoSoc Directive; C-5/08 Infopaq I para. 54

³⁹ Geiger, Frosio and Bulayenko (2018) p.8

First, regarding data collection phase, TDM could only be freely used on content unprotected by copyright, thus requiring no use authorization, or whenever that use fell within the scope of the mandatory exception for temporary acts of reproduction provided in Article 5(1) of Directive 2001/29/EC⁴⁰ (and eventually any other limitation to those rights permitted by article 5 of that Directive and transposed by a particular MS).

During this first phase, there could not be illegal access to protected content⁴¹, reproductions of protected content outside the scope of art. 5(1)'s exception (or other limitations transposed by the MSs), nor illegal distribution⁴².

The exception for acts of temporary reproduction was the sole mandatory exception from the Infosoc Directive and it is explicit in Recital 33 of the Infosoc Directive that the exception was meant to enable acts such as browsing and caching, as long as the intermediary did not modify the information and did not interfere with the lawful⁴³ use of technology to obtain data.

This exception was vastly interpreted by the European Courts, and in many decisions the Courts considered that the exception could encompass certain TDM actions, as it will be shown in this document.

⁴⁰ Recital 9 DCDSM

⁴¹ E.g. not searching for the corresponding rightholders and accessing the content without paying the respective license

⁴² e.g. if the data collector stores copyrighted content in a database with the purpose to transfer its ownership or part of it later on to another data collector. In Case C-572/17 para. 36 the CJEU clarified that, for there to be distribution infringement, national courts must establish if the purpose of the storage of copyrighted content (e.g. if the content was stored to be sold or transmitted afterwards, or otherwise). Right holders can enforce their right, even if the content is merely stored and yet not transmitted.

⁴³ The use was considered lawful when it was authorized by the rightholders or not restricted by law

In the Infopac I case, the CJEU considered that Infopaq's data capture process required five phases⁴⁴ and that some acts practiced in those phases affected certain content protected by the reproduction right and others did not. In this case Infopaq's use did not fall within Article 5(1)'s exception for acts of temporary reproductions, because "the last act in the data capture process at issue in the main proceedings, during which Infopaq prints out the extracts of 11 words, is not a transient act"⁴⁵ and due to that the CJEU decided that it was not necessary to consider if the other four acts/phases (which make up the data capture process) fulfilled the other conditions laid down in Article 5(1).⁴⁶

In Infopaq II, the CJEU clarified that for an act to fall within the scope of art. 5(1) it could not allow the creation of additional profit nor the modification of the work⁴⁷. The court emphasized again that the conditions to consider an act as an exception to copyright should be interpreted strictly because Article 5(1) is a derogation from a general principle of that directive that requires the rightholders authorization for any reproduction of a protected work. This CJEU decision also showed that whenever MSs legislation falls within the scope of the exemption for temporary purposes there is no need to analyze explicitly the three step test, because the text of the provision of article 5(1) had already taken into account all these steps.⁴⁸

However, in the Premier League case, acts of reproductions through satellite decoders and television screens did fall within scope of the exception for acts of temporary reproduction, an exception which the CJEU decided to not interpret in a strict manner in

⁴⁴ C-5/08 , Infopaq I, para. 16

⁴⁵ C-5/08 , Infopaq I, para. 70. According to para. 64, an act will be transient "only if its duration is limited to what is necessary for the proper completion of the technological process in question, it 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."

⁴⁶ C-5/08 , Infopaq I para. 73

⁴⁷ C-302/10, Infopaq II para. 54

⁴⁸ C-302/10, Infopaq II para. 56

order to enable its effectiveness and allow the observance of its purpose⁴⁹. The exception had to ensure the operation of new technologies and “safeguard a fair balance between the rights and interests of right holders, on the one hand, and of users of protected works who wish to avail themselves of those new technologies, on the other.”⁵⁰ Otherwise, if the exception were to be interpreted strictly, the consumers could have been deprived from receiving their broadcasts.

Academics have defended that the reproduction must involve the use of work “as a work”. If an actor uses a TDM process that involves acts of reproduction (e.g. to derive information or to check plagiarism or mistakes) and in that TDM process the work is a mere input for searching and identifying occurrences, and its original expressive features are not being used nor enjoyed by a public, that “use as a work” is seemingly not there.⁵¹

Additionally, the scope of protection of copyright law does not include words, considered in isolation, but the expression of the authors original intellectual creation within the copyrighted works or subject matters. “Through the choice, sequence and combination of those words (...) the author may express his creativity in an original manner and achieve a result which is an intellectual creation”⁵². It was considered in that decision that the author’s own intellectual creation could be expressed in 11 consecutive words, and that the national courts should determine “if that extract contains an element of the work which, as such, expresses the author’s own intellectual creation.”⁵³ Hence, EU IP law protects works or subject matters that encompass the original idea(s) expressed by the author(s) and copyrights violations will only occur if the process of TDM use goes beyond small-scale actions such as copying very small amounts of words or processing items individually without affecting the author’s intellectual creation.

⁴⁹ C-403/08 and C-429/08, Premier League, para. 163

⁵⁰ Ibid note 55 para. 164

⁵¹ Hugenholtz PB (2018) p. 227

⁵² C-5/08 Infopaq, para. 45

⁵³ Case C-5/08 Infopaq, para. 48

During the second phase, regarding data pre-processing (and in the final phase), when the collected data needs to be cleaned, extracted, selected and transformed into features, and afterwards uploaded somewhere, in order to become machine-readable, a breach of the rights of reproduction⁵⁴, extraction⁵⁵ and even the right to make adaptations and arrangements⁵⁶ might occur, because researchers need to store source data, as well as link them to their peers to allow them to confirm the TDM research results. This necessity is of course not protected by the temporary reproduction exception⁵⁷.

Depending on the chosen mining software, the methods used in the analytical phase, can trigger different exclusive rights infringement as well.

Finally, a part of the TDM output generally consists in descriptions of the results of the TDM analysis (e.g. through aggregate data, reports). This phase might trigger reproduction rights infringement if the output contains the whole or extracts of the work or database, as well as to a communication to the public. However, this is not a necessary occurrence in TDM outputs.

Before and after the DCDSM, the EU Copyright legal framework has been criticized for containing a broad set of copyrights and a closed list of narrow exceptions to those several rights, especially when compared with other legal systems that rely on open clauses, such as the Fair Use standard. Theoretically, it has been assumed that the EU framework is too strict to allow the judges of the European courts enough discretion to apply the law in accordance with the advancements of technology and refrain from making obsolete

⁵⁴ E.g. due to additional unauthorized copies of parts of the content

⁵⁵ However, the database creator cannot forbid a legitimate TDM user from extracting small/insubstantial parts of it in accordance with Articles 8(2) and (3) of the Database directive

⁵⁶ Articles 12 and 14 of the Berne Convention. In EU law, the right to make adaptations and arrangements has sometimes been regarded as an expression of the general right of reproduction. Infraction of this right might happen during the transformation of the features and latter publication of the transformed result.

⁵⁷ C-5/08 Infopaq I, paras 62 and 66

decisions. Because of this, it is conventionally thought that those Courts cannot decide flexibly nor consider as non-infringing the new uses of copyrighted works unless they are included in the closed catalogue of copyrights exceptions and definitions presented in the EU acquis and in accordance with the previous case law of the CJEU and ECJ.

Nevertheless, empirically, the decisions of these Courts regarding the new technological uses of copyrighted works often do not seem to reflect the theoretical discretion struggles attributed to the strict EU legal framework. “However, judgments like *Vorschaubilder*, *Megakini*, *Premier League*, and *GS Media* shed doubt on this established idea. They suggest that courts are sometimes willing to strip off the straitjacket, in order to deliver what they think is the most reasonable judgment in the circumstances at hand.⁵⁸” It seems that new technological uses are being considered as non-infringing to Copyright more often, and not only has the CJEU been able to circumvent the strictness of the broad scoped rights and narrow scoped exceptions through its flexible interpretation of the relevant legal precepts, but national courts have also adopted the three-step test to allow the judges “a way out” to adapt their decisions to the new uses of these works.

Accordingly, scholars have also defended that EU Courts every so often reach flexible decisions through non-deductive arguments.

Lasser has previously noted that, besides American judges, European judges also use non-deductive arguments to decide their cases. On the one hand, the U.S. Supreme Court opinions do not use the term “formalism” often and the decisions reveal interpretive disagreement while maintaining a certain degree of interpretive openness.⁵⁹ On the other hand, the ECJ decisions can also reveal interpretive disagreements however these end up

⁵⁸ Rendas T (2017)

⁵⁹ “The American judicial system does not possess or deploy some mode of hidden internal, or “behind-the-scenes” argument that differs significantly from its official and public form. [The ECJ] functions through a bifurcated discursive form [and] adopts a more argumentative approach whose highly personal and discursive American mode of legitimation the ECJ also cannot begin to reproduce. Lasser 2004, pp.17-23, 158-159

crushed with a “remarkably self-confident institutional judgment⁶⁰”, and this happens partly when that Court follows the opinions of Advocates General who reconstruct case law through purpose driven, non-deductive, interpretation.⁶¹ A parallel could be made in the GS Media decision, between the opinion of Advocate General Wathelet who defended it was “common knowledge” and “particularly obvious” that as a general rule the internet users are not aware and do not have the means to inspect if the initial communication of a protected work on the internet was posted with the copyright holder’s consent or not, and the decision of the ECJ that “followed” his opinion after stating that that sort of inspection can only be expected from persons who posted that hyperlink for profit.⁶² As a result, the “ECJ continues to stand firm—though fragile—magisterially handing down sweeping decisions composed in highly deductive argumentative shorthand.”⁶³

Similarly, Dreier found that wherever national courts consider a certain prohibition of a use infeasible or unwarranted, they often use safety valves to circumvent the EU copyright norms⁶⁴, such as analogies and even arguments such as freedom of expression⁶⁵, implied consent⁶⁶, good faith and abuse of rights⁶⁷. This, however, interferes with the legal security and certainty that is associated with the EU Copyright legal system.

⁶⁰ This judgment “partakes in significant measure of the characteristic French judicial formalism”. Lasser 2004 p. 246

⁶¹ Lasser 2004 p.158, 245-248

⁶² ECLI:EU:C:2016:221 Opinion of Advocate General Wathelet delivered on 7 April 2016, para. 78; and C-160/15, GS Media para. 51

⁶³ Lasser 2004, p. 22

⁶⁴ Dreier 2015, p. 141

⁶⁵ Case 05-14928, HFA v FIFA, French Court of Cassation (2007)

⁶⁶ Case I ZR 69/08, Vorschaubilder I (Thumbnails I), German Federal Court of Justice (2010)

⁶⁷ Case 172/2012, Megakini v. Google Spain, Spanish Supreme Court (2012)

A good representation of the antagonistic decisions of the European courts regarding Copyright can be found in the Svensson⁶⁸ and GS Media⁶⁹ decisions. Two cases with fairly similar facts, both related with the right of communication to the public⁷⁰ and not too spaced in time, yet with very different conclusions. At first sight, it seemed that the CJEU, in Svensson, had concluded that it was possible for anyone to hyperlink to another author's work uploaded on the internet, in general⁷¹. However, the act of hyperlinking to works could, in certain cases, fall within the scope of the right of communication to the public, as long as the copyright holders did not take in account a certain type of "public when they authorised the initial communication to the public" there would be a communication to the public⁷², and since copyright holders will (most likely) not take into consideration people accessing unauthorized works, the users accessing unauthorized works would correspond to "new public" and hyperlinking unauthorized information correspond to a "communication to the public". Thus, the CJEU concluded that, generally, hyperlinking would result in communication to the public. Differently, in GS Media, the ECJ decided that hyperlinking should be considered a "communication to the public" only if it was proven that the hyperlinker knew, or could have reasonably known, of the illegally uploaded works. This knowledge will be presumed if the hyperlinker posted the link for financial gain.⁷³ Meanwhile, the CJEU had previously mentioned that the lawful use of a work could only exist where there was previous authorisation, or it was not forbidden by law. Clearly, the ECJ's decision in the latter case contradicted (and practically overruled) the CJEU's first decision, thus reducing the array of negative consequences that could have derived from Svensson such as a big disruption of the balance between the protection of the interests of copyright holders v. the rights of users

⁶⁸ C-466/12 Svensson

⁶⁹ C-160/15 GS Media

⁷⁰ Art. 3(1) of the InfoSoc Directive

⁷¹ C-466/12 Svensson para. 32

⁷² C-466/12 Svensson para. 24

⁷³ C-160/15 GS Media para. 51 and 56

of protected objects, and the general interest⁷⁴, but revealing the legal uncertainty regarding the Courts' application of the same provisions.

Additionally, according to Hugenholtz and Senftleben, because of the transposition of the “three-step-test”⁷⁵ to the laws of Member States, European courts can reinterpret the statutory copyright exceptions in accordance with that open-ended norm and, thus, withdraw a permission that was previously granted by a national law to make certain use of a work.⁷⁶ This test was supposed to be an open-ended norm that aided courts and lawmakers to adapt their copyright system to new cases involving recent technologies; and created “an important link between continental European and Anglo-American copyright systems”⁷⁷ (the first system known for its prominent legal certainty and the second for its flexibility). However, when it is interpreted with the closed list of exceptions of article 5(1)-(4) of the Infosoc Directive, or articles 3 and 4 DCDSM, the three-step test frustrates the legal certainty objective, because “if national legislation adopts and further specifies exceptions listed in [those Directives], these specific national exceptions may still be challenged on the grounds that they are incompatible with the EC three-step test”⁷⁸.

⁷⁴ C-160/15 GS Media para. 31 and 46

⁷⁵ The “three-step test” can be found in article 9(2) of the 1886 Berne Convention and article 13 of TRIPS. Both treaties lay down minimum standards that their signatories' copyright law must respect (including the US and every EU MS). Article 5(5) of the Infosoc Directive incorporates the three-step test and clarifies that to pass this test, copyright exceptions/limitations can only occur on special cases where there is no conflict with the normal exploitation of the work and no unreasonable prejudices to the legitimate interests of the right holder.

⁷⁶ Hugenholtz PB, Senftleben M 2011, p. 9

⁷⁷ Geiger 2013, p. 613

⁷⁸ Senftleben 2010, p. 69

In *Megakini*⁷⁹, the Spanish Supreme Court clarified beforehand that Google’s use of protected works (making available cached copies of the works) did not fall within the scope of the exception for acts of temporary reproduction. Afterwards, however, the court concluded that Google’s use was non infringing by interpreting the three-step test in light of the abuse of rights doctrine and the principle of good faith, as well as applying by analogy the “*ius usus inoqui*” doctrine that is usually employed in matters of real estate and movable assets, to avoid “absurd extralimitations”.

Also, in *Vorschaubilder I*, Google was sued for showing on their results list certain thumbnails of works uploaded on a freely accessible website and the German Federal Court considered that Google’s use could fall within the scope of the right of “making available to the public” without falling in the scope of any of that right’s exceptions. However, because the plaintiff gave her “implicit consent” to make that work available on her website, Google’s unauthorised use was ultimately considered non-infringing.

4.2. CRITICS TO THAT LEGISLATION

It is important to analyse the formulation of the legislation that came before the current DCDSM, and its interpretations in EU case law, to understand the consequences and the critics it had. Only after studying it can we compare the before after of the formulation, see if the problems that were pointed out about it were or could be solved through the new EU Directive and the MS transposition of those norms.

The several cumulative requirements from article 5(1) of the Infosoc Directive were crucial to reduce the number of acts that were, in fact, exempted from the protection of copyright laws.

It was concluded in the EP’s review of the copyright framework that US researchers had a competitive advantage in the field of TDM.⁸⁰ Mainly due to the EU legal framework at the time, researchers in the EU used TDM tools less than those in the USA. This meant

⁷⁹ Case 172/2012 *Megakini*

⁸⁰ Reynolds 2015, pp. 20, 186-187, 263

that, if everything stayed the same, the EU could be depleted of its talent pool, investment and its place as a competitive research location, to other jurisdictions.⁸¹ It is also mentioned in that document that the deficient legal framework existent at the time caused several problems, in particular: high transaction costs because of the need to negotiate licenses in each Member State together with the intermediaries that usually exist in that MS⁸²; content that is not available in certain locations due to copyright territoriality; legal uncertainty regarding TDM, leading confused actors of that sector to have different interpretations of the EU's E&L and not knowing their rights and obligations, resulting in inefficiency costs and increasing their liability risk; actors who are able to achieve substantial market power in detriment of others; and a devaluation of the uses the internet can provide its users,⁸³ among others.

We can also find several critiques to the previous EU legal framework in the Recitals of the Directive 2019/790 of the European Parliament and of the Council.⁸⁴ Change was eminent because the legal framework at the time was not adapted to the current and future state of technology⁸⁵, and the decisions of the European Courts were an obvious indicator of this conditions. Because of this, the EU institutions wished to reach an agreement that

⁸¹ The appraised real value of research output obtained by the EU research budget could escalate up to €5.3 billion if full access to TDM tools was granted because these tools could boost productivity of research projects up to 2%, by augmenting analysis outputs while maintain labor inputs. Reynolds 2015 p. 263

⁸² McDonald 2012 pp. 27, 43

⁸³ Reynolds 2015, p. 25

⁸⁴ These can help us understand the reasons behind the votes of the EU Parliament MEP's and approval by the qualified majority and finally the acceptance of the Directive by the relevant EU institutions. They can also help us comprehend the motivations behind the literal formulation of the TDM related articles in the DCDSM, in particular the wording of Articles 2, 3 and 4 of the DCDSM.

⁸⁵ Recital 3 DCDSM

amended, updated and ultimately harmonised the EU Copyright legal system at the time⁸⁶ and had some benefits to all interested parties.⁸⁷

Thus, the Proposal for a Copyright Directive⁸⁸ was approved, after many rounds of negotiations, into the DCDSM and published in the OJEU on May 15th 2019, for all MSs to transpose into their national law by June 2021.⁸⁹

4.3. THE NEW EXCEPTIONS IN THE DCDSM

4.3.1. THE OBJECTIVES

The EU institutions that negotiated the DCDSM had several objectives behind the Copyright reform. On the one side, it is mentioned in the EP's Review that "one of the main objectives of introducing regulatory changes in the area of copyright would be to ensure a higher degree of horizontal harmonisation and coherence of various pieces of legislation"⁹⁰. The intended results included: a reduction of the need for a substantial body of case-law to be settled before the CJEU⁹¹; an increase of certainty for the market players, and an improvement in the functioning of the internal market following the new and more effective and efficient licensing processes; an upgrade in the fair remuneration for authors and the dissemination of the works in the EU; augmented growth, sustainable job creation and the competitiveness of the EU economy; and not stifling innovation coming from different sectors, whether research organizations or businesses.⁹² On the other side, three

⁸⁶ Recital 5 DCDSM

⁸⁷ Recital 6 of the DCDSM

⁸⁸ COM/2016/0593 final - 2016/0280

⁸⁹ For an updated compilation of the national consultations and transpositions of the DCDSM by country see <https://bit.ly/3idcvFH> last visited on 09-06-2020

⁹⁰ Reynolds 2015, pp. 263-264

⁹¹ Reynolds 2015, p. 23

⁹² Rosati 2018, p.9

general objectives have been identified in the EC's document⁹³, along with TDM related objectives mentioned in the same document, such as: solving the problem of legal uncertainty for research organisations who wish to carry out TDM on content they have lawful access to; proportionate transaction costs for preservation of works by cultural heritage institutions; increasing legal certainty and reducing rights clearance costs for research organisations; the maintenance of an unaffected right holders' subscription market⁹⁴.

4.3.2. THE TDM EXCEPTIONS

Before delving into the exceptions present in the DCDSM, it must be noted that no authorization is needed to apply text and data mining to data that is not protected by copyright law. Eventually, the application of certain TDM techniques might not involve acts of reproduction and/or those acts might fall under the mandatory exception for temporary acts of reproduction provided for in Article 5(1) Infosoc Directive, applicable to TDM use that does not involve making copies outside the scope of that exception.⁹⁵

ARTICLE 3 OF THE DCDSM:

The provision existent in Article 3(1) DCDSM encompasses a mandatory exception to the following rights:

- 1) Article 5(a) Database Directive
- 2) Article 7 (1) Database Directive

⁹³ Namely, allowing wider online access to protected content across the EU; facilitating digital uses of protected content for education, research and preservation in the single market; and ensuring the online copyright marketplace works efficiently for all players and gives the right incentives for investment in and dissemination of creative content. SWD 2016, 302 final, p. 1

⁹⁴ SWD 2016, 302 final, pp. 1-4

⁹⁵ Recital 9 of the DCDSM

- 3) Article 2 Infosoc Directive
- 4) Article 15(1) of DCDSM the exclusive rights of publishers of press publications established in a Member State to:
 - authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part for the online use of their press publications by information society service providers⁹⁶
 - authorise or prohibit the making available to the public, by wire or wireless means, in such a way that members of the public may access them from a place and at a time individually chosen by them for the online use of their press publications by information society service providers⁹⁷

According to Article 3(1) of the DCDSM every EU member state must implement this exception in its legal jurisdiction.

Although, as stated by the provisions from recital (12), Articles 2(1) to 2(3) and 3(1) of the DCDSM, this exception should allow the reproduction and extraction of content from works and other subject matters, only if:

1 - those works and subject matters are protected by the provisions from articles 5(a) and 7(1) Database Directive, Article 2 InfoSoc Directive or Article 15(1) DCDSM;

2 - the reproduction or extraction is made by research organizations, i.e. “universities or other higher education institutions and their libraries, also entities such as (...) hospitals that carry out research”⁹⁸, “a research institute or any other entity, the primary goal of which is to conduct scientific research or to carry out educational activities involving also the conduct of scientific research: (a) on a not-for-profit basis⁹⁹ or by reinvesting all the

⁹⁶ Article 2 Infosoc Directive

⁹⁷ Article 3(2) InfoSoc Directive

⁹⁸ Recital 12 DCDSM

⁹⁹For examples of industrial and commercial data use see <https://bit.ly/2ZoNSgQ>

profits in its scientific research; or (b) pursuant to a public interest mission recognised by a Member State; in such a way that the access to the results generated by such scientific research cannot be enjoyed on a preferential basis by an undertaking that exercises a decisive influence upon such organisation”¹⁰⁰ for example “because of structural situations, such as through their quality of shareholder or member, which could result in preferential access to the results of the research”¹⁰¹, or by cultural heritage institutions such as “a publicly accessible library or museum, an archive or a film or audio heritage institution”¹⁰²,

2.1 - In this case “a public-interest mission could, for example, be reflected through public funding or through provisions in national laws or public contracts.”¹⁰³

2.2 - Research projects that are carried with a possible commercial outcome, eg. in the context of PPPs could be included in this context¹⁰⁴ as long as the research organizations do not provide preferential access to the results of their research to commercial entities¹⁰⁵. It seems as TDM research could be done by private business(es) who collaborate with the research organization(s), seeing that the latter does not control the former.

3 - in order to carry out TDM as defined in Article 2(2) DCDSM;

¹⁰⁰ Article 2(1) DCDSM

¹⁰¹ Recital 12 DCDSM

¹⁰² Article 2(3) DCDSM

¹⁰³ Recital 12 DCDSM

¹⁰⁴ Recital 11 DCDSM; the OECD defines PPP as a “long term contractual arrangements between the government and a private partner whereby the latter delivers and funds public services using a capital asset, sharing the associated risks.” See European Court of Auditors, 2018, p.12

¹⁰⁵ Recital 12 DCDSM

4 - for the sole purpose of scientific research, that covers “both the natural sciences and the human sciences”¹⁰⁶ or “educational activities involving also the conduct of scientific research”.¹⁰⁷

As specified by Article 3(2) DCDSM, the copies of works and subject matters that fall under the mandatory exception must respect all the requirements already listed in article 3(1) plus two other requirements: to store them with an appropriate level of security and to retain them solely for the purpose of scientific research.

In consonance with Article 3(4) DCDSM, the provisions from Article 3(2) and 3(3) DCDSM consist of an obligation and measures, respectively. The MSs must not create a mandatory exception, similar to that of Article 3(1), but merely “encourage rightsholders, research organizations and cultural heritage institutions to define commonly agreed best practices”¹⁰⁸ for the appropriate level of security that must be respected for storing copies of works or other subject matter (retained only for purposes of scientific research)¹⁰⁹; and the necessary measures to ensure the security and integrity of the networks and databases hosting works and subject matters¹¹⁰.

In line with Article 3(3) DCDSM, the rightsholders of the works and subject matters mentioned in Article 3(1) cannot override that mandatory exception with contract restrictions or by “expressly” reserving them¹¹¹. Nevertheless, they will have available different types of technology to provide for “the security and integrity of the networks

¹⁰⁶ Recital 12 DCDSM

¹⁰⁷ Article 2(1) DCDSM

¹⁰⁸ Article 3(4) DCDSM

¹⁰⁹ Article 3(2) DCDSM

¹¹⁰ Article 3(3) DCDSM

¹¹¹ Unlike the exception from Article 4(1) DCDSM, where the use of works may be “expressly reserved by their rightholders in an appropriate manner”.

and databases” where their content is hosted, as long as they do not “go beyond what is necessary to achieve that objective”¹¹².

CRITICS TO ARTICLE 3

Whereas the previous EU legislation did not explicitly mention TDM use, but merely some of the acts that are part of the TDM process (e.g. making temporary copies of works), this copyright exception explicitly allows the beneficiaries of article 3, with lawful access to copyright protected works, to use TDM on them, in various situations that were disregarded in the previous legal framework. This of course has its benefits and drawbacks.

Now that there is an exception for TDM, it is clear that TDM use on copyright protected works outside of the scope of the exception(s) is illegal, because it would infringe the protected Copyright in question. Thus, arguably, Article 3 effectively reverses the presumption of the subsistence of copyright in a work, and this of course goes against the debated “right to read is the right to mine” doctrine¹¹³. This also restricts further the discretion of European Courts to interpret broadly new uses of copyrighted works.

Article 3 has been criticised because it is considered far too restrictive and narrow in scope.

First, there is a legal access limitation. The provision from Article 3(1) DCDSM requires “lawful access” to the content that the beneficiaries of that exception intend to mine. This “lawful access” could be obtained, among other ways, through subscriptions to publications or open access licences¹¹⁴ and content that is freely available online.¹¹⁵ However, this exception can “effectively be denied to certain users by a right holder who refuses to grant ‘lawful access’ to works or who grants such access on a conditional basis

¹¹² Article 3(3) DCDSM

¹¹³ Murray-Rust 2014, pp.27-29 and IFLA (2013) p. 1-2

¹¹⁴ Recital 10 DCDSM

¹¹⁵ Recital 14 DCDSM

only”.¹¹⁶ It is the prerogative of a private entities to decide if they allow TDM use in most cases and how much they will charge for that. As a result, this exception can cause small and medium-sized enterprises¹¹⁷ and start-ups to lose competitive power in various sectors due to the prohibitive transaction costs. For instance, “it is impossible to prove that the training sets used for AI technology are free from copyright material”¹¹⁸ and now companies without a dominant market position are required to purchase licenses, to have lawful access to big amounts of data, to develop AI technology, when it is practically impossible to make contracts with so many EU rightholders. Though Article 4 DCDSM was seemingly created to avoid this issue, the ease with which its exception can be “disabled” by rightholders makes this provision practically obsolete. Thus, the problem still remains.

Secondly, there is a limitation on the beneficiaries of this exception: mainly research organisations and cultural heritage institutions can benefit from it. This is a problem because “not only academic researchers but also not-for-profit organisations, governmental institutions or profit-seeking commercial companies can carry out scientific research”.¹¹⁹ The EP’s JURI analysis shows that “from a practical market-based perspective”¹²⁰ the narrow formulation of Article 3 could create several difficulties for start-ups and unaffiliated individual researchers that will work in the same market circumstances as the actors privileged by article 3, such as those who fall under the definition of “research organization”. In this case, not even the “public-private partnership” alternative would be viable for start-ups because the tasks that need to be performed in those partnerships are time intensive and almost impossible for small teams.

¹¹⁶ European Copyright Society 2017, p. 4

¹¹⁷ SMEs represent 99 % of all business in the EU. SWD(2016) 301 final pt 1/3, p.86

¹¹⁸ Smart Data Research Department 2017, p.33

¹¹⁹ EDPS 2020, p. 11

¹²⁰ EP Legal Aspects Juri committee, p. 21

Thirdly, its specific purpose limitation. This is a problem because distinguishing between commercial and non-commercial purpose may be very difficult in practice, especially in the case of PPPs. This exception is not applicable to those who wish to make a profit through TDM use nor for TDM researchers independent from research organizations or cultural heritage institutions. These unaffiliated researchers, e.g. independent data scientists or members of a think-tank, will not be able to benefit from the exception present in article 3.

The pairing of the “lawful access” and “specific purpose” requirements can generate legal uncertainty for TDM users, even for the actors protected under Article 3. For example, it is not clear if TDM users can be obliged to purchase a specific “scientific research” license once they have already purchased a different type of license for the same content, or if that could be considered a contractual override that violates Article 3(1). These conditions to use TDM make all the difference when an actor is evaluating if TDM use in that particular jurisdiction is viable or not, due to the risks of paying high transaction costs and potentially being liable for its use.

Fourthly, the increased costs of TDM research projects resulting from the formulation of this exception. Publishers can maintain the price of their subscription and provide an option for TDM authorization for an extra fee, or raise the total price of their subscription by including TDM access in the subscription program.¹²¹ Where the DCDSM requires actors to have lawful access to these works and subject matters, it recognizes that TDM is a transactional service that must be obtained for a price. Thus, TDM use depends on the economic power of its user to access the market. It seems implicit in the ratio of this exception that those who profit from the use of TDM techniques, including SMEs, are required to purchase a license to use TDM in copyright protected works from their respective rightsholders because they have the capacity to do so. However, on the one side, actors protected under the provisions of Articles 3 and 4, including research organizations and cultural heritage institutions, without the means to purchase legal access will not be able to perform TDM research. Only those with sufficient funds to

¹²¹ Recital 10 DCDSM

purchase the licenses of works, subject matters or databases rights will have lawful access and be able to use TDM on them. This can highlight not only the economic and scientific discrepancies between entities inside a Member-State, but also among the richer and poorer Member-States of the EU. On the other side, entities with enough economic power to lawfully access the content can be discouraged from performing TDM because they are not included in the Article 3 exception, and even if they fall under the scope of protection of article 4, rightholders can still use contracts and TPM to forbid the TDM use on their content. Not to mention the potential for anticompetitive behaviours that could be triggered by rightholders. Rightholders such as companies that control big datasets compiled from various sources can, for example, choose to directly deny their main competitors to access the data, or to celebrate a series of exclusive contracts solely with companies that help eliminate their competitors, or even force the purchase of other goods or services that come together with access to the dataset when signing a contract with them or paying for their subscription fee.¹²² Both Article 3 and Article 4 beneficiaries need to have legal access to the works (or database that hosts them) for the exceptions to apply to them and, in order to obtain this legal access they can either choose to work with solely open content material or to purchase one or several licenses for the desired content. However, obtaining a license can be severely difficult and costly for TDM users, as it was previously mentioned.

These indicators suggest that the quantity and quality of TDM use, and consequently research, in the EU will continue to be secondary to countries with fair-use models and open clauses.

This article is considered by some as a starting point that can put the EU in disadvantage in areas such as AI research, that rely heavily on TDM to develop. For example, for-profit enterprises benefit from the Fair Use doctrine existent at the moment in the US are often protected under that legal regime when making copies of copyrighted works and subject matters for TDM purposes, as they are not limited to a nonprofit “scientific research”

¹²² Carmona 2019, p. 8

purpose, nor to the status of the miner.¹²³ Businesses outside the EU can be discouraged from investing in TDM research in the EU because of the restrictiveness of the legal framework compared with legal frameworks existent in other countries. Even EU businesses might try relocating their TDM research to countries with jurisdictions that are more lenient towards TDM use. This can translate into an increase in the quantity and quality of TDM jobs outside the EU, that might become more appealing to the talented workers and researchers. Thus, resulting in the relocation of TDM related businesses and talented workers from EU, and all the economic and social losses that can result from that.

Another possible outcome is that this exception will equip European Courts with more instruments to protect those who wish to apply TDM uses to copyrighted works, maybe override past decisions on these topics, and even help “circumvent” strict copyright protection provisions, such as those applied in *GS Media* or *Megakini*, depending on the instruments that the Courts choose to apply. Even if that does not seem to be the case here, due to the new increased legal requirements to consider someone or some use protected by an exception to exclusive copyrights/database rights, the strict EU framework has not stopped EU Courts from making decisions that respond to the new technologic challenges until now, as *supra* indicated, and now these may even be the biggest hope for the majority of EU TDM users.

However, there are motives to worry about the many caveats of this exception, arguably too many for an effective attainment of the benefits of TDM. On the one side, a narrow scoped TDM exception can benefit legal certainty in the EU legal framework, compared with the previous state of no exceptions because of the enumeration of certain categories of actors and the explicit conditions to be respected for its application, and the legal expectations that actors create about the actions that are limited or not by the conditions prescribed in the provision of Article 3. However, the expectations for those outside the scope of article 3 are dire, especially because of how easy it is for rightholders to circumvent the provision article 4 through contracts and TPM’s, transforming it into an empty exception.

¹²³ See, for instance, *Authors Guild v. Google, Inc.* (2015)

Considering its narrow scope, the provision of Article 3 may not equip TDM users with effective safeguards against overrides created by the rightsholders through TPMs. In fact, the biggest safeguard that actors may find in the DCDSM against TPM seems to be in Recital 16 of the DCDSM.

TPM'S IN THE EU

TPM's are known tools for disabling the use of TDM techniques on a certain type of content. For example, nowadays there are TPMs that forbid the use of TDM on pre-contractual information of online websites and by doing that they eliminate the only viable way the users currently have to know what they are agreeing with in the midst of blocks of information.

Through the interpretation of Article 3(1) and the absence of a paragraph similar to 4(3) in Article 3 of the DCDSM, it appears that rights holders cannot override that mandatory exception with contract restrictions or by “expressly” reserving their use, but they may use various types of technological protection measures to provide for “the security and integrity of the networks and databases”¹²⁴ where their content is hosted, given that they remain proportionate to the risks involved, and do not exceed what is necessary to pursue the objective of ensuring the security and integrity of the system, nor undermine the effective application of the exception¹²⁵. However, the definition of “proportionate to the risks”, “risks”, “what is necessary” and “security and integrity of the system” are not explicitly defined yet. MSs are obliged to encourage the parties with different interests (rightsholders v. TDM users) to agree on what will be the “best practices” to follow when applying these TPMs.¹²⁶

¹²⁴ Article 3(3) DCDSM

¹²⁵ Recital 16 DCDSM

¹²⁶ Article 3(4) DCDSM

Dussolier¹²⁷ explained that in the Infosoc Directive there was a compromise to solve the controversial issue of the legal protection of TPMs. Such compromise was “built on intricate and cryptic provisions¹²⁸” and that the national lawmakers were left to decide the conflict between TPMs and copyright limitations. This author further explained that the Directive did not indicate what kind of appropriate measures the MSs should take to safeguard the exceptions, nor how their appropriateness would be evaluated. He predicted the failure of the objective of harmonization in the EU because rightholders would frequently use different TPMs depending on the MS.

The new DCDSM indicates that the protection of TPMs established in the Infosoc Directive should be maintained¹²⁹ but new requirements have been added. MSs must not decide, *prima facie*, about the appropriateness of the measures but encourage the different actors to agree on the “best practices” for their application, and only if no agreements are made nor voluntary measures decided should the MSs then take appropriate measures in accordance with the first subparagraph of Article 6(4) InfoSoc Directive.”¹³⁰ Here, the lack of harmonization and uncertainty remains because its the actors of each MS (or alternatively the MSs themselves) who define the “best practices” for TPMs.

Additionally, the DCDSM now indicates certain guidelines for TPMs application, such as them remaining “proportionate to the risks involved” and not exceeding “what is necessary to ensure security and integrity of the system”¹³¹, yet the exact meaning of these concepts is still not explicit and it is for the national legislators to decide and subsequently for the European courts to interpret the application of the transposed provisions. Hence, history seems to repeat itself, these alterations do not appear enough to solve the lack of legal harmonization and uncertainty regarding TPMs.

¹²⁷ Dussolier 2003, pp. 462-463, 473

¹²⁸ Dussolier was referring to the “anti-circumvention provisions” present in Article 6 InfoSoc Directive

¹²⁹ Recital 7 DCDSM

¹³⁰ Recital 7 DCDSM

¹³¹ Recital 16 DCDSM

The use of TPMs to ban TDM use, including the previously mentioned SDSs¹³² that are used to automatically analyse contracts, stops important data from reaching consumers who wish to make informed choices. In fact, a survey has shown that, very often, researchers cannot access several types of content, such as journals, newspapers, websites, ebooks and databases; that the mentioned block on content can take on average a month to solve, if it can be solved at all; and that final sanctions applied by publishers can frustrate whole communities as much as individual researchers when they involve, for example, threats to cut off access to content unless TDM is stopped, the suspension of campus-wide access to paid for electronic subscriptions, limiting downloads to one document only, requests for additional payments and the use of CAPTCHA technology to frustrate TDM.¹³³

As a pivotal tool for the evolution of AI and contemporary research, among other very important uses, TDM seems to have been limited by the several conditions present in Article 3.

ARTICLE 4 OF THE DCDSM:

Firstly, it is stated in Article 4(4) DCDSM that no paragraph of Article 4 will affect the application of the previously mentioned Article 3 of that Directive.

Article 4 of the DCDSM contains a broader mandatory TDM exception on large datasets, by reducing the number of situations where TDM users have to obtain a license to mine the content. This exception is seemingly directed at sectors such as AI, machine learning and commercial data analytics. Every MS must transpose this exception to its national law because this exception is mandatory, but rightsholders may opt out if they do not want the exception to be applied on their content. That opt out option enables rightsholders to override TDM: through contracts, in cases where the content is licensed to

¹³² See section 3.1 of this thesis

¹³³ See the results of a survey carried out by LIBER, available at: <https://bit.ly/2ZqI7iy>

organisations e.g. distributed offline or through private networks; or by means of TPMs that block TDM in cases where the content is available online.

Because of the opt out option this provision is not as efficient protecting TDM use in copyrightable content as the exception from article 3 mentioned above. Additionally, it is stated in Article 7 (1) DCDSM that “any contractual provision contrary to the exceptions provided for in Articles 3, 5 and 6 shall be unenforceable” but it does not include article 4. In accordance with Article 4(3) DCDSM, this provision reinforces the fact that there is no protection for article 4 beneficiaries against contractual provisions of rightsholders who opt out of the exception in article 4 for their content available online.

Problems arise due to this opt out option. If TDM users overall are not granted free access to the works and subject matters many sectors will suffer, economically and socially. This will reduce the capacity that universities have to put the findings of their research projects on the market and make a profit out of it. Discoveries made in the public interest can suffer from this because, practically, research projects endorse innovation and innovation has inherent commercial value.

The provision displayed in Article 4(1) DCDSM obligates the EU MSs to implement an E&L to the previously mentioned rights provided for in Article 5(a) and Article 7(1) of Database Directive, Article 2 Infosoc Directive, and Article 15(1) DCDSM, as well as Article 4(1)(a) and (b) of the Computer Programs Directive.

In line with the provisions from Article 4(1) and 4(3) of the same Directive, this exception should allow the reproduction and extraction of content from works and other subject matters, only so long as:

- 1) those works and subject matters are protected by the provisions from articles 5(a) and 7(1) of the Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 15(1) of the DCDSM and Article 4(1)(a) and (b) of Directive 2009/24/EC;
- 2) those works and subject matter are lawfully accessible;
- 3) both reproductions and extractions are done for the purposes of text and data mining.

- 4) the rightholders of those works and subject matters did not expressly reserve them, “in an appropriate manner, such as machine-readable means in the case of content made publicly available online.”

As per the provision of Article 4(2) DCDSM, the reproductions and extractions mentioned in Article 4(1) can be retained, “for as long as is necessary for the purposes of text and data mining”.

CRITICS TO ARTICLE 4:

At first sight, because every EU MS must transpose it to its national law, the broad exception present in this article helps all TDM users who are not protected by the exception of article 3(1) DCDSM. This would seemingly reduce the number situations where those actors need to obtain a special license to mine content without the exclusive research purposes, after they accessed it lawfully.

Indeed, EU countries are obliged to create a TDM exception, but they are also obliged to provide an option for the rightholders to opt out from this exception through “appropriate” means (contracts and TPMs included). The option to override TDM techniques through contracts is allowed so long as their content is licensed through private networks or offline methods of digital supply.

Where the works and subject matters are publicly available online, the contractual overrides are not allowed but TPM’s are, i.e. the terms and conditions of a website can validly forbid the use of TDM on the website’s content.

Therefore, the actors that are not protected by the provisions from article 4 may still be required to purchase licenses to mine content. This of course includes the controversial cases of licenses for (profitable) development of AI and machine learning.

5. US LEGAL FRAMEWORK

It is often mentioned in the US, in doctrine and jurisprudence, that copyright law is connected with the First Amendment rights and should be understood within that context.¹³⁴

The very purpose of copyright in the US jurisdiction can be found in the US Constitution, where it is stated that “The Congress shall have power [to] promote the progress of science and useful arts, by securing for limited times to authors and inventors the exclusive right to their respective writings and discoveries.”¹³⁵ To fulfil that purpose, the US law and respective Courts protect the authors’ exclusive rights and create leeway, in the form of limitations to those rights, for the promotion of Progress of Science and useful Arts.

The copyright owner has several exclusive rights enumerated in 17 USC section 106.

US Copyright law can protect various forms of literary and artistic demonstrations. However, even if they fall under one of these forms, ideas themselves cannot be copyrightable, only the author’s expression of an idea. Judge Learned Hand pointed out that nobody has been (and never will be) able to fix the boundary between an idea and its expression¹³⁶ because “obviously, no principle can be stated as to when an imitator has gone beyond copying the “idea” and has borrowed its “expression”. [D]ecisions must

¹³⁴ Netanel 2001, pp.30-47 and Rubinfeld (2002) pp.5-27; also *Suntrust v. Houghton Mifflin Co.*, 268 F.3d 1257 (11th Cir. 2001), in this case the court concluded that there had been no copyright infringement, and ruled there was a Fair Use of the work, after expressing their preoccupation with the protection of First Amendment rights regarding freedom of speech, among others concerns with the use of copyrighted works.

¹³⁵ US Constitution, art. I, § 8, cl. 8

¹³⁶ *Nichols*, 45 F.2d at 121.

therefore inevitably be ad hoc.”¹³⁷ Nevertheless, US Courts have made attempts to define this boundary through the years.¹³⁸

For the work to be copyrightable it must also have a certain degree of originality and be “fixed in a tangible medium of expression”. Once the work is fixed, the copyright protection starts. The copyrightable works are registered in the Copyright Office even though it is not required to register them for it to be valid. However, if a US author intends to file an infringement suit she/he must register their works before doing so.

It must be noted that copyright does not provide the rightholders the exclusive right to inhibit people from using, making or selling their works, but only the right to prevent unauthorized copying of the protected works and some of the uses of those works, e.g. public performances and display.

According to the legal framework it is legal for someone to independently develop an identical or similar work¹³⁹. Therefore, to find out if there is copyright infringement, the law must distinguish between the legal and the illegal copying of a work, in accordance with 17 USC § 501(a). Traditionally, there is a legal interpretation of 17 USC § 106 that goes as follows: “a copyrighted work would be infringed by reproducing it in whole or in any substantial part, and by duplicating it exactly or by imitation or simulation. Wide departures or variations from the copyrighted works would still be an infringement as long as the author's “expression” rather than merely the author's “ideas” are taken.”¹⁴⁰

¹³⁷ Peter Pan Fabrics F.2d 487, 489

¹³⁸ In the Whelan case 797 F.2d at 1235. “the purpose or function of a utilitarian work would be the work’s idea, and everything that is not necessary to that purpose or function would be part of the expression of the idea”.

¹³⁹ In the Arnstein v. Porter case it can be seen that copyright protection is not an obstacle to independent creation.

¹⁴⁰ US House of Representatives Report No. 94-1476, 94th Congress, 2nd Session, p. 61 (1976)

The defendant can be liable for copyright infringement, whether it is a literal reproduction or a substantially similar reproduction of the protected expressions. Direct proof of illegal copying is rare so, usually, US courts infer that copying exists after there is proof the defendant accessed the plaintiff's work and evidence that the defendant's work is "substantially similar"¹⁴¹ to the protected expressions in the plaintiff's work.

To put it simply, it is required: proof of copying and substantial copying of protected expressions. However, it is difficult to determine if the defendant is infringing copyrights of the plaintiff because often copyrightable works mix original expression with ideas or public domain elements, or facts, or other elements that are not copyrightable. In these cases, a defendant can recognize that she created a work while knowing the plaintiff's work but that her work is not "substantially similar"¹⁴² to the protected expression in the plaintiff's work. US courts used this method to assess copyright infringement in various cases¹⁴³.

5.1. US COPYRIGHT LIMITATIONS

Copyright infringement is normally unrelated with the intent of the actor who copies the work. The US copyright laws do not distinguish between actors who copy with a commercial purpose or produce large-scale copying, they apply to all copying.¹⁴⁴ However, the rightsholder's exclusive right to copy is not absolute and it is not protected in all circumstances by Copyright laws. The 1976 Act approved copyright liability exemptions through the codification of the fair use doctrine developed until then by the courts, it recognized new compulsory licensing regimes (for musical compositions, cable

¹⁴¹ In the Peter Pan Fabrics case the Court decided there is "substantial similarity" where the "ordinary observer, unless he set out to detect the disparities [between two works], would be disposed to overlook them, and regard their aesthetic appeal as the same"

¹⁴² If the part from the work that was taken is "substantial", the "fair use" defence cannot be used in that regard

¹⁴³ See *Arnstein v. Porter*; *Gaste v. Kaiserman*; *Arnstein v. Edward*; *Marks v. Leo Feist*; *Wilkie v. Santly Bros.*; *Nichols v. Universal*.

¹⁴⁴ Menell 2018, p. 230

television and webcasts) and it forestalled several state and common law defences that affected the federal copyright protection, thus reducing the number of IP safeguards existent until then.

The Fair Use doctrine will be the only limitation further discussed in this document, due to its frequent mention in US Courts cases with claims of copyright infringement by the use of TDM tools.

5.2. FAIR USE IN THE US

“The ultimate test of fair use [is] whether the copyright law’s goal of “promoting the Progress of Science and useful Arts,” US Const., art. I, § 8, cl. 8, “would be better served by allowing the use than by preventing it.”¹⁴⁵

Jurisprudence along the years has developed the fair use doctrine not only as a safety net for freedom of expression, by creating more leeway for creative content such as commentary, criticism, and research, but also to balance the interests of first authors and the authors who draw on their content for cumulative creativity.

As it will be shown in this document, various decisions have adapted the interpretation of the fair use doctrine and the balance it brings to the protection of the authors and the creative arts, having proven to be a doctrine that will continue to evolve and accommodate the innovations in society.

Nowadays, US Courts apply a four factored test to ascertain if copyrighted content can be used without the authorization of the copyright owner. The authorization to use copyrighted material can be given by a copyright owner, its agent or the law¹⁴⁶. Fair Use was traditionally interpreted as an affirmative defense that excused an otherwise

¹⁴⁵ Castle Rock at 141, quoting *Arica*, 970 F.2d at 1077

¹⁴⁶ 17 USC §512(3)(A)(v)

infringing behavior, however, recently US Courts have viewed it as an authorization under the law, considering section 107 of the USC.

Fair use is considered a Limitations on Exclusive Rights in 17 USC § 107, where it is stated that “Notwithstanding the provisions of sections 106 and 106A, the fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright. In determining whether the use made of a work in any particular case is a fair use the factors to be considered shall incorporate (1) the purpose and character of the use, including whether such 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.”¹⁴⁷

Judges may still rule in favor of Fair Use even if they decide unfavorably on one or more of the factors because the four factors are considered altogether. This test has been applied in several US Court Cases, as it will be shown in this document.

The general intention behind this provision can be found in the Congress report¹⁴⁸ that accompanies the 1976 Act.

The application this test and the US courts interpretation of its four factors has varied over the years, as it can be found in the following three examples:

In *Harper & Row v. Nation Enterprises*, the Court did not find Fair Use in the Defendants’ unauthorized reproduction of parts of a work regarding unpublished memoirs whose

¹⁴⁷ 17 USC § 107(4) obliges courts to evaluate how much the market will be harmed because of the actions of the alleged infringer. This “must take account not only of harm to the original but also of harm to the market for derivative works.” *Harper & Row, supra*, at 568.

¹⁴⁸ H.R. Rep. No. 94 -1476, at 66

publishing rights had been sold to the Plaintiff, after applying the four factor test in the following manner.

The first factor (the Purpose and Character of the Use) weighed in favor of the Plaintiff because the court considered that despite the Defendant having “every right to seek to be the first to publish information” it did however go “beyond simply reporting uncopyrightable information and actively sought to exploit the headline value of its infringement, making a news event out of its unauthorized first publication of a noted figure’s copyrighted expression.”

With regards to the second factor, the Nature of the Copyrighted Work, the court considered that the unpublished nature of the work in question was an important factor weighing in favor of the Plaintiff and against the fair use defense; additionally the copied extracts did more than merely convey facts to the point where there was little separation between the idea and the expression where there were “subjective descriptions and portraits of public figures”.

The third factor, regarding the Amount and Substantiality of the Portion Used, weighed in favor of the Plaintiff because even though the excerpts were an “insubstantial portion” of the Plaintiff’s work they were “the heart of the book” and considered important to the determination of an infringing work, because of their expressive value.

Finally, the Court decided with regards to the fourth factor, the Effect on the Market, that the effect on the potential market is the “single most important element of fair use” and that, because the defendant’s infringement resulted in the cancellation of Time magazine’s serialization of the work with inherent costs of 12.500 dollars, that represented a direct competition for a share of the market and was “clear-cut evidence of actual damage.”

Differently, the *Matt Hosseinzadeh v. Ethan and Hila Klein* is an interesting case where broad copyright protection was forfeited in favour of cumulative creativity. The court applied the four-factor test and decided the Defendants’ made Fair Use of the Plaintiff’s original work.

With regards to the first step, the court considered that the Defendants' video was "quintessential criticism and comment" therefore having a different purpose from the original work.

However, the Defendants' video was "entirely scripted and fictional" so the second step weighed in favor of the Plaintiff.

The third factor did not weigh in favor of any of them because the Defendants' could and did use clips of the plaintiff's original work that were "plainly necessary" and "reasonable to accomplish the transformative purpose of critical commentary" in order "to comment on and critique a work" but, nevertheless, "a great deal of plaintiff's work was copied."

The fourth step favored the Defendants', since their video "transforms [the Plaintiff's] video from a skit into fodder for caustic, moment-by-moment commentary and mockery", therefore not being "a market substitute" for the Plaintiff's video.

Following this interpretation of 17 USC § 107, US Courts have also reiterated that it is necessary to apply the fair use doctrine in a flexible manner¹⁴⁹ and on a case-by-case basis¹⁵⁰. This way, the Fair Use doctrine can adapt to changes in society and in technology, authorizing new uses that were unthinkable at the time the Copyright Law was last amended. Thus, if the users act under the fair use clause when they use copyrighted content they have no need to obtain a previous authorization from the right holder for that use.

Along with the application of this four factor test, doctrine and jurisprudence have come to develop the "Transformative use doctrine" and, nowadays, the effective transformative use of a work is critical to determine the existence of Fair Use of a copyrighted work.

Usually taken into consideration amid the analysis of the first statutory factor¹⁵¹, the Court may conclude there is transformative use if "the secondary use adds value to the original—if the quoted matter is used as raw material, transformed in the creation of new

¹⁴⁹ Perfect 10 v Amazon at 1166

¹⁵⁰ In Castle Rock at 141

¹⁵¹ 17 USC § 107(1)

information, new aesthetics, new insights and understandings [because] this is the very type of activity that the fair use doctrine intends to protect for the enrichment of society.”¹⁵²

The scope of application of Fair use was enlarged through the employment of the transformative use doctrine in jurisprudence, thus revealing its importance for the public benefit. Fair use started to protect those who used someone else’s original work but with a different purpose, thus promoting the progress of Science and useful Arts.

Criticism has been made to the transformative use doctrine.¹⁵³ However, this transformative use element has become essential in the analysis of the existence of fair use. According to a study made by Neil Netanel, from 2006 to 2010 around 95.83% of US Courts considered “Transformativeness” and 100% of the defendants won when the Court found their use of the copyrighted content was transformative.¹⁵⁴

Because there is no standard application of general principles to all cases, the jurisprudence surrounding the analysis of Fair Use is rich and diverse. Consequentially, lawyers must pay attention to several fair use related cases to reach their conclusions. Not only that, but they will have to face the uncertainty regarding the judge or panel hearing the case and juris who may have subjective views in the matter, and act accordingly depending if the client is being sued in a Regional Circuit, or the Federal Circuit in case there is a patent matter involved in the case.

Fair use cases may fall within the categories listed in section §107¹⁵⁵, many of which are connected with First Amendment protections. However, that would be only the first

¹⁵² Leval, 1990, p.1111

¹⁵³ See *Kienitz v Sconnie Nation* and the corresponding response in *Authors Guild v Google*

¹⁵⁴ Netanel 2011, p. 755

¹⁵⁵ Namely, criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research 17 USC § 107

step to considered the existence of Fair Use because, “the fair use enquiry often requires close questions of judgment as to the extent of permissible borrowing in cases involving parodies (or other critical works)”¹⁵⁶ This means that even if the purpose of the use is listed in section 107, that use must involve a parody or other critical work.

US Courts are now faced with a modern era of Fair Use cases. These appeared after the normalization of the use of digital tools that allow authors to create projects that were unthinkable years ago due to the high implied costs. This resulted a series of relevant fair use jurisprudence covering the evolution of technology and its implications in copyright law.

From the Kelly case¹⁵⁷ onwards, courts have ruled in several decisions that reproduction of works, under TDM use and database development through TDM use¹⁵⁸, constitutes Fair Use. Diverse purposes are often mentioned in these cases, everything from scholarly research to investigating plagiarism. The courts started recognizing that TDM was not a substitute for original works, but a tool that allowed new purposes. The use of databases for TDM and TDM use started being considered a “highly transformative” use and the public benefit that can only be offered by TDM use was, and still is, a major factor for these rulings.¹⁵⁹ As mentioned by Circuit judge T.G. Nelson “Arriba's use of Kelly's images promotes the goals of the Copyright Act and the fair use exception. The thumbnails do not stifle artistic creativity because they are not used for illustrative or

¹⁵⁶ See Campbell

¹⁵⁷ Kelly v. Arriba-Soft – Arriba-Soft made exact replicas of Kelly’s original images, however, the Arriba’s thumbnails had a purpose different than that of Kelly’s images. That led to the ruling of the court where the use of Kelly’s images in a search engine was transformative and represented Fair Use.

¹⁵⁸ US courts usually differentiate between productive use and reproductive use. Whilst the productive use of a work increases the amount of new works, the reproductive use merely increases the amount of “copies” of a work.

¹⁵⁹ See A.V. v. iParadigms

artistic purposes and therefore do not supplant the need for the originals. In addition, they benefit the public by enhancing information-gathering techniques on the internet.”¹⁶⁰

Other examples of TDM related decisions in US Courts include cases such as *Perfect 10 v. Amazon* (2007)¹⁶¹, *A.V. v. iParadigms* (2009)¹⁶², *White v. West* (2014)¹⁶³, *Authors Guild v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014) – In collaboration with Google, HathiTrust (a group of universities) digitized works into their database, the HathiTrust Digital Library, a repository with indexing/searching tools enabled by data mining and textual analysis, that also provided full access to copyrighted works to sponsors with certified print disabilities. The court concluded that those two features qualified as Fair Use, especially because the database had effective security measures that blocked unauthorized access to the works and because the plaintiffs, individual authors and authors’ associations, were not able to establish a (non-speculative) risk that the database could create replacement copies of their copyrighted works. The Court asserted not only that full-text search posed no harm to any existing or potential traditional market for the copyrighted works but also that the creation of a full-text searchable database was fair use

¹⁶⁰ *Kelly v. Arriba Soft*

¹⁶¹ the court also ruled in favor of Fair Use, noting that using thumbnail copies of images protected by copyright in the internet search results constitutes transformative use, since thumbnail copies have a different purpose/function from that of the original images.

¹⁶² in this case *iParadigms* constructed a database to check for plagiarism. The database accessible by teachers who were able to compare their students’ papers with previous papers present in the database and information from the Internet. Even if the access to the database was paid, and therefore had commercial nature, the court determined it a “highly transformative” use.

¹⁶³Two publishers created and licensed two databases (*Westlaw* and *LexisNexis*). They copied legal filings from the *PACER* database and inserted them into their own databases, they transformed them into “text-searchable” documents and added Metadata to those copies to enhance their searchability. Thus, creating a legal search tool with search results that included the full text of the copied legal filings. The copyright owner, a lawyer who had the copyright registrations for briefs submitted to the *PACER* database sued the publishers for infringement, accusing them of unauthorized use of his briefs for their commercial databases. The court considered that the unauthorized copying and the display of full legal filings in the search results qualified as Fair Use because there was transformative use and three of the four statutory factors weighed in favor of Fair Use.

and that it was a “quintessentially transformative” use because “the result of a word search is different in purpose, character, expression, meaning, and message from the page (and the book) from which it is drawn.” The court concluded the ruling, observing that the copies were reasonably necessary to help the public use the database’s services and to reduce the risks of data loss.

In *Authors Guild v. Google*, the court pointed out the transformative use in the Google Books project and its importance to TDM because it “transformed the book text into data for the purpose of substantive research, including data mining and text mining in new areas, thereby opening up new fields of research. Words in books are being used in a way that they have not been used before”. A settlement agreement with the copyright owners was ultimately denied by the court and, finally, after a decade of litigation, Judge Leval delivered the final opinion on this case. Not only did the Court highlight the transformative use and irrelevance of the profit motivation behind the Google Books Project to the determination of the existence of Fair Use but also the benefit it would bring to the public knowledge without affect the copyright owners interests though a substantial substitute of their works.

As it can be seen in the cited case law, unauthorized and non-transformative use of a work resulting in content that competes with the commercialization of the original work will usually not be considered as a Fair Use defense by the US Courts, thus protecting the exclusive rights of the authors and copyright holders of and their original first expression. The public interest in the access to information has also been a significant factor weighing in favor of “fair use” often cited by the US Courts with regards to the application of the “fair use” defenses for TDM use.¹⁶⁴

A brief look into the recent development in US caselaw reveals that fair use has often privileged text and data mining in research. “Quietly, invisibly almost by accident,

¹⁶⁴ Okediji 2017, p. 32

copyright has concluded that reading by robots doesn't count. Infringement is for humans only; when computers do it, it's fair use.”¹⁶⁵

5.3. TPMS IN THE US

The Digital Millennium Copyright Act (DMCA) recognized, on the one side, anti-circumvention and anti-trafficking bans to protect copyright holders, while guaranteeing that the TPMs placed by them work effectively, and on the other side, various of safe harbours for Internet Service Providers (ISPs) and telecom companies.

The two main categories of TPMs in this Act are TPMs that manage the access to copyrighted content¹⁶⁶, laid down in 17 USC section 1201(a), and TPMs that manage the reproduction, copying, and similar copyright rights after lawful access has been obtained¹⁶⁷, laid down in 17 USC section 1201(b).¹⁶⁸ In generic terms, the first category forbids acts of TPM circumvention and the second category forbids the marketing and trafficking of circumvention devices.

Breaching the DMCA does not imply copyright infringement, the US courts consider it a separate offense. It is stated in Section 1201(c)(1), title 17 US Code, that “nothing in this section shall affect rights, (...) or defenses to copyright infringement, including fair use” but courts have previously ruled that, even if the DMCA does not hamper the application of Fair Use, it can diminish the effectiveness of that defence by permitting copyright holders to file a non-copyright claim.¹⁶⁹

¹⁶⁵ Grimmelmann 2016, p. 658

¹⁶⁶ e.g. the “robots.txt” file that indicates what pages of a website robots can visit or not.

¹⁶⁷ e.g. a lock that blocks users from copying software of a computer to another computer, or from copying the music of a CD to another device.

¹⁶⁸ This section does not forbid circumvention acts but merely the marketing and trafficking of circumvention devices, in order to enable users to make fair use of content after obtaining legal access to it.

¹⁶⁹ *Studios v. Metro-Goldwyn-Mayer* - a software provider was sued because it sold a DVD ripping software that permitted DVD owners to copy parts of DVDs to CDs, including DVDs encoded with a TPM denominated Content Scramble System. The court held that the software provider was liable for TPM infringement. The court also ruled that

To counter the negative impact that Section 1201 inflicts on users who wish to make fair use of copyrighted works, the DMCA allows TPM circumvention only for reverse engineering of computer programs that have the “sole purpose of identifying and analyzing those elements of the program that are necessary to achieve interoperability of an independently created computer program”¹⁷⁰, and allows some copyrighted works to be exempted from this rule by the Librarian of Congress whenever “persons who are users of a copyrighted work are, or are likely to be in the succeeding 3-year period, adversely affected by the prohibition under subparagraph (A) in their ability to make noninfringing uses under this title of a particular class of copyrighted works.”¹⁷¹

it was possible to claim Fair use under the DMCA even if that was difficult without circumventing the TPMs.

¹⁷⁰ 17 US Code Section 1201(f)(1)

¹⁷¹ 17 US Code Section 1201(a)(1)(C)

6. CONCLUSIONS

To answer the first research question, articles 3 and 4 of the DCDSM, appear to be insufficient to complete most of the objectives referred to in section 4.3.1. of this thesis.

To answer the second research question it must be mentioned that, despite of being subjected to a more restrictive body of Copyright law, EU Courts have done a fantastic work, adapting the interpretation of the provisions in accordance with the technological changes, even if they do not do it as transparently as US Courts. Despite this, the US Copyright framework continues more appellative for TDM users.

Both conclusions were reach due to the following reasons.

Where Copyright law is used to prohibit the dissemination of relevant information extracted from copyrighted content it works against its purpose. Whilst the US legal system can rely on the Fair Use doctrine to better identify what copyrights should be protected in accordance with their purpose, in a particular case¹⁷², the current EU legal system can rely on two TDM exceptions to the exclusive copyrights.

One exception in article 3, limited to research organisations and cultural heritage institutions for the purposes of scientific research; and another exception in article 4, that includes TDM use in a commercial context, limited to works and other subject matter that have not been expressly reserved by their rightholders in an appropriate manner. In cases where copyright holders have expressly reserved their works or other subject matter and TDM users are not research organisations and cultural heritage institution/ do not wish to use it for scientific research purposes, some authors have wondered if it would be possible

¹⁷² E.g. in *Authors Guild v. HathiTrust*, the court ruled that there was Fair Use in the reproduction of works into a searchable database (through TDM), because “the result of a word search is different in purpose (...) from the page (and the book) from which it is drawn”.

to invoke the implicit requirement that the act of reproduction must involve a “use as a work”, and that such use cannot exist when TDM is performed. Using this an argument to allow EU and national courts to consider the acts of reproduction (and extraction) during TDM a non-infringing use of copyrighted material in several other instances, besides the ones mentioned in the DCDSM.

Where the EU’s TDM exceptions allow different uses, that could trigger copyright infringement of exclusive rights otherwise, they appear akin to the US’s fair use doctrine. However, they differ where one is more flexible than the other. Whilst the fair use doctrine is based on open norms, hence flexible enough to adapt to new cases not envisaged by lawmakers and courts at the time of the creation of those norms, the TDM exceptions are shaped by an assortment of previously enumerated acts. There are, of course, pros and cons to both legal options. Though limited by the previous decisions and the formed precedent, the fair use doctrine may be flexible enough to allow courts to respond more efficiently in new unpredictable cases at first sight. The EU legal systems’ closed number of statutory exceptions may not cover certain uses that might be of effective public interest and it could also create uncertainty for creators of copyrightable works and TDM users where no previous decisions have elaborated on the ground rules to make fair use of a work.

Arguably, European Courts will interpret these new exceptions in harmony with what they believe are the advancements of new technologies and uses that *prima facie* might infringe copyright, as they have done in the past. However, the strict formulation of the TDM exceptions might be an added obstacle to this exercise due to its vast and precise list of conditions. Thus, creating another step forward towards increased legal certainty in the EU yet one step backwards regarding the flexibility to adapt the EU norms to technological developments in our current reality.

The current EU legal framework shows a clear improvement for a specific category of TDM users when compared with the previous EU legal framework, due to the explicit

recognition of two specific TDM exceptions¹⁷³ that can now serve as ground rules for MSs to implement in their legal systems. There is an expected¹⁷⁴ reduction in market inefficiencies caused by high transaction costs in the scientific research sector, after the MSs transpose the Directive to their national legislations. Thus, facilitating the gains of trade between rightholders and the creation of markets that otherwise could not have formed due to the inherent costs, also reducing social loss caused by market power, hold-ups or positive externalities.

The exception in Article 3 greatly benefits its recipients because they are faced with few requirements, besides the mandatory scientific research purpose, to legally reproduce and extract content from copyrighted works through TDM and the rightholders do not have the prerogative to impede this through contracts or TPMs. The actors benefited by the exception in article 4 have now opportunities that were not presented to them before, despite being limited to “works and other subject matter referred to in that have not been expressly reserved by their rightholders”, with fewer opportunities than the beneficiaries of Article 3. However, it seems that the sought-after objectives of legal certainty and harmonization of EU copyright law, with regards to TDM, have not been reached yet. It still does not seem to provide the necessary legal environment for a society where TDM and machine legibility can be executed so as to develop automated systems to treat information.

The players that are disregarded/disadvantages from the exception include for-profit businesses and independent researchers. EU for-profit firms and independent researchers will suffer from this disadvantage competing with their American equivalents, among other companies from countries with legal systems that allow more freedom to perform TDM analyses. The fact that most “profitable uses” (excluding the possibility of profitability in PPP’s) are banned from the exception in Article 3, and the exception from Article 4 appears to be easily overridable by rightholders through either contracts or TPMs, making Article 4 practically obsolete, does not appear as a good enough solution

¹⁷³ Articles 3 and 4 of the DCDSM

¹⁷⁴ Langu 2013, p. 73

for the EU legal framework and for the various actors that should be taken into account the EU, like they are in the US.

When it comes to the creation and use of TDM tools in the EU, the threat of the previous Directive remains. The formulation of the current TDM related Articles and Recitals of the DCDSM does not eliminate the temptation to relocate to other research jurisdictions that are more advantageous for researchers and text and data miners in general. They are still likely to lead to the depletion of talent and investment from EU. The incentives given to actors such as European businesses, especially SME's and start-ups, individual researchers and even cultural heritage institutions and research organisations with low resources, especially those who rely on access to big amounts of data, do not seem enough to maintain and to generate new investments regarding the use and creation TDM tools in the EU when they are competing in the FinTech ecosystem based in Silicon Valley who work under the Fair Use clause. Even if the EU institutions grants subsidies to those businesses and researchers for them to start their projects, many may have to choose between terminating their work or relocating due to the high transaction costs and liability risks to perform TDM in the EU.

Additionally, the adoption of this legislation can create a false sense of security regarding the developments in the area of TDM, and the benefits that it can bring, merely because there was an improvement in the legal framework. EU legislators might think that they have improved enough and think that they will not have to alter the legislation for a long time, despite it having grey areas, creating disadvantages for important players in the TDM field and, ultimately, becoming practically obsolete in a short amount of time after being approved.

Complacency regarding this legislation might leave the EU behind in the technological sector in general, including fast developing markets such as the AI and data science line of businesses. The fact that countries like the USA have jurisdictions that allow for various types of businesses and researchers to develop TDM activities not only creates a disadvantage for the EU's for-profit businesses and independent researchers, but for the general TDM community because of the lack of think tanks and a concentration of diversity of businesses and clients that wish to use those services and end choosing a more

competitive market, where they can find better professionals maybe even cheaper prices – because there are more TDM fuelled businesses, more offer and same type of demand, creating therefore cheaper prices. The feared loss of talent and investment to more favorable research locations is still eminent, especially these times where AI and machine learning represent a booming activity in the economy that requires the automated analysis of big datasets.

After the analysis of the both legal systems with regards to the non-infringing use of TDM on copyrighted works and subject matters, through both the Fair Use doctrine based on open ended norms and the system of closed statutory exceptions, ultimately leads to the conclusion that the benefits of having the currently “narrow” EU TDM exceptions, do not surpass the benefits of a broader TDM exception in the EU. The various requirements necessary for their application to TDM user in the EU make it difficult to reach the intended objectives that were explicated by the EU institutions.

The TDM exception(s) should not have been as narrow as they were formulated to be. In accordance with the analysis presented in this document, one “broader” TDM exception could have better suited the objectives promoted by the EU institutions and benefited the public interest of the EU. Publicly accessible data is fundamental for the advancement of science, arts and technology, but also because it promotes civil initiative in these fields and influences the surrounding economy.

TDM users that are not one of the enumerated beneficiaries of article 3 (and all actors that are uncertain if they are included in this closed list, i.e. if they conform to the definition of research organisation or cultural institution, or if they are an entity whose “primary goal” is to conduct scientific research or educational activities, or if their results will be enjoyed “on a preferential basis” by an undertaking that has “decisive influence” upon it) will be subjected to a considerably narrower exception¹⁷⁵, due to the condition in article 4 that restricts legal TDM use to works where rightsholders have not expressly opposed

¹⁷⁵ In comparison with the exception in Article 3 DCDSM

“in an appropriate manner” to that TDM use, even if the users have previously acquired the access permission for that work from the same rightsholders.

This will still translate in high transaction costs, as mentioned above, because of the need to find and negotiate the mining licenses to that content. Additionally, the data can only be legally stored for as long as it is justified by that particular mining purpose, thus preventing the storage of previously gathered data for further use for new purposes, despite the costs of gathering it in the first place.

During the DCDSM implementation period in the MS, publishers must lay down their terms and conditions regarding TDM use and actors in the TDM sector will have to evaluate how they will access the text and data they need, in accordance with this new EU copyright framework.

More importantly, the creation of both Article 3 and 4 provisions clearly reverse the presumption of copyright subsistence in a work by asserting that text or data mining on a copyrighted work, without falling into one of the exceptions, will equate to copyright infringement. Thus, these exceptions appear eerily displaced in the midst of the current copyright *acquis Communautaire* that follows that presumption.

7. REFERENCES

Aggarwal CC (2015) Data Mining: The Textbook. Springer International Publishing Switzerland. Available at: <https://bit.ly/38ifk3N>

Carmona AF, Lombardo AGQ, Pastor R, Quirós CT, García JPV, Muñoz DR, Martín LC. (2019) Competition issues in the Area of Financial Technology (FinTech), In-Depth analysis Requested by the ECON committee of the European Parliament, available at: <https://bit.ly/2YQpx17>

Dreier T (2015) Thoughts on Revising the Limitations on Copyright under Directive 2001/29, JIPLP, V. 11, I. 2 (2016). Available at: <https://bit.ly/2BeP4vj>

Dusollier S (2003), Tipping the Scale in Favour of the Right Holders: The European Anti-Circumvention Provisions. Available at: <https://bit.ly/2VxCU7I>

EDPS (2020) European Data Protection Supervisor, A Preliminary Opinion on data protection and scientific research. Available at: <https://bit.ly/3gdrQnS>

European Copyright Society (2017) General Opinion on the EU Copyright Reform Package. Available at: <https://bit.ly/2NjNvIk>

European Court of Auditors (2018) Special Report, Public Private Partnerships in the EU: Widespread shortcomings and limited benefits”, Available at: <https://bit.ly/3g6gJgu>

Filippov S, Hofheinz P, Interactive Policy Brief Issue 20/2016 (2016). Available for download at: <https://bit.ly/38a4Rr7>

Flynn S, Carroll M, Sag M, Guibault L, Margoni T, Butler B, Souza AR, Jančić MB, Jaszi P, Quintais JP, Geiger C, Ncube C, White B, Scaria AG, Botero C, Craig C (2020) Joint Comment To Wipo On Copyright And Artificial Intelligence. Comment endorsed by 16 members of the Global Expert Network on Copyright User Rights. Available for download at: <https://bit.ly/388mYxG>

Gantz J, Reinsel D (2012) The Digital Universe In 2020: Big Data , Bigger Digi Tal Shadows, and Biggest Growth in the Far East. Available at: <https://bit.ly/2VJIB2v>

Geiger C, Frosio G, Bulayenko O (2018) The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects, IN-DEPTH ANALYSIS For the JURI committee. Available at <https://bit.ly/2F6FQBd>

Geiger C, Gervais D, Senftleben M (2013) The Three-Step Test Revisited: How to use the Test's Flexibilities in National Copyright Law, PIJIP Research Paper no. 2013-04. Available at: <https://bit.ly/3eMppZd>

Grimmelmann J (2016) Copyright for Literate Robots. 101 Iowa Law Review
Hugenholtz PB, Bechtold S, Dusollier S, Kretschmer M, Ohly A, Poor J, Quintais JP, Rognstad OA, Strowel A (2018) Copyright Reconstructed: Rethinking Copyright's Economic Rights in a Time of Highly Dynamic Technological and Economic Change, Kluwer Law International. Available at: Google Books <https://bit.ly/3eIXhWQ>

Hugenholtz PB, Senftleben M (2011) Fair Use in Europe: In Search of Flexibilities. Available for download at: <https://bit.ly/2BLvwym>

Hugenholtz PB, Senftleben M (2011) Fair use in Europe: In search of flexibilities. Available at: <https://bit.ly/2YKnsqE>

IFLA (2013) Statement on Text and Data Mining. Available at <https://bit.ly/3i9BkCq>

ISA² (2019) ISA² Programme 2019 Detailed Description of Actions Part ½ Available at: <https://bit.ly/2NE1teD>

Judge Leval PN (1990) Toward a Fair Use Standard, 103 Harv. L. Rev. 1105

Kelly E (2016) New EU text and data mining proposal would not benefit everyone. Available at: <https://bit.ly/2NQZtQx>

Kelly E (2018) MEPs vote to grant researchers greater freedom to text and data mine. Available at: <https://bit.ly/2ZmF2QL>

Knight W (2020) Researchers Will Deploy AI to Better Understand Coronavirus. Available at: <https://bit.ly/2YMi86l>

Langu G, Neven D, Shier G (2013) Assessing the economic impacts of adapting certain limitations and exceptions to copyright and related rights in the EU, Prepared for the European Commission and DG MARKT.

Lasser MS (2004) Judicial Deliberations: A Comparative Analysis of Judicial Transparency and Legitimacy, Oxford University Press. Available at: Google Books <https://bit.ly/2ZlQRq7>

Lunden I (2019) Factmata gets backed by eyeo, maker of Adblock Plus, and takes over its Trusted News app. Available at: <https://tern.ch/31oggSV>

Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH (2011) Big data: The next frontier for innovation, competition, and productivity. Full report available at: <https://bit.ly/3ga2esc>

McDonald D (2012) Intelligent Digital Options and Ursula Kelly, Viewforth Consulting, JISC, The Value and Benefit of Text Mining to UK Further and Higher Education, Digital Infrastructure. Available at: <http://bit.ly/jisc-textm>

Menell PS, Lemley MA, Merges RP (2018) Intellectual Property in the New Technological Age, Volume II: Copyrights, Trademarks & State IP Protections, Clause 8 Publishing

Ménière Y, Rudyk I, Valdes J, (2017) Patents and the Fourth Industrial Revolution (EPO), no. December: 100. Available at: <https://bit.ly/2BNrTbi>

Miner G, Elder J, Fast A, Hill Y, Nisbet R, Delen D (2012) Practical Text Mining And Statistical Analysis For Nonstructured Text Data Applications. Academic Press

Murray-Rust P, Molloy J, Cabell D (2014) “Open Content Mining”. Available at: <https://bit.ly/2NHXSMG>

Netanel NW (2001) Locating Copyright Within the First Amendment Skein. Stanford Law Review, V. 54. Available at: <https://bit.ly/2YLyNqh>

Netanel NW (2011) Making Sense of Fair Use, Lewis & Clark Law Review (2011) Available at: <https://bit.ly/31sW6ap>

Okediji RL (2017) Copyright Law in an Age of Limitations and Exceptions. Available at: <https://bit.ly/2BpePsJ>

Prosser M (2020) How AI Helped Predict the Coronavirus Outbreak Before it Happened. Available at: <https://bit.ly/3dK54lQ>

Radev D, Otterbacher J, Winkel A, and Blair-Goldensohn S (2005) News In Essence: Summarizing online news topics. Available for download at: <https://bit.ly/35TrL3W>

Reinsel D, Gantz J, Rydning J, (2018) The Digitization of the World From Edge to Core, An IDC White Paper. Available at: <https://bit.ly/2YGMDu2>

Rendas T (2017) Copyright, Technology and the CJEU: An Empirical Study, 49(2) IIC – International Review of Intellectual Property and Competition Law 153 (2018). Available at: <https://bit.ly/3eXI2cZ>

Reynolds S (2015) Review of the EU copyright framework European Implementation Assessment, EPRS European Parliamentary Research Service, Ex-post Impact Assessment Unit. Available at: <https://bit.ly/2CXX7Nz>

Rosati E (2018), The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Technical Aspects, Briefing Requested by the JURI committee. Available at: <https://bit.ly/3dJNaQq>

Rubinfeld J (2002) The Freedom of Imagination: Copyright 's Constitutionality, Yale Law Journal V. 112:1. Available at: <https://bit.ly/31wwzgw>

Senftleben M (2010) The International Three-Step Test: A Model Provision for EC Fair Use Legislation. JIPITEC 1/67. Available at: <https://bit.ly/2ZnarSR>

Shu C (2018) Amazon's newest service uses machine learning to extract medical data from patient records. Available at: <https://tcrn.ch/2NGcU5q>

Singh N, Hu C, Roehl WS (2007) Text mining a decade of progress in hospitality human resource management research: Identifying emerging thematic development. International Journal of Hospitality Management, Volume 26, Issue 1. Available at: <https://bit.ly/2BTmVtl>

Smart Data Research Department (2017) Data as an economic asset, European data economy or rights to data? Available at <https://bit.ly/2VA6HfP>

Stieg C (2020) How this Canadian Start-Up Spotted Coronavirus Before Everyone Else Knew About it. Available at: <https://cnb.cx/3g5vvEg>

Sunstein C (2012) Informing Consumers through Smart Disclosure. Available at: <https://bit.ly/38fyqaH>

SWD (2016) 301 final part 1/3, European Commission Staff Working Document, Impact Assessment on the modernisation of EU copyright rules Accompanying the document Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market and Proposal for a Regulation of the European Parliament and of the Council laying down rules on the exercise of copyright and related rights applicable to certain online transmissions of broadcasting organisations and retransmissions of television and radio programmes. Available at: <https://bit.ly/2BQo30X>

SWD (2016) 301 final Part 3/3, European Commission Staff Working Document, Impact Assessment on the modernisation of EU copyright rules Accompanying the document Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market and Proposal for a Regulation of the European Parliament and

of the Council laying down rules on the exercise of copyright and related rights applicable to certain online transmissions of broadcasting organisations and retransmissions of television and radio programmes. Available at: <https://bit.ly/3dNH81c>

SWD (2016) 302 final, European Commission Staff Working Document, Executive Summary Of The Impact Assessment on the modernisation of EU copyright rules Accompanying the document Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market and Proposal for a Regulation of the European Parliament and of the Council laying down rules on the exercise of copyright and related rights applicable to certain online transmissions of broadcasting organisations and retransmissions of television and radio programmes, Brussels, 14.9.2016. Available at: <https://bit.ly/3dLqOxV>

Thomas J, O'Mara-Eves A (2011) Reconceptualising searching and screening: How new technologies might change the way that we identify studies," in 19th Cochran Colloquim. Available for download at: <https://bit.ly/2Abge1Y>

Yarkoni, T (2010) Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. Journal of Research in Personality, Volume 44 Issue 3, Available at: <https://bit.ly/3dLNNZA>

ANNEX 1: CASE LAW UTILIZED

- A.V. ex rel. Vanderhye v. iParadigms, L.L.C., 562 F.3d 630 (4th Cir. 2009)
- A&M Records, Inc. v. Napster, Inc., 239 F.3d 1004 (2001)
- Arica Institute, Inc. v. Palmer, 970 F.2d 1067
- Arnstein v. Edward B. Marks Music Corp., 2 Cir., 82 F.2d 275, 277
- Arnstein v. Porter, United States Court of Appeals for the Second Circuit 154 F.2d 464 (2d Cir. 1946) Frank, Circuit Judge
- Authors Guild v. Google, Inc. - 804 F.3d 202 (2d Cir. 2015)
- Bond v. Blum, 317 F.3d 385 (4th Cir. 2003)
- C-160/15, GS Media BV v Sanoma Media Netherlands BV, Playboy Enterprises International Inc., Britt Geertruida Dekker
- Campbell v. Acuff-Rose Music, Inc. Supreme Court of the United States 510 US 569 (1994)
- Case I ZR 69/08, Vorschaubilder I (Thumbnails I), German Federal Court of Justice, I ZR 69/08, 29 April 2010
- Castle Rock Entm't, Inc. v. Carol Publ. Group, Inc., 150 F.3d 132 (2d Cir. 1998)
- Field v. Google, Inc., 412 F.Supp. 2d 1106 (D. Nev. 2006)
- France, HFA v FIFA, Court of Cassation, 2 Oct. 2007, 214 R.I.D.A. 338, appeal number 05-14928
- Gaste v. Kaiserman, 863 F.2d 1061 (2d Cir. 1988)
- Harper & Row, Publishers, Inc. v. Nation Enterprises, 471 US 539 (1985)
- Kelly v. Arriba-Soft, 336 F.3d 811 (9th Cir. 2003)
- Kienitz v. Sconnie Nation, 766 F.3d 756 (7th Cir. 2014)
- Lenz v. Universal Music Corp., 815 F.3d 1145 (9th Cir. 2015)
- Marks v. Leo Feist, Inc., 2 Cir., 290 F. 959,
- Matt Hosseinzadeh v. Ethan Klein and Hila Klein, 276 F. Supp. 3d 34 (S.D.N.Y. 2017).
- Megakini.com v. Google Spain, Spanish Supreme Court, No. 172/2012, 3 April 2012
- Nichols v. Universal Pictures Corporation United States Court of Appeals for the Second Circuit 45 F.2d 119 (2d Cir. 1930)

Perfect 10, Inc. v. Amazon.com, Inc., 487 F.3d 701 (9th Cir. 2007),

Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1166 (9th Cir. 2007)

Peter Pan Fabrics, Inc. v. Martin Weiner Corp., 274 F.2d 487, 489 (2d Cir. 1960).

Studios v. Metro-Goldwyn-Mayer, Inc., 307 F. Supp. 2d 1085 (N.D. Cal. 2004)

Svensson and Others v Retriever Sverige AB (C466/12)

Whelan Assocs., Inc Associates, Inc. v. Jaslow Dental Laboratory, Inc., 797 F.2d 1222
(3d Cir. 1986)]

White v. West Pub. Corp. No. 12-cv-1340-JSR (S.D.N.Y. Jul. 3, 2014)

Wilkie v. Santly Bros., 2 Cir., 91 F.2d 978