UNIVERSITY OF TRENTO
Faculty of Law

lawtech

Trento Law and Technology Research Group
Student Paper n. 74

THE ROLE OF COPYRIGHT IN INNOVATION: A COMPARATIVE ANALYSIS OF THE LEGAL FRAMEWORK OF TEXT AND DATA MINING

EUGENIO DE BIASI

About the author

Eugenio De Biasi (eugenio.debiasi@alumni.unitn.it) graduated in Law, *magna cum laude*, at the University of Trento, under the supervision of Prof. Paolo Guarda (September 2021).

The opinion stated in this paper and all possible errors are the Author's only.

Informazioni sull'autore

Eugenio De Biasi (eugenio.debiasi@alumni.unitn.it) ha conseguito la Laurea in Giurisprudenza, magna cum laude, presso l'Università di Trento con la supervisione del Prof. Paolo Guarda (Settembre 2021).

Le opinioni e gli eventuali errori contenuti sono ascrivibili esclusivamente all'autore.

THE ROLE OF COPYRIGHT IN INNOVATION: A COMPARATIVE ANALYSIS OF THE LEGAL FRAMEWORK OF TEXT AND DATA MINING.

Abstract

This work aims to analyse the relationship between Text and Data Mining (TDM), *i.e.* those techniques that allow to derive useful information from the computational analysis of data and texts, and intellectual property law, in particular copyright and database protection law.

A first chapter, which will introduce the reader to the subject of TDM and some of its technical features, as well as its economic potential and practical applications and the main legal issues encountered in its use, namely intellectual property and data protection law, is followed by a second chapter that deals specifically with copyright and its relationship with technological innovation. The chapter will start by introducing the rights that copyright provides, and the justifications for them, their exceptions and the difficult balancing of interests in regulating them. In the second part of the chapter, the history of copyright will be investigated, starting from its origins, going through the different phases of its evolution, up to the present day, concluding how the initially well-considered balance of interests has been lost, gradually shifting towards the private interests of authors and publishers, to the detriment of the public interest.

The third chapter focuses on the problems and obstacles to TDM and on the solutions that were provided in the European context, focusing in particular on the relationship between TDM activities and the rights and exceptions to copyright and database protection as provided by the InfoSoc Directive and Database Directive. The second part deals with the long process that finally led to the recent introduction of two exceptions for TDM by the Directive on Copyright in the Digital Single Market of 2019, thus analysing in detail these clauses, presenting their numerous criticisms and advancing some alternative solutions proposed by European scholars to the issue.

The fourth chapter looks at the phenomenon from a comparative perspective, analysing and comparing the different approaches adopted by the main world powers, such as, among others, the United States, Japan and China.

KEY WORDS

Text and Data Mining - Copyright Exceptions - Database Law - CDSM Directive - Fair Use

IL RUOLO DEL DIRITTO D'AUTORE NELL'INNOVAZIONE: UN'ANALISI COMPARATA DEL QUADRO GIURIDICO DEL TEXT AND DATA MINING.

Abstract

Questo lavoro si propone di analizzare il rapporto tra il Text and Data Mining (TDM), ovvero quelle tecniche che permettono di ricavare informazioni utili dall'analisi computazionale di dati e testi, ed il diritto della proprietà intellettuale, in particolare il diritto d'autore e quello in materia di protezione dei database.

Ad un primo capitolo, che introdurrà il lettore al tema del TDM e ad alcune delle sue caratteristiche tecniche, nonché alle sue potenzialità economiche e applicazioni pratiche e alle principali questioni giuridiche che si incontrano nel suo utilizzo, ovvero il diritto della proprietà intellettuale e la protezione dei dati personali, segue un secondo capitolo che tratta specificamente del diritto d'autore e del suo rapporto con l'innovazione tecnologica. Il capitolo inizierà introducendo i diritti che il diritto d'autore fornisce, e le giustificazioni degli stessi, le loro eccezioni e il difficile bilanciamento degli interessi nella loro regolamentazione. Nella seconda parte del capitolo si indagherà la storia del diritto d'autore, partendo dalle sue origini, passando per le diverse fasi della sua evoluzione, fino ai giorni nostri, concludendo come l'equilibrio di interessi inizialmente ben ponderato si sia perso, spostandosi gradualmente verso gli interessi privati di autori ed editori, a scapito dell'interesse pubblico.

Il terzo capitolo si concentra sui problemi e gli ostacoli al TDM e sulle soluzioni che sono state fornite nel contesto europeo, concentrandosi in particolare sul rapporto tra le attività di TDM e i diritti e le eccezioni al diritto d'autore e alla protezione dei database come previsti dalla Direttiva InfoSoc e dalla Direttiva Database. La seconda parte affronta il lungo processo che ha portato infine alla recente introduzione di due eccezioni per il TDM da parte della direttiva sul diritto d'autore nel mercato unico digitale del 2019, analizzando quindi nel dettaglio queste clausole, presentandone le numerose critiche e avanzando alcune soluzioni alternative proposte dalla dottrina europea alla questione.

Il quarto capitolo esamina il fenomeno da una prospettiva comparata, analizzando e confrontando i diversi approcci adottati dalle principali potenze mondiali, quali, tra gli altri, Stati Uniti, Giappone e Cina.

PAROLE CHIAVE

Text and Data Mining – Eccezioni al Diritto d'Autore – Normativa sulla Protezione delle Banche Dati – Direttiva CDSM – Fair Use

TABLE OF CONTENTS

CHAPTER 4

COMPARATIVE ASSESSMENT OF THE RELEVANT NON-EUROPEAN LEGISLATIONS ON TEXT AND DATA MINING

INTRODUCTION

"*Data in the 21st Century is like Oil in the 18th Century: an immensely, untapped valuable asset. Like oil, for those who see Data's fundamental value and learn to extract and use it there will be huge rewards*"[1].

For several years now, we have been hearing that "*data are the new oil*". Actually, this statement, as hyperbolic as it may sound, is reductive. Oil is a fungible commodity, a limited resource, and it is nowadays unanimously recognised the urgency to replace it as an energy source as quickly as possible. Data, on the other hand, are not fungible, considering that pieces of data are different from each other, they will never run out and, on the contrary, will increase exponentially over time, thus increasing their value. Moreover, for many industries, governments and researchers, data are becoming increasingly indispensable.

The wealth and prosperity of companies and countries in the near future will therefore partly depend on the ability to exploit and process data in order to derive, through this exploitation, new and previously non-existent value.

Since data and text analysis involves various legal issues such as, primarily, intellectual property and personal data protection, the demarcation line between the success of a country and its failure will thus heavily depend on the creation of a valid legal framework, granting a proper balance of interests in the regulation of such activities and capable to protect the interests of rights holders and, at the same time, to take advantage of the potentialities of these new techniques, without unnecessarily hindering their exploitation.

This work tries therefore to investigate, from a legal perspective and above all focusing on the relationships with intellectual property, the phenomenon of Text and Data Mining (TDM), a term currently used to describe "*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*"[2].

The topic, it has to be said, is particularly challenging. It is a still relatively unexplored field of study, about which little has been written, but which nevertheless intercepts several legal areas, different from each other but equally important such as: copyright and database law, contract law, EU law, international law, competition law, as well as ethical issues, such as the protection of personal data.

The work at hand is built as follows.

The first introductory chapter explores the phenomenon of text and data mining in its "*technical*" aspects.

After a general overview that traces the roots of the matter at hand back to historically very remote times and, more recently, to the birth of the Internet, the advent of big data, the worldwide growth of interest in artificial intelligence (AI) over the last 30 years, an attempt will be made to estimate the value of the industry in both micro and macro-economic terms. The potential of these techniques in the most diverse fields of scientific, business and industry research will be briefly enunciated, in a list of applications that is by no means intended to be exhaustive.

This brief premise will be followed by a technical explanation of the type of data that can be used in a TDM project and of the process that is usually performed.

Finally, we will address in more detail the legal issues that such techniques can raise in relation to intellectual property law (above all copyright and database protection) and/or in the field of personal data protection.

---

[1] J. TOONDERS, *Data Is the New Oil of the Digital Economy*, Wired. Available at: https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/

[2] Copyright in the Digital Single Market Directive, art. 2(2): Definitions.

Copyright, probably the legal barrier that more powerfully can hinder text and data mining activities, will be the main subject of the second chapter and will be discussed in depth, starting from a short historical introduction about the recurring clashes that have taken place over the last 300 years between copyright protection and technological innovation in general. It will become evident how this troubled relationship (also known as "*copyright wars*") has always been characterized by the convergence of a great variety of interests at stake and by a markedly cyclical pattern.

A general overview of copyright will then be given, outlining the main features of the matter, its growing importance and weight up to the today's society.

In this general framework a particularly important explanation will be made of the different rationales given for copyright protection within the two legal traditions of civil law and common law, generally known and summarized as "*natural law*" and "*utilitarian theory*". Here we find an underlying difference of approach that has great implications for the construction of a copyright system, in the shaping of rights themselves and especially of exceptions, which, as a result, are distinguished between open ended exceptions and specific exceptions. These aspects are notably important and they will help us to better understand the rather different courses of action adopted among non-European countries in regulating text and data mining.

Focusing then on the history of copyright from its very beginning up to current times, we will try to detect the reasons and identify the main events that led to the today's result, where an activity such as text and data mining, which seems to make use of protected material in a way that is often referred to as "n*on-expressive*" or "*non-consumptive*", ends up being unquestionably included within the scope of the rights provided by copyright.

In chapter three we will analyze the framework of rights and exceptions to copyright as well as the database directive in force in the European Union.

At the outset, an analysis of the main rights provided by copyright and database law in Europe will hopefully permit a better understanding of why they have been identified as potential barriers to text and data mining activities. This analysis will be carried out taking into consideration the actions performed in a typical text and data mining process, as previously described in chapter one.

We will subsequently turn our attention to the exceptions that the InfoSoc and Database Directives introduced in the *acquis communitaire* of the European Union, in order to check whether such activities can be included in some of them.

Then, will follow a brief presentation of the legislative initiatives undertaken by some European countries over the past decade, introducing various specific exceptions devoted to text and data mining activities.

In the second part of the chapter, we will deal with the two different courses of action that the European Union decided to undertake in order to regulate text and data mining: the first, a licensing approach, which turned out to be unsuccessful; the second, an approach based on exceptions.

An in-depth analysis will be carried out about the legislative process that resulted in the Digital Single Market Directive and, through it, to the final introduction of two specific exceptions for TDM activities in the European system of copyright: Article 3, addressed to subjects involved in scientific research, and Article 4, which benefits a wider range of users.

We will eventually devote thorough attention to the reactions and criticisms levelled at these two exceptions, analyzing the difficult process of their implementation among the member states and alternative solutions for reforming European copyright advocated by some scholars.

The final chapter addresses this phenomenon from a comparative perspective. We will investigate the different approaches followed by several countries outside Europe, starting with the United States. U.S.A. certainly embodies the idea of innovation and technological development in the collective imagination and is the motherland of the major computer platforms. Here text and data mining activities, perceived as the nerve center of many innovative achievements, seems to be protected and allowed by the open-ended clause called "*fair use theory*". Through the analysis of the various and famous US jurisprudence cases involving similar issues about the use of material protected by intellectual property in the framework of strategic and highly innovative contexts such as, for instance, search engines, we will see how this open regulation has gradually taken shape over time.

Particular attention will be paid to Japan, a country that, embracing an alternative approach to an open clause, for more than a decade now has been particularly supportive of text and data mining practices, therefore receiving the appellation of "*text and data mining paradise*". We are referring to the particularly well-drafted specific exception introduced in 2018 and unanimously appreciated in doctrine.

The focus will then shift to some countries that, even if not belonging to the common law tradition, decided to introduce a fair use clause in their copyright legislation such as Taiwan and South Korea and, mainly, Israel, a country which in 2007 moved decidedly from a fair dealing system to a fair use clause modelled on the American one, and that, in some respects, seems to have improved upon, the American system.

Two further countries, Australia and Canada, will be analyzed. Despite their similar approach to copyright exceptions, based on Anglo-Saxon-inspired fair dealing, they differ from each other in the degree of open-mindedness among their courts, where the Australian Courts have shown a certain reluctance to consider uses of protected material as fair dealing, while the Canadian ones have ruled so more often, to the point that Canadian legislation is now substantially based on fair use system disguised as fair dealing.

Finally, for the sake of completeness, we will analyze the approach that China is trying to adopt despite the lack of a provision allowing text and data mining, wavering between the willingness to adopt a fair use approach and a specific clause approach.

CHAPTER 1

TEXT AND DATA MINING: NOTION, ORIGINS, APPLICATIONS, COMMON PROCESS, LEGAL BARRIERS

1.1. Text and Data Mining: overview, definition, goals and methods

1.1.1. Background and Overview

The advent of the digital age between the end of the 20th and the beginning of the 21st century can be certainly enumerated as one of the most disruptive technological revolutions. Since then, the world has experienced an economic and demographical boom, which eventually brought through the last two decades to the creation and accumulation of an enormous amount of data. This surge in data creation was surely made possible by the incredible increase of the storage capacity since the first rudimental computers were built in the last century. Not only, the augmented computing power of the computers themselves, thus the speed at which computers can process the data used as inputs for their calculus, increased at the same pace.

We now produce data at a pace that was unthinkable until the second part the last century. According to the McKinsey Global Institute's (MGI) "*Big Data*" report, back in 2014, the generation of information and data was predicted to increase at a rate of 40% annually[3]. This growth depends above all on the proliferation of the internet, the rise of the user-generated contents and the digitalization of traditional industries and services[4].

Businesses collect trillions of bytes of information on customer transactions, suppliers, internal operations and indeed competitors; moreover, as early as 2014 the global research community was estimated to generate over 1.5 million new scholarly articles per annum[5]; and social networking sites such as Facebook, Instagram and Twitter enable users to share billions pieces of contents per day.

This growth was further eased at the beginning of this century by the so called "*technological convergence*" given by the birth of the modern smartphones, in all respects viewable as minicomputers. A lot of the activities we used to make with different means and devices now can be made with less (if not only one) technological devices, and every time we do something we produce data that are collected by someone around the world (sending an email, writing a message, sending a photo, watching a movie).

This phenomenon is now generally called "*Big Data*", term that is often used in a business setting in two ways: the first to describe big data sets that are unmanageable by purely human analytics; the second as a metaphor for all business processes that support the usage of Big Data through a process of conversion into business value[6]. In the first instance,

---

[3] J. MANYIKA, M. CHUI, B. BROWN, J. BUGHIN, R. DOBBS, C. ROXBURGH, A. HUNG BYERS, *Big data: The next frontier for innovation, competition and productivity*, McKinsey Global Institute, May 2011, 26. *Report* downloadable from the website: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation#.

[4] S. FILIPPOV, *Mapping Text and Data Mining in Academic and Research Communities in Europe* in *Lisbon Council Special Briefing: Text and Data Mining*, 2014, 2.

[5] D. MCDONALD, U. KELLY, *The Value and Benefits of Text Mining, JISC, The Higher Education Funding Council for England*, 2012, 7.

[6] J. STRYCHARZ, *Trend analysis, future applications and economics of TDM*, Future TDM, 2014, 8. Available at: https://www.futuretdm.eu/blog/legal-policies/trend-analysis-future-applications-and-economics-of-tdm/.

it usually means that the data set is being developed and maintained in accordance with the three Vs:

- Volume (because in many areas volumes of available facts are higher than ever before),
- Velocity (they are also expanding quicker)
- Variety (they materialized in many different forms than small, well-structured data sets from the past)[7].

All these newly created data, taken individually, are of little use. But if brought together, aggregated, and analyzed, they could be a lot more useful for a number of different purposes (see paragraph 3).

In the last decades the entire world experienced an exponentially increase in the demand and value given to these data (thoroughly defined the "*Crude Oil of the Digital Age*"[8]), and this favoured the birth of the modern Data Analysis, and the so called "*Text and Data Mining"*, the set of techniques that try to give to this amount of data, different kind of data, a sense, in order to extract useful information for upgrading the decision making process in a given context.

The economic potential of Text and Data Mining is indeed enormous. 2016 estimates suggested that a total of $23.8 billion was spent globally on the Big Data Market. A considerable part of this sum, $6.4 billion, was spent on Text and Data Mining (particularly in software purchases, support and training), of which the European market counted for $2.5 billion, that was expected to rapidly grow to $10.3 billion in 2021[9].

Given the ubiquitous presence of data, the increasing economic potential (both under the micro and the macro perspective) and, as a consequence, the recent proliferation of these techniques (particularly in the US, which has significantly higher exploitation levels of data than in Europe), this phenomenon started to draw the attention of the first European countries, of the EU policy makers and of the legal community, that tried to discipline their use, in the difficult task to balance their potentialities and numerous benefits with other (previously existing) potentially conflicting interests that could be sacrificed throughout the processes.

The above mentioned, still ongoing, technological revolution, as highlighted by Powell and Snellman, drastically changed the way in which the world economy works, bringing a substitution of the old tangible assets-dominant economies to the creation of new knowledge-based information economies, in which greater reliance is put on intellectual capabilities than on physical inputs or natural resources, combined with efforts to integrate improvements in every stage of the production process[10]. The success of the near future knowledge-based economies will be thus heavily based on the best implementation of policies capable of capitalizing the most from the potential deriving from information contained in data, and thus to make informed and data-driven decision both at micro and at macro level.

---

[7] S. BUCHHOLTZ, M. BUKOWSKI, A. ŚNIEGOCKI, *Big and open data in Europe: A growth engine or a missed opportunity?*, the Warsaw Institute for Economic Studies (WISE Institute), report commissioned by demosEUROPA – Centre for European Strategy Foundation within the "Innovation and entrepreneurship" programme, 2014, 10. Available at: http://wise-europa.eu/en/2014/02/26/big-open-data-in-europe-a-growth-engine-or-a-missed-opportunity/

[8] FILIPPOV, *Mapping Text and Data Mining in Academic and Research Communities*, cit., 2.

[9] STRYCHARZ, *Trend analysis, future applications and economics of TDM*, cit., 7.

[10] W. W. POWELL, K. SNELLMAN, *The Knowledge Economy*, in *Annual Review of Sociology*, 30(1), 2004, 201.

The awareness of the large public on Text and Data Mining will indeed for sure be fundamental to boost a still stagnant innovation in every sector of the European economy and can be seen as a tool to reignite the European stalling economic engine.

Without a properly drafted regulation on Text and Data Mining the EU member states could not be able to keep up with the old and newly affirming global powers, that, as will be seen later on, are already extensively exploiting these newly born economic assets at their best potential.

This work will thus particularly focus on the most important legal problems that can arise from the undertaking of a TDM project, assess their potential overridability under the current European legal framework, then comparing this with other non-european major countries, trying to highlight the best policies and to propose improvements for our recently reformed legal context on the subject.

1.1.2. Concept and Definition of Text and Data Mining

Before going forward to discuss and analyse what text and data mining is and its various and interesting applications, it is useful to try to give a definition to such a broad and, maybe for the most, unknown concept. In order to do so, it would be better to start from a brief definition of the phenomenon that can give the reader a tasting of what will be discussed about in the next chapters, trying then to go deeper into the single parts that compose the definition. In the next chapters special attention will be given to the importance, in particular when drafting provisions regulating the subject, to give an as much as possible "*technologically neutral*" definition, capable of standing the test of time and further technological innovations.

"Text and Data Mining", often just called "*Data Mining"*, is the common used term in the research field, but it is worth noting that Data mining is just a (perhaps the most essential) phase, among the others that will be presented in Chapter 1 paragraph 5, of the process of "*Knowledge Discovery from Data* (KDD)"[11].

A correct definition of Text and Data Mining (hereinafter referred to as "TDM") can be found in the "*Study of the legal framework of Text and Data Mining*". In it, after having given different definitions from various authors, the authors propose the following definition of TDM:

"*The automated processing of digital materials, which may include texts, data, sounds, images or other elements, or a combination of these, in order to uncover new knowledge or insights*"[12].

The word "*Automated*" means that TDM is made applying automated techniques to a set of selected digital materials. At the present, computers are one of the automated techniques available, but in the future, there may be other automated techniques capable to achieve the same tasks.

The act of "*processing*" is made through different actions, such as extraction, copy, comparison, classification or other statistical analysis, etc. of data, or a mix of them.

"*Digital materials*" are all kind of contents which can be found in digital formats. "*In order to uncover new knowledge or insights*" describe the scope of TDM, thus analyzing a large

[11] J. HAN, J. KAMBER, J. PEI, *Data Mining: Concepts and Techniques, 3rd Edition*, Morgan Kaufmann, 2011, 6.

[12] J.-P. TRIAILLE, , J. DE MEEÛS D'ARGENTEUIL, A. DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* Directorate-General for Financial Stability, F. S. and C. M. U., Directorate-General for Internal Market Entrepreneurship and SMEs, I., Commission, 2014, 17.

amount of material in order to discover something new or until then unknown, i.e. relationships, patterns, otherwise difficult to find[13].

The commonly used term "*Text and Data Mining*" can be then further divided in two parts: Data mining and Text mining[14]. Data mining is the broadest concept of the two, consisting on the automated processing of various and different types of data. On the other hand, Text mining only uses textual sources to derive useful information. The common characteristic of the two processes, thus, can be identified in the extraction, made on one or different kind of sources, which can lead to the discovery of previously unknown or unnoticed knowledge, in different forms (as said in Chapter 1, paragraph 3).

Of course, although TDM is certainly possible with small amount of data, the bigger the data set the better will be the accuracy of the outcome of the process.

As obvious as it could seem, it should be pointed out that TDM, as will be seen soon, is certainly a powerful tool that helps to determine and find useful relationships and patterns within data, but it does not work by itself and does not eliminate the requirement for preparing and understanding data and analytical methods. Indeed, TDM extracts hidden information from data but is not able to assess the value of information[15]. For this reason, TDM must be viewed as a "*cooperative effort of humans and computers*" in which best results are achieved by balancing the knowledge of human experts in describing problems and goals with the search and processing capabilities of computers[16]. Thus, analysts must be able to know not only the skills necessary to properly perform TDM in the specific application domain, but also the legal context in which this activity is carried out, in order to avoid the consequences deriving from an unlawful activity.

It certainly should be policy makers' duty to implement a clear legal framework in order to boost and not hinder innovation and the related benefits that comes from these activities.

1.1.3. Goals, methods and techniques of TDM

In order to better contextualize and understand the numerous applications of TDM that will be presented in paragraph 3, it could be useful to cite the goals and the most known and frequently used methods to achieve them, providing examples for each of them.

The two main goals of TDM are generally identified with the terms "*prediction"* and "*description*".

Prediction involves using some variables or fields in the data set to predict unknown or future values of other variables of interest. On the other hand, description focuses on finding human-interpretable patterns describing the data[17].

Although the boundaries between prediction and description are not clear cut (some of the predictive models can be descriptive, to the degree that they are understandable, and vice versa), the distinction is useful for understanding the overall discovery goal.

The relative importance of prediction and description for particular data-mining applications can vary considerably.

---

[13] *Ibid.*, 17.

[14] As suggested by J. CLARK, *Text Mining and Scholarly Publishing, Report Commissioned by the Publishing Research Consortium (PRC)*, Amsterdam, 2013, p. 5.

[15] P. BHATIA, *Data mining and data warehousing: Principles and Practical Techniques,* Cambridge University Press, 2019, II, 19.

[16] M. KANTARDZIC, *Data Mining: Concepts, Models, Methods, and Algorithms, 3rd Edition*, Wiley-IEEE Press, 2019, I, 2.

[17] U. FAYYAD, G. PIATETSKY-SHAPIRO, P. SMYTH, *From Data Mining to Knowledge Discovery in Databases*, in *AI Magazine*, 1996, 17(3), 43.
Available at: https://ojs.aaai.org//index.php/aimagazine/article/view/1230.

The goals of prediction and description in TDM applications can be achieved using a variety of particular data-mining techniques, for the following "*primary data-mining tasks*"[18].

These five tasks are considered fundamental to the mining process, as they are encountered repeatedly in the context of many and the majority of data mining applications. They are:

- Link analysis or dependency modeling or association pattern mining
- Database segmentation or data clustering
- Outlier detection or deviation detection
- Summarization
- Predictive modeling or data classification

From these "*building blocks*", other specific and specialized tasks can be derived.

1. The first, "*link analysis*", is a descriptive task that aims to establish links, called associations, between the individual record, or sets of records in a database[19]. There are three specializations of link analysis:
   a. "*Associations discovery*", used to locate items that imply the presence of other items in the same events, used to define associations.
   b. "*Sequential pattern discovery*" finds patterns between events such that the presence of one set of items is followed by another set of items in a database of events over a period of time.
   c. "*Time sequence discovery*" is used to determine whether links exist between two sets of data that are time dependent.[20]

   These techniques are used, for example, by Amazon recommendation system. For instance, the usage of these methods might reveal that customers who bought a cocktail shaker and a cocktail recipe book also often buy Martini glasses. These types of findings are often used for targeting coupons/deals or advertising. Similarly, this form of data mining is behind Netflix movie recommendations[21].
2. The second task, "*data clustering*", is the descriptive method with which data are segmented into groups or clusters based on its features or attributes in order to create a group of similar records that share a number of properties[22].

   Again, another example could be made in the e-commerce field. In fact, purchasing habits of different hobbyists would look quite different from each other: gardeners, fishermen and model airplane enthusiasts would all be quite distinct. Machine learning algorithms can detect all of the different subgroups within a dataset that differ significantly from each other. This in order to make better recommendations to different kind of customers[23].
3. "*Outlier detection*", belonging to the descriptive tasks, is based on identifying the outliers in the database, which indicate deviation from some previously known expectation and norm[24].

   This method could be for example used by governments and tax agencies, which could model typical tax returns and use anomaly detection to identify specific returns that differ from this for review and audit[25].

---

[18] *Ibid.*, 37.
[19] BHATIA, *Data mining and data warehousing: Principles and Practical Techniques,* cit., 24.
[20] *Ibid.*
[21] A. FURNAS, *Everything You Wanted to Know About Data Mining but Were Afraid to Ask*, The *Athlantic*, 2012.
[22] BHATIA, *Data mining and data warehousing: Principles and Practical Techniques,* cit., 24.
[23] FURNAS, *Everything You Wanted to Know About Data Mining but Were Afraid to Ask*, cit.
[24] BHATIA, *Data mining and data warehousing: Principles and Practical Techniques,* cit., 24.
[25] FURNAS, *Everything You Wanted to Know About Data Mining but Were Afraid to Ask*, cit.

4. "*Summarization*" is a descriptive task that involves methods for finding a compact description for a set (or subset) of data[26].

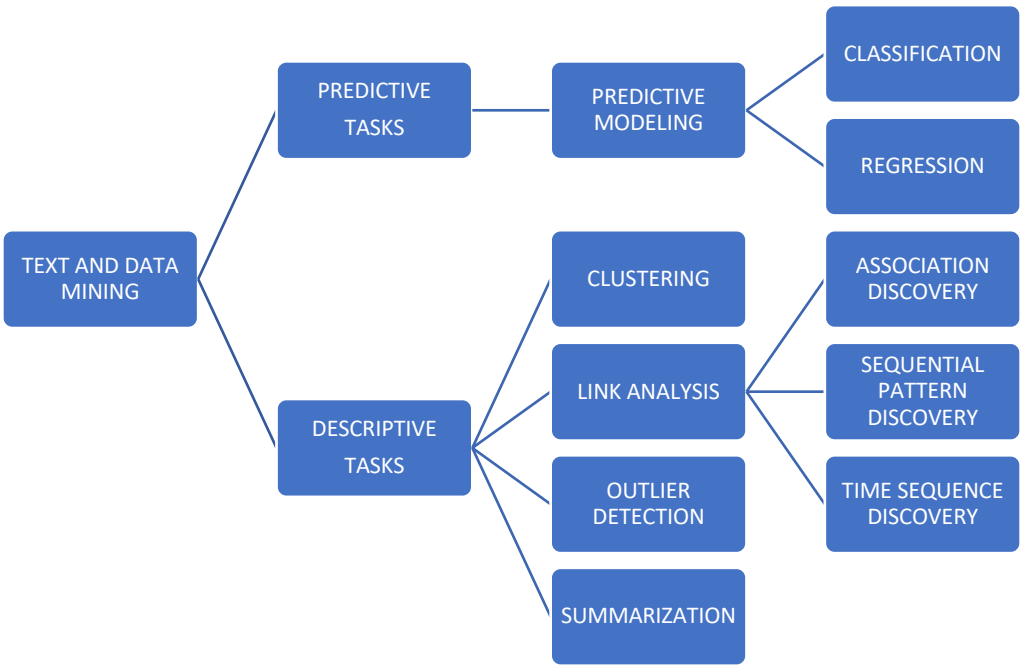
5. "*Predictive modeling*" is based on predicting the outcome of an event. Similar to the human learning experience, it uses observations to form a model of the important characteristics of some tasks. It can be divided in two types: "*classification*" and "*regression*".

a. "*Classification*" involves the discovery of a model or a predictive learning function that classifies a data item into one of several predefined classes or concepts[27].

Spam filters are a great example of this method. Indeed, large sets of emails that have been identified as spam have enabled filters to notice differences in word usage between legitimate and spam messages and classify incoming messages according to these rules with a high degree of accuracy[28].

b. "*Regression*" involves the discovery of a predictive learning function that maps a data item to a real-value prediction variable[29].

Facebook, for example, might be interested in predicting a user's future engagement based on past behaviour, from information on the platform itself, such as the amount of personal information shared, the number of photos tagged, friend requests sent or received, comments, likes, etc., and the amount of information that has been collected. Over time, this model could be refined to include or weight things differently as Facebook compares how predictions differ from observed behaviour. These results could also be used to help in site design to encourage more of the behaviours that seem to lead to greater engagement over time[30].



[26] KANTARDZIC, *Data Mining: Concepts, Models, Methods, and Algorithms,* cit., 3.
[27] *Ibid.*, 3.
[28] FURNAS, *Everything You Wanted to Know About Data Mining but Were Afraid to Ask*, cit.
[29] KANTARDZIC, *Data Mining: Concepts, Models, Methods, and Algorithms, 3rd Edition*, cit., 3.
[30] FURNAS, *Everything You Wanted to Know About Data Mining but Were Afraid to Ask*, cit.

Figure 1: Major Text and Data Mining tasks

In order to fulfil these tasks, many different data mining techniques are exploited. These techniques are closely related to some of the machine learning techniques, developed over the last fifty years, while others are related to techniques developed in statistics. For example, machine learning in the form of supervised learning is used in predictive modeling, or in database segmentation, under the form of unsupervised learning. Deviation detection is made through the use of statistics and visualization techniques[31].

Along time these techniques have been modified to deal with larger amounts of data, due to the fact that larger the data the better the chance of finding something novel and interesting[32].

1.2. Origins and History of TDM

To better understand the concept and, later on, the common process of TDM, it could be useful to outline a brief background about the rise of this discipline.

Even if one may be led to think so, the birth of Text and Data Mining isn't related just to the outbreak of the digital age at the end of the last century. Indeed, TDM owes its birth to the evolution and combination of a number of further and former events and discoveries that took place over the last three centuries.

TDM is a highly interdisciplinary, application-driven domain which derives its techniques from a variety of disciplines ranging from data science, statistics, database theory, machine learning, pattern recognition, data warehouse systems, information retrieval, visualization, algorithms, high performance computing, and many application domains[33]. Therefore, events which apparently do not refer directly to TDM, such as the invention and development of statistics, are commonly indicated as first contributions to its birth.

These are some of the historical steps that compose the overall basis to the coming up of TDM.

Referring to statistics, a milestone event dates back to 1763, when Thomas Bayes' published a paper containing the namesake theorem, which is fundamental to data mining because it describes the probability of an event based on the prior knowledge of the conditions that might be related to the event[34]. Shortly after, in 1805, Adrien-Marie Legendre and Carl Friedrich Gauss applied regression (one of the key methods used in data mining) to determine the orbits of comets and planets about the Sun[35].

Another ring in the chain leading to nowadays data mining is, of course, the advent of the so-called "*computers age*", the birth of which could be traced back in 1936, with the publication of Alan Turing's paper "*On Computable Numbers*". Here the famous mathematician introduced the idea of a Universal Machine capable of performing computations, more or less like our modern-day computers. For the first time a modern discovery made possible the collection and processing of large amounts of data[36]. The

---

31 BHATIA, *Data mining and data warehousing: Principles and Practical Techniques,* cit., 25.

32 G.K. GUPTA, *Introduction to Data Mining with Case Studies*, 3rd edition, PHI learning Private Limited, 2014, I, 24.

33 HAN, KAMBER, PEI, *Data Mining: Concepts and Techniques,* cit., 23. For further and in detail explanation of the techniques used in data mining see Chapter 1 paragraph 5 of *Data Mining: Concepts and Techniques.*

34 R. LI, *History of Data Mining*, KDnuggets. Available at: https://www.kdnuggets.com/2016/06/rayli-history-data-mining.html

35 *Ibid.*

36 *Ibid.*

Turing's theoretical machine (named "*Universal Machine*") could solve any problem described by simple instructions encoded on a paper tape, such as calculating square roots or solving Sudoku puzzles. One single machine solving any problem and performing any task for which a program could be written[37].

In 1943 Warren McCulloch and Walter Pitts created a conceptual model of a sort of neural network consisting of a multitude of artificial neurons each one capable of carrying out three tasks: receive inputs, process inputs and generate outputs[38].

These ideas finally led to the foundation, starting from the 1960', of the first companies specifically applying evolutionary computation to solve real-world problems[39].

During the 1970s sophisticated database management systems started to make possible the storage and query of terabytes and petabytes of data. In addition, the institution of the first "*data warehouses*" led to a more analytical way of viewing the data, even if extracting sophisticated insights was still a very limited mission[40].

In the 1980s sophisticated algorithms began to "*learn*" relationships among data, allowing more and more in-depth experts in the subject to reason about what the relationships meant.

In 1989 the term "*Knowledge Discovery in Databases (KDD)"* was coined by Gregory Piatetsky-Shapiro[41].

In the 1990s the term "*data mining*" was increasingly in use within the database community. Retail companies and financial communities began using data mining in order to analyze data and recognize trends with the aim of increasing their customer base, predict fluctuations in interest rates, stock prices and customers' demand[42].

Bernhard E. Boser, Isabelle M. Guyon and Vladimir N. Vapnik suggested an improvement on support vector machines obtaining a supervised learning approach capable of analyzing data and recognizing patterns used for classification and regression analysis[43].

At the beginning of this century data science began to be perceived as an independent discipline.

The specific events and discoveries described by now are the antecedents that led to the uprising of data mining. During the following decade a combination of reasons brought to the exponential growth in importance of data mining that we are experiencing nowadays.

These reasons include:

- growth in data
- decline in the cost of processing
- growth in data storage capacity
- competitive environment
- availability of data mining software[44].

The increase in the processing power of digital devices and the decline in their costs have been remarkable during the last forty years.

---

37 I. WATSON, *How Alan Turing Invented the Computer Age*, 2012. Available at: https://blogs.scientificamerican.com/guest-blog/how-alan-turing-invented-the-computer-age/

38 LI, *History of Data Mining*, cit.

39 *Ibid.*

40 *Ibid.*

41 *Ibid.*

42 *Ibid.*

43 *Ibid.*

44 GUPTA, *Introduction to Data Mining with Case Studies,* cit., I, 26.

The forecasts that Gordon Moore formulated in 1965, predicting that the amount of power in a processor would have doubled approximately every eighteen months, today prove to be correct and this trend is likely to continue also in the future[45].

The decline in the cost of disks since the 1980s (when a gigabyte disk used to cost one million dollars) is another key factor leading to the increase in data storage capacity[46].

The globalization of trade determines competition growth too, so that it's crucial in current business to get more efficiency in finding new customers and, at the same time, maintaining the old ones. Modern knowledge management is therefore becoming central in the formulation and implementation of market strategies[47].

Among the reasons previously enlisted, however, growth in data is undoubtedly the first in relevance. It is commonly recognized that this phenomenon comes from a series of causes. Just to mention a few: digitalization of industries and services, remote devices, websites (Google, Yahoo etc.), user generated contents, activities of national agencies capturing and storing data in the field of national security, social network sites like Facebook, Twitter, YouTube, storing billions and billions of data drawable from posts, images, videos uploaded by individuals. The data uploaded to these sites are stored in data centers consisting of huge buildings and connected "*cooling towers*" containing thousands of servers capable of handling billions of searches every day[48].

Even the widespread use of credit cards and Automated Teller Machine (ATM) and mobile phones is an important cause of data growth and the same can be claimed about e-commerce, a glaring example of the broader trend to trade dematerialization, which encompasses other economies like online trading in finance and online banking.

Not to mention the quite recent development of low-cost wearable sensors and smart devices that can communicate with one another (Internet of Things), that is estimated to now exceed the number of people on the planet in 2008[49].

Worth mentioning is also the increasing quantity of "*text data*" deriving from dematerialized books and publications, including both already existing-newly digitized paper-texts and new digital texts published over the last decades. This, of course, has a great impact on the work of the research community, which will be of our interest in the next chapters. Indeed, nowadays, the global research community generates over 2.5 million new scholarly articles per annum[50]. Scholarly journals are more and more available in electronic form and thus they are theoretically available to researchers. By the way, availability doesn't necessarily mean that these text data are easily at hand in order to be analysed. There is too much literature to read and, on the other hand, you can't take for granted that the search terms that are used as keywords in the search have the same meaning in the documents retrieved. Therefore, text mining offers, with its techniques, a real solution to these problems, thus supporting innovation and development of new knowledge[51].

---

[45] See: G. E. MOORE, *Cramming more components onto integrated circuits*, in Electronics, Volume 38, Number 8, April 19, 1965. Available at: https://newsroom.intel.com/wp-content/uploads/sites/11/2018/05/moores-law-electronics.pdf

[46] GUPTA, *Introduction to Data Mining with Case Studies,* cit., 4.

[47] *Ibid.*

[48] *Ibid.*

[49] C. C. AGGARWAL, *Data Mining: The Textbook*, Springer International Publishing, 2015, I, 5.

[50] M. WARE, M. MABE, *The STM Report: An overview of scientific and scholarly journal publishing, International Association of Scientific, Technical and Medical Publishers*, 2015, 6. Available at: https://digitalcommons.unl.edu/scholcom/9/

[51] MCDONALD, KELLY, *The Value and Benefits of Text Mining,* cit., 13.

In short, the amount of data created over the next three years will be more than the data created over the past 30 years, and the world will create more than three times these data in the next five years than it did in the previous five[52].

This sea of data has significant potential economic and societal value and therefore the techniques of text and data mining are required to exploit this huge and ever growing potential.

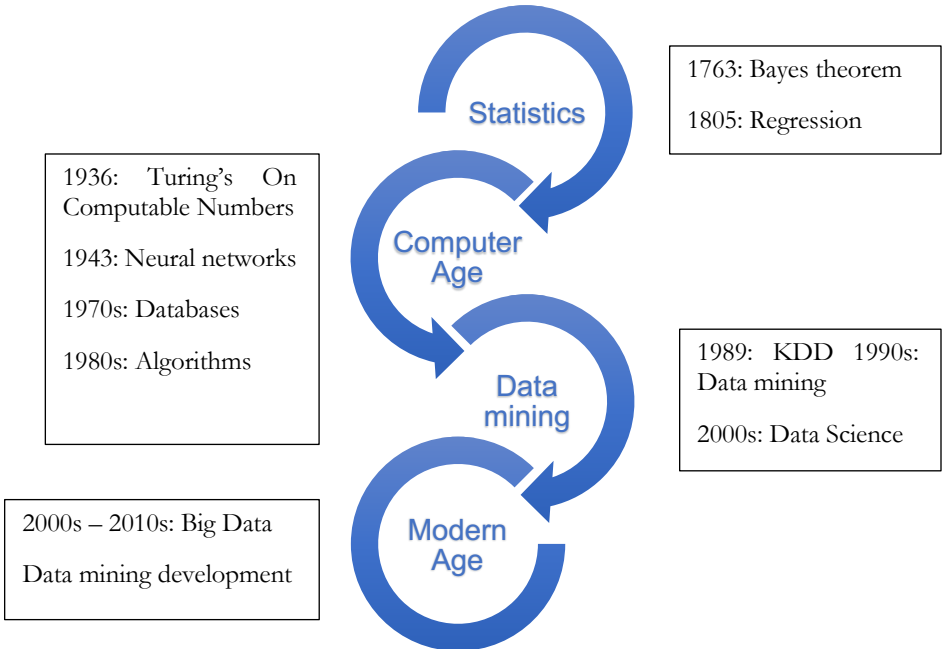


Figure 2: Birth and development of TDM[53]

1.3. Usefulness of Text and Data Mining and applications

In this paragraph the economic value and impact of TDM in the economy will be first highlighted, trying then to give some examples of real and already experimented TDM applications in different fields, such as in business and research.

1.3.1. Possible impact of TDM on the European Economy

As previously mentioned in paragraph 1, the European value of Big Data and, consequently of TDM, as a tool used to abstract value from them, is rapidly increasing at a Compound Annual Growth Rate (CAGR) of 25.7 %[54].

This activity, as will be seen soon after, among the numerous applications that are implemented, just to mention a few, in science, biology, law and education, could be also used by businesses and startups in their daily operativity, thus extracting and utilizing new insights from data.

Exploiting this potential could largely benefit the single businesses at a micro level, and the entire European economy at a macro level, that since long time is lagging behind and

[52] "*IDC's Global DataSphere Forecast Shows Continued Steady Growth in the Creation and Consumption of Data*", May 2020.
Available at: https://www.idc.com/getdoc.jsp?containerId=prUS46286020

[53] For the realisation of this figure, inspiration was taken from LI, *History of Data Mining*, cit.

[54] STRYCHARZ, *Trend analysis, future applications and economics of TDM,* cit., 61.

can't keep up anymore with countries like China and USA (just taken as examples), in terms of economic prosperity, which are already extensively using these techniques.

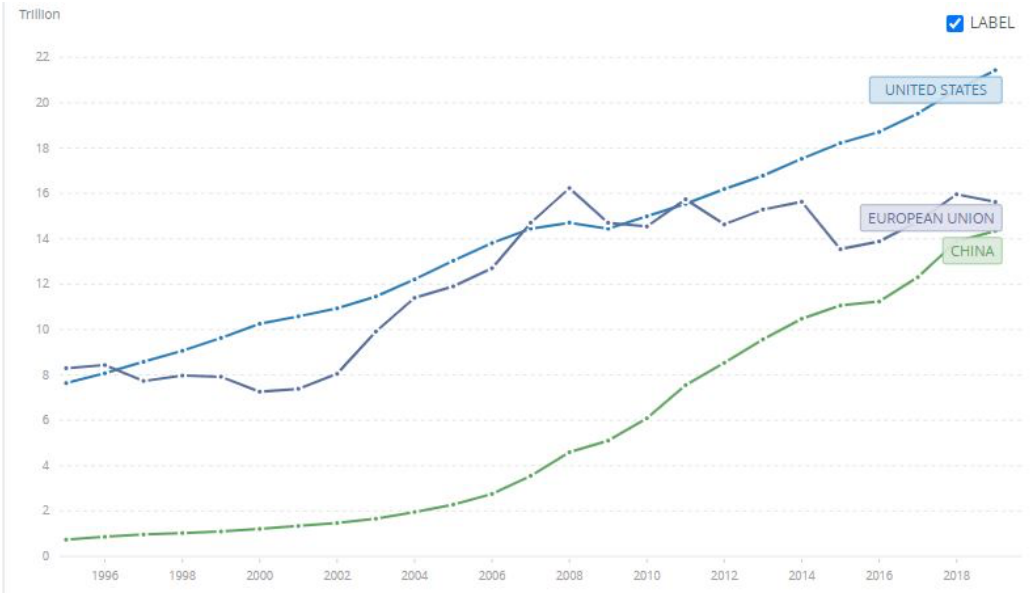


Figure 3. GDP (current US$) - United States, China, European Union. (World Bank national accounts data, and OECD National Accounts data files).

We will now try to quantify these benefits, and in doing so we will use one of the few studies made on the topic, "*Big and open data in Europe: A growth engine or a missed opportunity?*", made by the Warsaw Institute for Economic Studies (WISE Institute), and commissioned by demosEUROPA – Centre for European Strategy Foundation within the "*Innovation and entrepreneurship*" programme[55].

Even if this study focuses on the benefits brought by Big and Open data, this also encompasses TDM, as a process needed to translate big data into actionable intelligence and new value. Indeed, Big Data is a *sine qua non* condition for TDM activity to exist[56].

In this study the authors claim that Big Data are the logical prolongation of the improvements of technology observed since the second part of the last century, but they moreover think this Information and Communication Technology (ICT) revolution has many features in common with the so-called "*General Purpose Technologies (GPT)*": those technologies that form the prerequisite for other inventions, thus imposing a substantial, long term impact on our wealth and quality of life, exerting influence on the entire world economic system (e.g. exactly as electricity did in the last century). These GPTs have three characteristics, that seem to be present in the Big Data revolution. In particular:

- they are pervasive: used in most branches of the economy;
- they are potentially technologically improvable, due to cost reductions when basic technology is developed and becomes more available;
- reflect complementarity towards other innovations, enabling, creating and producing new products and progresses[57].

[55] BUCHHOLTZ, BUKOWSKI, ŚNIEGOCKI, *Big and open data in Europe: A growth engine or a missed opportunity?,* cit.
[56] STRYCHARZ, *Trend analysis, future applications and economics of TDM,* cit., 65.
[57] T. F. BRESNAHAN, M. TRAJTENBERG, *General purpose technologies 'Engines of growth'?* in *Journal of Econometrics*, 65(1), 1995, 83-108.

ICT revolution, with its Big Data prolongation, seems to be the 20th-21st century GPT.

Starting the analysis of the economic impact of TDM from a micro perspective, the economic gains that businesses can derive from thoroughly made analysis of data, can be distinguished into three categories:

- Resource efficiency improvements.

  The big amount of information produced and stored by businesses concerning production, distribution and marketing can be reduced by mining them for valuable information, capable of creating new value for the business.

- Product and process improvements through innovation.

  The data analysis could be exploited in the physical world in order to get insights about possible research & development activities and to keep track of day-to-day process monitoring and consumer feedback.

- Management improvements.

  Companies can create new value by trying to shift their corporate culture from an instinct-driven decision-making process of management to an evidence-based and data-driven one.

Indeed, according to a widely quoted study, companies that adopt data-driven decision making are on average more productive by 5-6% than their intuition-driven competitors[58]. Of course, the proportion of the gains deriving from each of the above-mentioned categories largely varies in relation with each company size and the industry in which the company operates.

The study then develops a bottom-up macro-economic model, the "*BOUDICA model*" (which stands for Big and Open Data Universal Assessment), that, starting with sectoral impact assessments, then tries to aggregate them to the entire economy. According to the estimation of this model, that dates back to 2016, the economic gains deriving from the exploitation of Big and Open Data in EU may have surpassed 200 billion EUR in 2020, that at the time counted for 1.9 % of the EU GDP, in practice an equivalent of one full year of economic growth in the EU. These gains, will mostly derive from increased business processes efficiency (made possible by sensors usage, better supply chain management in retail trade and fraud/error detection in finance, insurance and government agencies), increased competition and business creation (thanks to lower barriers to entry as a consequence of opening of public sector data), and improved allocation of production factors, resulting from the shift to a more data-driven decision making process of business managements.

To conclude, these studies make clear the indispensable need to pay special attention to this newly emerged trends and to implement a proper legislation that doesn't hinder innovation, balancing the conflicting interests and trying not to impose unnecessary and binding restrictions to these technologies.

The first necessary prerequisite, highlighted by the study as a recommendation to policy makers, is the openness of data and data availability.

In the following chapters we will try to understand if this demanding task has been successfully achieved by the European Union, through the regulation recently issued about the subject, taking into consideration the harsh criticism encountered from many quarters among the main stakeholders.

As the study in question notes: *"identifying truly transformational technologies the moment they emerge is a daunting challenge. Only after a dozen years or several decades does the economic impact of new*

[58] E. BRYNJOLFSSON, L. M. HITT, H. H. KIM, *Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?,* 2011.

solutions become discernible"[59]. And this impact is already visible in other countries, which have significant less restrictions to this kind of activities.

1.3.2. TDM applications

"*Where there are data, there are data mining applications*"[60].

It is generally acknowledged that text and data mining techniques are becoming indispensable resources in a really wide range of fields. Their exploitation tends to catch on worldwide for commercial and non-commercial purposes and, sometime, with the aim of increasing public welfare[61].

There is undoubtedly a huge number of potential benefits that such tools could deliver, if adequately implemented, in terms of innovation and new knowledge in fields such as biomedical sciences, chemistry, social sciences and even humanities, not to say about enhancements of productivity and competitiveness in the field of business, allowing additional values for consumers too.

Although it might be probably impossible to enumerate all the TDM applications, it could be useful, for the purpose of this work, to shortly cite the main and most interesting ones.

According to a recent poll of the KDnuggets' website, the areas in which data mining is currently most used are enumerated as follows: CRM/Consumer analytics, Health care, Banking, Finance, Science, E-commerce, Education and Fraud detection[62].

The uses of TDM can therefore be roughly divided into two main domains:

- business related applications
- scientific related applications

The first and most foreseeable application of TDM for business purposes is functional to the so-called "*knowledge management"*. In this area of interest, retailers, facilitated by the massive availability of electronic data stored through the general barcoding of goods, use such techniques to detect the customers' habits and, in such a manner, understand their tastes and needs, becoming able to predict their future behaviors with increasing accuracy. As a result, enterprises can take more aware decisions when planning future trade strategies[63].

Many modern businesses, such as telecommunications or mass media companies, on the other hand concentrate on the *relationship* with the customer (*Customer Relationship Management* CRM) who comes into consideration not just as a potential buyer, but for his/her "*lifetime value*". In order to build durable relationships, enterprises need increasingly accurate knowledge of their current and/or potential customers, in terms of profiling and/or segmentation within homogeneous groups on the basis of distinctive features like age, interests, attitudes, behaviors, etc[64].

Through TDM means, enterprises can therefore discover "*sales triggers"*, find issues that determine the customer's "*loyalty"* and, thanks to the already cited method "*cluster analysis"*,

---

[59] BUCHHOLTZ, BUKOWSKI, ŚNIEGOCKI, *Big and open data in Europe: A growth engine or a missed opportunity?,* cit, 9.
[60] HAN, KAMBER, PEI, *Data Mining: Concepts and Techniques,* cit. 27.
[61] See e.g. J. REICHMANN, R. L. OKEDIJI, *When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale*, in *96 Minnesota Law Review* 1362 *(2012)*, at p. 1370, stating that: "*Enlightened policymakers view these upstream data and information resources as public goods that need to be widely shared in order to produce more downstream commercial applications that advance public welfare*".
[62] M. MAYO, "*Where Analytics, Data Science, Machine Learning Were Applied: Trends and Analysis*", KDnuggets. Available at: https://www.kdnuggets.com/2019/03/poll-analytics-data-science-ml-applied-2018.html.
[63] GUPTA, *Introduction to Data Mining with Case Studies,* cit., 15.
[64] *Ibid.*

identify customers suitable for "*cross-selling products"* etc. In this latter respect, for instance, by means of "*Market Basket Analysis"*, *associations* can be spotted in the customers' habits, discovering that sometimes the purchase of a particular good tends to occur frequently in association with the purchase of other goods. It must be clarified that "*Market Basket Analysis"* is just a particular application of A.R.M. ("*Association Rules Mining"*), a technique widely known and exploited in marketing, customer segmentation, medicine, electronic commerce, classification, clustering, web mining, bioinformatics and finance[65].

About the potential impact of the exploitation of these techniques in commerce and the capacity of TDM to lead the development of technology businesses there is general recognition, and not for nothing market success of certain companies around the world is increasingly linked to their capacity to mine information about their current and/or potential customers' habits and tastes[66].

Moreover, in finance Banks use TDM techniques to assess the credit worthiness of their customers by mining, for example, a customer's historical business transactions coming from many sources with the final aim to predict the chances of a customer paying back a loan[67].

Among the numerous applications in business "*fraud detection*" is worth mentioning for its relevance and frequency. Fraud detection through TDM can be performed (with the *outlier detection* method) by analyzing data of organizations in order to find patterns and relationships that deviate from an expected norm, finding outliers that may be possible locations for fraud, and thus object to further investigations[68].

Moving on to science related applications, especially in biology or medicine, text mining is already extensively used as a tool applied to wide corpora of scientific literature capable of automatically generate series of *hypothesis* that, later on, can be tested by scientists in laboratory[69].

In a study committed to JISC (Joint Information Systems Committee), a United Kingdom not-for-profit company whose role is to support institutions of higher education and research, about the Value and Benefits of Text Mining, six broad categories of use of Text Mining have been identified. The list doesn't, after all, claim to be exhaustive. The categories, referring to all the possible uses of TDM techniques in research process, can be shortly summarized as follows:

- Systematic reviews of literature, in order to identify literature that should be reviewed by researchers in a particular field
- Developing of new hypotheses, where articles are text mined to identify interesting intermediate topics and linkages
- Testing of hypotheses, by mining documents to see if their content confirms a hypothesis
- Generation of reusable representations, like concept maps which present distilled knowledge in a concise form.
- Assessment of the quality of documentation,
- Enhancement of usability of the research base.[70]

---

[65] *Ibid.*

[66] See, for example: "*How Tesco became Britain's top supermarket*", MoneyWeek, 2007. Available at: https://moneyweek.com/31267/how-tesco-became-britains-top-supermarket

[67] BHATIA, *Data mining and data warehousing: Principles and Practical Techniques,* cit., 20.

[68] *Ibidem.*

[69] M. BORGHI, S. KARAPAPA, *Copyright and Mass Digitization: a Cross-Jurisdictional Perspective,* Oxford University Press, 2013, 48 (quoting 'Text Mining and IP', Submission to the Independent Review of Intellectual Property and Growth from the National Centre for Text Mining, University of Manchester, May 2001, 2).

[70] MCDONALD, KELLY, *The Value and Benefits of Text Mining,* cit., 15.

For example, in the biomedical sciences, D.R. Swanson used text mining-like approaches in order to hypothesize how pre-existing drugs could be used to target different diseases including Alzheimer's Disease and these hypotheses were later on validated experimentally[71].

It has been underlined that text mining goes far beyond the mere "*information retrieval*", the task of which is just providing many documents to wade through[72]. Statistical analysis of co-occurrence of biomedical concepts describing genes, drugs, and diseases in a large number of articles allows the arising of novel relationships between concepts that have high probability of being biologically valid[73]. In biomedicine are available applications capable of carrying out "*deep searches*" making automatic connections between data contained in articles, adding precious information about genes and diseases. Text mining applications are more and more implemented in the search led through digital storage places including scientific literature and other possible textual resources, as long as they are in digital format. It can be claimed that these techniques allow researchers to see things that are not visible by the mere reading and analyzing of the single texts that compose the same compounds. It has been underlined that repositories containing large amounts of textual resources own the "*collective intelligence of a library"* which is different and much more than the mere summary of the individual texts[74]. Here lies the difference between "*information retrieval*" and "*value extraction*" provided by such text mining techniques, characterized by a general underlying rule "*the larger the data the better the chance of finding something novel and interesting"* and, meanwhile*, "large amounts of data provide more confidence in the achieved discoveries"*[75].

Computational linguistic analysis on large corpuses of texts is currently used in the field of *humanities* by means of software applications of both research and commercial value[76] that allow the extraction of quantitative and qualitative data about cultural and linguistic trends over the time and through the space, that is currently defined with the neologism "*culturomics"*[77].

Interesting applications of TDM go under the name "*Educational data mining*" and represent a substantial effort towards the achieving of increasing efficiency in academic activity. This discipline applies data mining techniques to datasets operative in the educational field, such as online education platforms or the so-called "*MOOC*" (Massive Open Online Courses). These techniques try to extract useful information from the data produced by students while using the platform, for instance: time spent on the platform, interactions and activities on the pages, mouse usage and eye movement, just to cite a few. This subfield of data mining promises to bring benefits to both students and teachers, not considering the undeniable advantages that the Universities themselves can gain from such an analysis. For example, it can predict school drop-out and thus give the university the opportunity to undertake the necessary countermeasures. Moreover, students can customize their learning experience and improve their productivity and Teachers have, at the same time,

---

[71] S. ANANIADOU, D.B. KELL, J. TSUJII, *Text mining and its potential applications in systems biology,* in *Trends in biotechnology* 24(12), 2007 (quoting D. SWANSON, N. SMALHEISER, *Assessing a gap in thebiomedical literature: magnesium deficiency and neurologic disease*. In *Neurosci. Res. Commun.15*, 1–9).

[72] BORGHI, KARAPAPA, *Copyright and Mass Digitization: a Cross-Jurisdictional Perspective*, cit., 4.

[73] *Ibidem*, 48 (quoting R. FRIJTERS, *Literature Mining for the Discovery of Hidden Connections between Drugs, Genes and Diseases*, in *PLoS Computational Biology 6/9*, 2010).

[74] BORGHI, KARAPAPA, *Copyright and Mass Digitization: a Cross-Jurisdictional Perspective,* cit., 49.

[75] GUPTA, *Introduction to Data Mining with Case Studies,* cit., 24.

[76] For instance: *Google Ngram Viewer,* outcome of cooperation between Google and researchers from Harvard University, consisting of a graphic tool that enables the tracking of the frequency of given letters combination across 5.2 million digitized books.

[77] BORGHI, KARAPAPA, *Copyright and Mass Digitization: a Cross-Jurisdictional Perspective,* cit., 49.

the opportunity to monitor their teaching effectiveness and to take measures to improve it, thanks to the feedback received by students[78].

Just recently, also the legal community began to propose the possibility of applying TDM techniques to the legal context, in particular to legal research, in order to "*identify rules, correlations, and patterns that may point towards new hypotheses, the validation of untested theories or the visualization of results in a way that can offer novel insights*"[79]. We are talking about a newly born approach to legal research proposed as an autonomous item in the list of the legal methodologies, known as "*autonomous legal systems*". It involves a real paradigm-shift from the "*deductive method*" (according to which one derives certain conclusions from a preconceived set of principles, categories and rules, usually taught to the young lawyers at university), to the newly proposed, statistical and data-driven oriented "*inductive method*"[80].

Other applications related to the legal research field can be enumerated: discovery procedures in trials, automatic summarization or argumentation extraction of court decisions, knowledge extraction from legal statues, or assisting documents and contract drafting[81].

1.4. Data to be mined

TDM can be applied to any kind of data, as long as data are meaningful with a target application. The most basic forms of data used for mining applications are database data, data warehouse data and transactional data (e.g. customer's purchase, flight booking or user's clicks on a webpage), but given the significant development of the discipline since its birth and its newly emerged applications also other forms of more complex data are now disposable. To give an example:

- data streams (e.g. video surveillance and sensor data, continuously transmitted),
- ordered/sequence data (e.g., historical records, stock exchange data),
- graph or networked data (e.g., social and information networks),
- spatial data (e.g., maps),
- text data,
- the World Wide Web[82].

The types of data used in TDM projects largely depend on the application domain to which TDM is applied. For example, in business applications the most used data type are transactional data. This kind of data allow companies to trace their customers' buying habits and can be easily translatable into actions to reaching higher sales targets. Other types of data widely used not only in business applications, but also in scientific related applications are text data, which seem to be used far more often than the others[83].

---

[78] G. PASCUZZI, *Le valutazioni degli studenti le fa l'algoritmo (a proposito di Educational Data Mining)*, in *L'Adige*, 13 ottobre 2020. Available at: https://www.giovannipascuzzi.eu/2020/10/13/le-valutazioni-degli-studenti-le-fa-lalgoritmo-a-proposito-di-educational-data-mining/.

[79] T. MARGONI, *Text and Data Mining in Intellectual Property Law: Towards an Autonomous Classification of Computational Legal Methods*, CREATe Working Paper 2020/2021. Available at http://eprints.gla.ac.uk/215611/.

[80] *Ibid.*

[81] M. TRUYENS, P. VAN EEECKE, *Legal aspects of text mining,* in *Computer Law & Security Review 30 (2)*, 2014, 154.

[82] HAN, KAMBER, PEI, *Data Mining: Concepts and Techniques,* cit., 8.

[83] For an overview of the most popular used data see: G. PIATETSKY, "*What Data You Analyzed – KDnuggets Poll Results and Trends*", KDnuggets. Available at: https://www.kdnuggets.com/2017/04/poll-results-data-analyzed.html

Since TDM is a continuously evolving computer science branch, it will certainly embrace other and new data types as they will emerge in the future. It is important to keep in mind that, in many applications, multiple types of data could be present. For example, in web mining, there often exist text data and multimedia data (e.g., picture and videos), on web pages, graph data, and map data, leading to fruitful findings due to the mutual enhancement and consolidation of such multiple sources, but making the process more difficult for the data cleaning and data integration part, as will be seen[84].

Data used for TDM purposes may be "*closed*" or "*open*" to a varying degree.

In order for TDM process to properly function, without being hindered, data used to perform TDM should have some ideal characteristics of "*openness*". In particular, starting from an economic point of view, they should be available without any imposition of monetary constraint to the users (e.g. open web content). Moreover, they should be published in a way enabling reuse for both commercial and non-commercial application, without any legal constraint. They should be provided in machine-readable (automatically readable and processed by computers) formats (such as CSV, JSON, XML etc.). Finally, they should be well described by correct metadata, or tags, in order to more easily perform data linking in certain TDM applications which uses these kind of methods[85].

As will be seen in the next paragraphs, data used in TDM projects seldom have these characteristics. In particular, this work will focus on the first and the second characteristics, highlighting the problems that arise when trying to access and collect the material to be mined, where usually permissions or licenses are required, and more importantly the legal problems relating to the first part of the data processing, particularly intellectual property and copyright.

1.5. Common procedure of TDM process: access to content, extraction and/or copying of content, mining of text and/or data and knowledge discovery

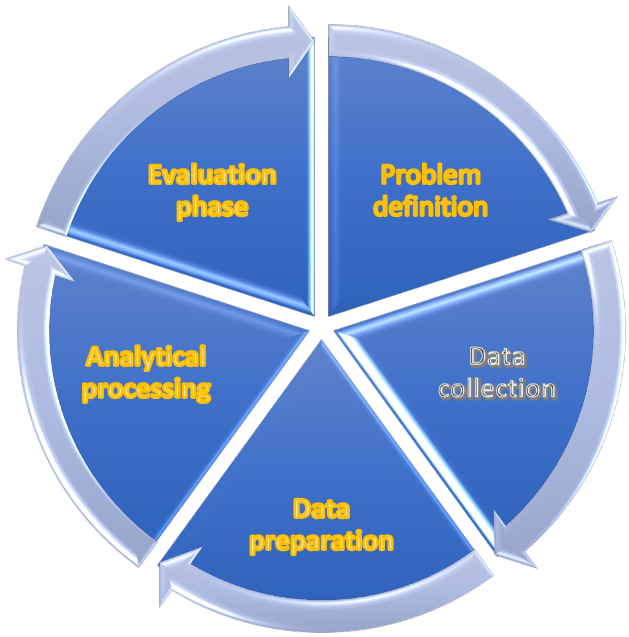


Figure 4: Typical TDM process phases

[84] HAN, KAMBER, PEI, *Data Mining: Concepts and Techniques,* cit., 8. For further explanation about the different types of data.

[85] STRYCHARZ, *Trend analysis, future applications and economics of TDM*, cit., 13.

Before passing to the main subject of this work, that is to say the legal issues that can arise in undertaking a text and mining activity, their possible solutions and the peculiar relationship between TDM and intellectual property and, more specifically, copyright protection, it could be useful to underly a "*technical*" aspect of TDM, relating to the procedure or, in other words, "*how*" TDM works. It has to be stressed that it doesn't exist and wouldn't be possible to spot a single procedure commonly shared by all the TDM techniques that, as a scope driven discipline, mainly depend on the object and on the target pursued. Nonetheless, generalizing, we can pinpoint some steps than can be identified and considered as belonging to the majority of text and data mining processes.

We will start this paragraph by giving a more technical and analytical description of the common process of TDM. Then we will try to conceptualize and to summarize the various phases in order to make it suitable for a study from a legal perspective, finally going to present the major legal issues discussed in the next paragraphs that each macro-phase of TDM poses.

As said before, depending on the different goals pursued and on the different techniques deployed, TDM works in a wide range of ways, the only common element being the analysis and extraction of associations and relationships among concepts and the identification and verification of patterns. Nevertheless, data analysts use a pipeline of processing, containing many different phases.

The workflow of a typical data mining application includes the following phases:

1. Problem definition and formulation of hypothesis.

   The focus of this phase is to understand the requirements and objectives of the project. Once the problem is defined, it can be formulated as a data mining problem (e.g. the business problem: "*how can I sell more of my products to customers?",* can be translated into a data mining problem such as "*which customers are most likely to buy my product?*")[86].

   Of course, each TDM project is performed in a specific application domain. For this reason, knowledge about the domain in which TDM is applied is necessary in order to come up with a proper problem statement, and for this purpose it would be better to build a strict interaction between the application expert and the data mining expert. About the hypothesis, there may be several of them for a single problem[87].

2. Data collection

   This is the phase in which data are generated and collected. Its importance is generally undervalued, but in contrast it is critically important because good choices at this point can benefit and have a big impact on the entire process. In this step, two scenarios are possible. The first, named as "*designed experiment"*, occurs when the data-generation process is under the control of an expert. On the other hand, the so-called "*observational approach",* in which the expert cannot influence the data-generation process but just observe it. The most widely used approach among the various applications is the latter[88].

   For the purpose of collecting the data it may be necessary to use specialized hardware such as a sensor network, manual labor (e.g. collection of user surveys), or software tools, such as Web document crawling engine to collect documents. After the collection phase, the data are generally stored in databases or in data warehouses for processing[89].

---

[86] BHATIA, *Data mining and data warehousing: Principles and Practical Techniques,* cit., 22.
[87] KANTARDZIC, *Data Mining: Concepts, Models, Methods, and Algorithms*, cit., 7.
[88] *Ibidem.*
[89] AGGARWAL, *Data Mining: The Textbook,* cit., 3.

3. Data preparation

This phase generally consumes about 90% of the time of a TDM project[90]. When all the data needed to perform TDM are collected they rarely are in a form suitable for processing. For example, data may be encoded in complex logs or free-form documents, or different types of data are mixed together. In order to make these data suitable for processing, three steps are generally followed by analysts:

a. Feature extraction

Once collected, the data at the disposal of the analyst are raw and there is little guidance on how they should develop into database features usable and meaningful for processing. Obviously, it's up to the analyst to abstract out and identify the features that are most useful, considering the aim of the processing. This implies that the analyst should be deeply capable of understanding the application domain in which he works, in order to detect and extract the relevant features[91].

b. Data cleaning

Within the data preprocessing activities, we can pinpoint "*data cleaning*".

Indeed, the extracted data may have erroneous or missing entries and therefore some records may need to be dropped, missing entries may need to be estimated and inconsistencies may need to be removed[92].

Cleaning data usually includes a common task called "*outlier detection"*, where the word outlier refers to unusual data values inconsistent with most observations that can seriously affect the work of the analyst so that outliers have to be removed or, alternatively, the method deployed has to ben insensitive to them[93].

c. Feature selection and transformation

Among the data preprocessing activities also feature selection and transformation can be enumerated.

Since high dimensional features could be noisy or may add error to the data mining process, in order to make data amenable for the analyst, it may be necessary to remove irrelevant or transform the current set of features to a new data space[94].

Data transformation, on the other hand, involves the transformation of data set with a particular set of attributes in another set of attributes of the same or of a different type (e.g. attribute: age, may be partitioned into ranges)[95].

4. Analytical processing and algorithms, estimation of the model

It's the phase in which data mining algorithms are applied to the processed data in order to achieve the goals set in the problem definition phase. Usually, in practice, the implementation is based on several of the different models and techniques already presented in paragraph 1, and selecting the best is an additional task, so that the success of the operation mainly depends on the ability of the analyst.

5. Evaluation phase, interpretation of the model, and conclusions

During this phase, the model results are evaluated to determine whether or not they satisfy the originally stated goals.

---

[90] BHATIA, *Data mining and data warehousing: Principles and Practical Techniques,* cit., 23.
[91] AGGARWAL, *Data Mining: The Textbook,* cit., 5.
[92] *Ibidem.*
[93] KANTARDZIC, *Data Mining: Concepts, Models, Methods, and Algorithms*, cit., 7.
[94] AGGARWAL, *Data Mining: The Textbook*, cit., 5.
[95] *Ibidem.*

This procedure is the procedure commonly used in Knowledge Discovery in Databases (KDD) projects. However, a number of other processes may apply that usually don't differ too much from each other, the only difference being the stress given to particular phases among others, depending on the application domain in which TDM techniques are applied.

For sure, the most used process[96], especially in business domain applications, is the so-called CRISP-DM process, which stands for Cross-Industry Standard Process for Data Mining, that was created in 1998 from a consortium of vendors and users including DaimlerChrysler, SPSS and NCR. They claimed that this process was more practical, successful and widely adopted. It consists of six steps that are: business understanding, data understanding, data preparation, modelling, evaluation, deployment[97][98].

To conclude the "*technical*" part of this work, to sum up the concepts treated in these first paragraphs and in order to make the whole process and some of the above mentioned methods clearer to the reader, it could be useful to give a practical example of a typical data mining process in a business application domain, extracted from a Data Mining textbook[99].

> "*Consider a scenario in which a retailer has Web logs corresponding to customer access to Web pages at his or her site. Each of these Web pages corresponds to a product, and therefore a customer access to a page may often be indicative of interest in that particular product. The retailer also stores demographic profiles for the different customers. The retailer wants to make targeted product recommendations to customers using the customer demographics and buying behavior.*
>
> *The first step for the analyst will be the collection of the relevant data from two different sources: the set of Web logs at the site and the demographic information within the retailer database that were collected during Web registration of the customer.*
>
> *The log may contain hundreds of thousands of such entries. The customer from a particular IP address can be identified using the previous login information, by using cookies, or by the IP address itself, but this may be a noisy process and may not always yield accurate results. The analyst would need to design algorithms for deciding how to filter the different log entries and use only those which provide accurate results as a part of the cleaning and extraction process. Moreover, the raw log contains a lot of additional information that is not necessarily of any use to the retailer. In the feature extraction process the retailer decides to create one record for each customer, with a specific choice of features extracted from the Web page access. For each record, an attribute corresponds to the number of accesses to each product description. Therefore, the raw log need to be processed, and the accesses need to be aggregated during this feature extraction phase. Attributes are added to these records for the retailer's database containing demographic information in a data integration phase. Missing entries from the demographic records need to be estimated for further data cleaning. This results in a single data set containing attributes for the customer demographics and the customer accesses.*
>
> *At this point, entering the analytical processing phase, the analyst has to decide how to use this cleaned data set for making recommendations. He or she decides to determine similar groups of customers and make recommendations on the basis of the buying behavior of these*

[96] According to a poll, "*What main methodology are you using for your analytics, data mining, or data science projects? Poll*", KDnuggets. Available at: https://www.kdnuggets.com/polls/2014/analytics-data-mining-data-science-methodology.html

[97] GUPTA, *Introduction to Data Mining with Case Studies,* cit., 37.

[98] For a more detailed explanation of the CRISP process phases, see P. CHAPMAN, *CRISP-DM 1.0 Step-by-step data mining guide*, 2000. Available at: https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2015-16/kdd/files/CRISPWP-0800.pdf

[99] AGGARWAL, *Data Mining: The Textbook,* cit. 4-5.

similar groups. In particular, the building block of clustering, among others, can be used to determine similar groups. For a given customer, the most frequent items accessed by the customers in that group are recommended".

Nevertheless, starting to look at the TDM process from a legal (and less technical) perspective, three common steps have been aptly stressed in TDM processes[100] with the specific aim of identifying the possible legal issues that might arise in connection with each of them. The generally renown steps are the following:

1. access to content;
2. extraction and/or copying of content;
3. mining of text and/or data and knowledge discovery.

The work will now draw attention to the first of these steps, access to content, in order to determine in which cases TDM process is (at the moment, just considering this phase) freely doable or, on the other hand, could be hindered or delayed due to the necessity of obtaining the permits of undertaking such activities from the relevant right holders.

Later on, in the following paragraphs, the phase of extraction and/or copying of content will be covered, the one into which legal issues are more likely to arise during a TDM process.

1.5.1 Different levels of access to data

Access to content, whether text or data, is the first and necessary step to TDM activities and it provides, through search, the matter to be mined.

On the other hand, variable restrictions in access to data are paramount to determine the legal framework (e.g. intellectual property rules) into which these activities can lawfully be performed. Intellectual property rights may, in certain circumstances (e.g. clauses in contracts with publishers that explicitly permits or exclude TDM) be supplemented and reinforced or, on the contrary, derogated by contractual clauses[101]. Indeed, contents are sometimes freely accessible and some others not (e.g. access permissions, such as licenses, may be required). Of course, even if one finds that content to be mined is freely accessible it might be possible that legal restrictions will apply during the next phase, namely extraction and/or copy of data for the purpose of TDM activities, as will be seen soon[102].

Four different levels of access to data are currently distinguished, depending on the increasing degree of restrictions that disciplines the access to data[103]:

- all to all (data freely accessible on the web)
- many to many (data created and shared on social networks)
- one to many (publishers' data with access restricted by contractual clauses);
- one to one (regarding confidential data accessible by means of confidentiality or nondisclosure agreements NDAs containing restrictions about the use of data).

The latter level of access doesn't come into consideration in the field of TDM. The constraints binding the contractual parties in these cases are indeed directly related to the

---

100 E. ROSATI, *Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and its Role in the Development of AI Creativity* in *Asia Pacific Law Review*, 2019, 27(2), 8.

101 TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* cit., 20.

102 ROSATI, *Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and its Role in the Development of AI Creativity*, cit., 8.

103 TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* cit., 18.

high degree of confidentiality of the subject at hand, which is frankly irreconcilable with the TDM purposes.

Starting from the first level of access (all to all) it is generally acknowledged that the application of TDM techniques in order to extract information from web data such as web documents, hyperlinks between documents or usage logs of websites, is a freely disposable activity which doesn't require authorization from relevant copyright holders[104]. In fact, considering the possibility of the existence of terms of uses governing the access to the page, it seems difficult to prove the consent of the user, who should be somehow asked to explicitly accept terms and conditions, for example via a link on another page or with a "*click*", which in many jurisdictions could not be considered binding. Terms and conditions could, for example, prohibit the crawling of the pages by a robot, but if visitor does not need to accept them before continuing on his visit to the website, they could not guarantee the enforceability of the clauses. The only exception, which could happen in a limited number of cases, would be the case of a website designed in a way that the user could not pretend that he has not seen, read and accepted Terms and Conditions[105].

There are several worth mentioning private originated initiatives (sometimes also entered in public organizations), such as the "*Budapest Open Access Initiative"* and the "*Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*", that aim at promoting, on a voluntary basis, open access to data not only to the research community but to the public at large too, in order to make research increasingly free and available to anyone with a computer or an internet connection. These initiatives are per se pro-TDM, because they grant to the public the right to use protected works to the fullest extent permitted by law, including for data analysis purposes[106].

"*All to all*" level of access will probably increasingly grow in importance in the near future also as a result of the provisions of the "*Digital Agenda for Europe*" set out by the European Commission. It has been established as an action to be achieved in 2020 "*to open up public data resources for re-use*" by the private sector[107], for commercial or non-commercial purposes[108], in order to facilitate re-use of existing documents held by public sector bodies. These provisions tend, therefore, to enlarge the corpus of works available for re-use (the definition of which, given by PSI Public Sector Information European Directives, is broad and includes data analysis) imposing specific obligations on public administrations to deliver documents in standard open and machine-readable digital formats together with their metadata[109]. The public fields falling under these obligations include key-sectors such as libraries, museums, universities, health and social security administrations, the activities of which are financed by public money. Even if the right to re-use public data, under strict rules

---

104 ROSATI, *Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and its Role in the Development of AI Creativity,* cit., 10.

105 TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN*, Study on the legal framework of text and data mining (TDM),* cit., 21.

106 *Ibid.*, 25-26.

107 Made possible by Directive (Eu) 2019/1024 of The European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, which amended the previous version (2013/37/EU) of the first PSI directive, dating back to 2003.

108 Art 3 Directive (Eu) 2019/1024, General principle: "Subject to paragraph 2 of this Article, Member States shall ensure that documents to which this Directive applies in accordance with Article 1 shall be re-usable for commercial or non-commercial purposes in accordance with Chapters III and IV".

109 Article 5 Directive (Eu) 2019/1024 Available formats: "Without prejudice to Chapter V, public sector bodies and public undertakings shall make their documents available in any pre-existing format or language and, where possible and appropriate, by electronic means, in formats that are open, machine-readable, accessible, findable and re-usable, together with their metadata. Both the format and the metadata shall, where possible, comply with formal open standards."

posed by EU in order to limitate price to merely cover the costs involved, can be subject to payment and exceptionally fall under the above mentioned third level of access (one to many), the general rule should allow access on a "*no condition basis*", so that the "*all to all*" category is the one that comes into account[110].

When access is considered "*restricted*" (level 2, many to many, and 3, one to many, of access), contractual clauses agreed to by the parties in the first place and intellectual property law for the remainder, interplay in regulating TDM. In these categories, as said, the interplay between intellectual property rights and contractual clauses is important, and need to be further analyzed, in order to determine if TDM activity by the user is from that clauses further restricted or, on the contrary, relieved by intellectual property (copyright or database legislation), thus, giving the user the extra burden of considering if IP legislation permits contractual overridability of its provisions by contract or not, as will be seen.

Regarding the second level of access, namely "*many to many*", relating to data that are created and shared on social networks, the availability of data to the analyst is strictly dependent on the private account settings of the users (public/private/custom)[111], and the Terms of Use[112] of the platforms[113].

The last level of access, "*one to many*", the most important for the purpose of this work, usually refers to data covered by contracts with publishers and/or repositories. These contracts may contain contractual clauses which can restrict or limit access to data and, in our case, prohibit data analysis or other text or data mining activities. Access is allowed just to the authorized users, that have previously accepted terms of use[114]. Some contracts, moreover, explicitly exclude TDM activities to the accessed content (e.g. "*automated processes may not be used to systematically retrieve batches of articles from Europe PMC web site […]". Europe PubMed Central*)[115]. However, some countries (e.g. UK) explicitly prohibit any contractual restriction to the right of the user with lawful access to data to mine them, as will be seen.

When a license is required the licensee is entitled to undertake activities on the contents put at his disposal, the range of which mainly depends on the scope of the license that cannot be exceeded.

Relevant problems could arise when the identification of the *relevant right-holder* is needed in order to ask permission, and this is a task that not always can be easily achieved. Works and subject matters protected by copyright, e.g. in a not negligible percentage of cases within libraries, relate to "*non identified"* or "*non located"* rights holders. This phenomenon is worldwide known as "*orphan works"* and it poses not negligible questions such as costs of identification of right holders or the estimation of risks of legal actions, that become crucial issues when, for example, institutions have to decide whether or not to undertake digitization projects of works. This source of uncertainty led to the adoption of legislative initiatives at European and national level such as *EU Orphan Works Directive,* providing new mandatory exceptions.

It has to be taken into consideration that freedom of access doesn't necessarily mean that contents are devoid of legal restrictions, depending on the legal rights pending on them, as will be seen immediately soon.

---

110 TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* cit., 22-24.

111 For an example of a platform's account settings, see Facebook profile privacy policies: https://www.facebook.com/about/basics/manage-your-privacy/profile

112 Facebook Terms of Service: https://www.facebook.com/terms.php

113 TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* cit., 18.

114 *Ibid.*, 19.

115 *Ibid.*, 20.

1.6. Legal and practical barriers to Text and Data Mining activities

The aim of this paragraph is to give the reader a short summary of what will be discussed and analysed in details in the remainder of the work (intellectual property), and, on the other hand, briefly discussing other legal and non-legal barriers that are encountered in different kind of TDM projects.

The barriers to TDM could be divided in two: on one side the legal barriers related to Intellectual property protection (comprehending copyright law, database law, and contract law), while on the other side data protection law[116]. They differ for the fact that intellectual property laws aim at protecting the economic interests of the right-holder, while data protection law aims at protecting the privacy of the data subject.

1.6.1. The difficult relation between Text and Data Mining and Intellectual Property Protection

With the birth of the modern data analysis and TDM techniques, a number of different business and scientific applications that derive new value by extracting knowledge from data (whether text or other types of data) saw the light.

These newly born applications can operate just if they have unrestricted access to big amounts of data and, however, they can be potentially hindered by the existence of legal barriers. Among these (others will be discussed soon after) copyright and the related database law are the most important ones, but, before discussing them, contract law also needs to be addressed.

Even contract law can indeed come into consideration as a legal barrier, because contractual clauses signed by the users that seek access to content to mine and publishers, can alleviate or, on the contrary, reinforce the copyright legal provisions. As seen before, one can get access to data through different means and, especially when TDM is conceived for scientific purposes on publications, data/text access may be somehow restricted by rightsholders. Legal troubles can therefore arise because these new applications can possibly clash against already long-lasting existing business models[117].

Copyright usually works granting automatically exclusive rights to the authors of scientific literature. The same authors commonly transfer their rights to commercial publishers[118]. The typical publishers' business model depends on paid access to their databases, either under single payment of a subscription or license fee[119]. When TDM is performed relying upon materials obtained by publishers, different kinds of obstacles can arise. In the first place, publishers can charge prohibitive fees to their licenses to preserve these benefits to researchers. The ability to negotiate favorable license agreements varies of course from publisher to publisher, forcing the prospective TDM researchers to deal with a patchwork of different rules they must abide for each content source they wish to include in

116 Other non-legal barriers to TDM can be found in M. CASPERS, L. GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, FutureTDM report, 2016, 107. Available at: https://project.futuretdm.eu/publications/

117 ROSATI, *Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and its Role in the Development of AI Creativity*, cit., 6.

118 H. REICHMANN, L. OKEDIJI, *When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale*, cit., 1369.

119 FILIPPOV, *Mapping Text and Data Mining in Academic and Research Communities in Europe*, cit., 20.

their corpus. At other times, vendors might require researchers to ask permission to conduct TDM on a case-by-case basis and this may implicate additional obstacles[120].

We will see later-on how policymakers, when drafting exceptions to copyright law, can limit their overridability by contractual clauses meant to restrict these activities.

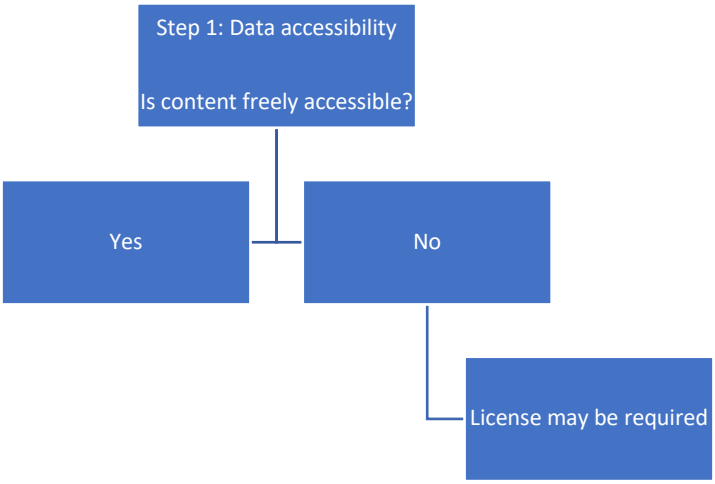


Figure 5 Issue regarding contract law[121]

Moving on to the main subjects of this work, namely copyright and related rights, we must consider that these rights come into consideration mainly in relationship with the previously mentioned TDM process phases of data collection, data preprocessing and/or extraction/copying phase. Along these phases a number of different activities need to be undertaken in TDM projects in order to prepare contents for the subsequent mining phase. Indeed, during a typical TDM process, as said before, the preprocessing phase may involve deeds of reproduction, extraction or copying of content that could infringe the rightsholders' prerogatives from copyright law and database law.

It is assumed that these acts fall within the area of the rights recognized to the beneficiaries of these intellectual property protection regimes. However, it must be said that not every TDM process necessarily involves these sensitive activities (e.g. Web Mining, where data mining techniques extract knowledge directly from web data)[122].

When carrying out these actions, copyright may come into play with respect to both individual works that are part of mined collections and to the collections as a whole. The latter can themselves be protected as a work under copyright law, regardless of any copyright in the contents of the collections. Sui generis database law, on the other hand, exists only with reference to the collections, regardless of any copyright in the contents or the collections themselves[123].

In that case content to be extracted and preprocessed could be subjected to different disciplines, depending on its inclusion or not into a database.

In particular, we can imagine two different scenarios:

1. Content to be extracted is actually included in a database.

[120] K. K. COURTNEY, R. SAMBERG, T. VOLLMER, *Big data gets big help: Law and policy literacies for text data mining*, in *College & Research Libraries News 81 (4) Scholarly Communication*, 2020. Available at: https://crln.acrl.org/index.php/crlnews/article/view/24383

[121] This figure is contained in ROSATI, *Copyright as an Obstacle or an Enabler?*, cit., 10.

[122] ROSATI, *Copyright as an Obstacle or an Enabler?*, cit., 10.

[123] CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit.,17.

In this case both copyright and database law might apply, protecting the content independently from both sides.

2. Content to be extracted is not included in a database.

Here copyright law may still be relevant.

Of course, in both the above-mentioned categories legal issues other than copyright and database law could arise, e.g. in the field of the already mentioned contract law and data protection law[124].

We will later-on deal with this subject in a more extensive manner, but here we anticipate and introduce some general concepts that will be further developed in the next chapters.

Oversimplifying, considering the content accessed (which could be or not included in a database), copyright law automatically grants protection over new creative works created or recorded by anyone in any manner. This protection gives the owner of these works exclusive rights, such as: the right to authorize, or to refuse to authorize, certain "*restricted acts*" by any third party. These acts include copying, adaptation of the work, re-dissemination, all, or a "*substantial part*" of the original copyright work, as well as translation of the work into other languages[125].

Among data to be mined in TDM processes, some are more likely to be protected by copyright law than others. These data, referred to as "*high level data"*, usually are "*rich*" data, e.g. newspapers or journal articles, books, music or photographs. On the other hand, *low level data*, that can be regarded as "*raw*" data, are less likely to be protected by copyright. These generally are measurement data, names, phone numbers and address data, financial data[126].

Moreover, if the content to be mined is included in a database, the latter could itself be protected under copyright law, if "*arranged in a systematic or methodical way and individually accessible by electronic or other means*"[127].

Sui generis database law, on the other hand, applies to those databases that show "*that there has been qualitatively and/or quantitatively a substantial investment*" in the obtaining, in the verification or presentation of their contents[128].

When TDM projects and activities are undertaken within the area of some of the exclusive rights granted by copyright or database law, one could choose to address the issue in 3 different ways, plus another one.

Firstly, the analyst can opt to get the "*authorization*" to use the content by the rightsholder, for example by obtaining a license enabling him to perform TDM on the protected content. Secondly, he can opt to abandon the project or to go on with it violating copyright law, assuming the risk of being sued by the rightsholders. The analyst, depending on the application domain in which TDM is applied and by the country/region in which it operates, could also, among other various variables, rely on policy-maker drafted exceptions or limitations to copyright and database law[129].

---

124 ROSATI, *Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and its Role in the Development of AI Creativity*, cit., 12.

125 M. BROOKS, P. MURRAY-RUST, C. OPPENHEIM, *The Social, Political and Legal Aspects of Text and Data Mining (TDM)*, D-Lib Magazine, Volume 20, Number 11/12, December 2014. Available at: http://www.dlib.org/dlib/november14/brook/11brook.html

126 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 19.

127 Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, Art. 3.

128 *Ibid*, Art. 7.

129 TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* cit., 87.

This macro-issue, thus copyright and sui generis database rights with their exceptions and limitations, including the newly drafted exceptions provided by Directive on Copyright in the Digital Single Market 790/2019 (CDSM Directive), will be thoroughly explained in the next chapters.

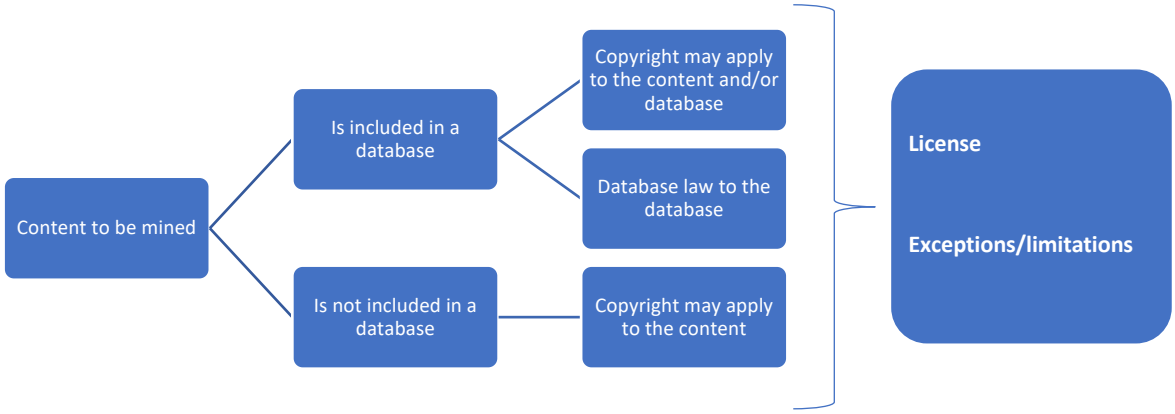


Figure 6 Copyright and Database law issues and solutions[130]

1.6.2. Privacy and Data protection and other legal issues relevant to data analysis

One of the most relevant legal issues is certainly that of the protection of personal data, and in order to make this work complete, a brief and concise discussion of the problem cannot be neglected.

Sometimes TDM projects involve the collection and usage of personal data.

The internet, through the last two decades, evolved from a medium of consumption into a medium of informatory participation, where users interactively use web services to access dynamic content or, instead of consuming, produce their own content for others to access, sharing not only opinions but also an incredible amount of personal information. This, given the increasing value that data have been given in the recent times, could eventually be used by individual or companies for purposes that the author/interested party never intended[131].

Especially among the consumers, whose personal features, habits and behaviors are the main target of these activities when directed toward commercial purposes, concern grows about the risks involved, such as:

1. Data loss or theft of identity (identity fraud);
2. Collection, sharing and usage of data in ways non previously foreseen or agreed;
3. Inappropriate usage of customers' data and subsequent risk of so-called nuisance contacts;
4. Discriminatory usage of personal data.

As a consequence, *privacy protection*, even if not at the center of this work, can undoubtedly be recognized as one of the main legal issues highlighted by the exploitation of personal data, as it is the case in most of the TDM applications.

---

[130] This figure is contained in ROSATI, *Copyright as an Obstacle or an Enabler?*, cit., 13.

[131] B. BREMERT, *Legal aspects of text mining publicly available data*, 1. Available at: https://www.datenschutzzentrum.de/uploads/projekte/itesa/Bremert-Legal_aspects_of_text_mining_publicly_available_data-paper.pdf

Although concern about personal data protection has become a crucial topic among the public just recently, after the recent scandals about misusage involving the exploitation of American citizens' personal data for unethical purposes other than the purposes for which they were collected, the history of privacy protection dates back in time.

The borders of privacy protection have greatly developed and changed through the time in direct relationship with the technological progress.

The famous article *The Right to Privacy* written by the Bostonian lawyers Samuel Warren and Louis Brandeis, published in 1890 in the Harvard Law Review[132], is generally recognized as the theoretic foundation of the modern idea of privacy. This article summarizes the increasing need of protection, perceived in the first place among the common law systems, depending on the invention and diffusion of portable cameras and the subsequent risk for the people portrayed, especially when belonging to elites, to lose control of their photographs.

With specific reference to Italy, the same need was initially perceived just after the WWII, in the 1950', for the same reasons of the US, when the rise of the first mass media raised the concern among famous people of a possibility of intrusion in their intimate sphere. After some famous jurisprudential pronouncements denying the existence of such a right and the consequent possibility of its protection, finally in 1975[133], the Supreme Court ruled that the legal system recognizes and protects everyone's interest in not disclosing confidential facts or events without his or her consent. This judgment thus claimed that the disclosure of images or events not directly relevant to public opinion constitute an infringement of privacy, and this right was anchored to the many rules in which emerged the will of the legislator to ensure the personal and family privacy and Article 2 of the Constitution and in the recognition of the inviolable rights of the person[134].

The initial concept of privacy is therefore, in the thought of the first commentators, mainly a protection against intrusion by the State or by third parties into the individual existential sphere, so that the right to privacy comes into consideration basically as a "*right to be let alone"*.

The concept of privacy develops toward a more "*informational*" type of protection with the birth and evolution of information technology. In fact, in the 70's computers were still few, voluminous and cost too much, remaining therefore almost an exclusive prerogative of the governments. During the Cold War, the western countries began to collect an increasing amount of substantial information, gathered into massive databases, about citizens habits and political preferences, so that the threat to privacy is now represented by the possible government control and the risks of discrimination associated with their collection of sensitive data[135]. In this period the first national regulations started to spread, as for example in the Nordic countries, already in the 70's, and to develop (in Italy only in 1996[136] and then in 2003[137]).

The informational side of protection in the field of privacy grows more and more with the development and the wider use of computers, that at this point got cheaper. This marks the shift from a conception of privacy as a "*right to be left alone"* to a "*right to retain control over*

132 S. WARREN, L. BRANDEIS, *The Right to Privacy*, 4 *Harv. L. Rev.* 193 (1890).
133 Cass. Civ., 27 maggio 1975, n. 2129, in *Foro Italiano* 1976, I, C.2895.
134 G. PASCUZZI, *Il diritto dell'era digitale*, 5th edition, Il Mulino, Bologna, 2020, 78.
135 *Ibid.*, 79.
136 Legge n. 675 del 31 dicembre 1996 "Tutela delle persone e di altri soggetti rispetto al trattamento dei dati personali".
137 Decreto Legislativo, n. 196 del 30 giugno 2003"Codice in materia di protezione dei dati personali".

one's own information"[138], and culminates with the advent of the World Wide Web (WWW) in its 2.0 version, where people are allowed to carry out on line daily activities such as communications, professional performances, purchases of goods, relocated from the physical to the virtual environment and thus producing immense quantities of personal data for others to access, an incredible amount of personal information the protection of which has become a challenging problem for policy makers.

These new technologies have created a further problem, arising from the aterritorial characters of these new technologies, given the difficulty of giving effectiveness to this new protection in an international context, which has prompted the legislators to actions of harmonization of the protection on the processing of personal data, resulting, within the European Union in Directive 95/46/EC, and, most recently, in the *General Data Protection Regulation* (GDPR)[139].

1.6.1.1. Data protection in Europe

In order to deal with the issue in the European context, one should start from the most recent *EU regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data*, and repealing Directive 95/46/EC (GDPR) UE 2016/679, which came into force in 2016[140].

This regulation has certainly brought some positive features to the EU Data protection rules, harmonizing a former fragmented discipline which, being a directive, was subject to interpretation and implementation of the single Member States.

A thorough regulation is now provided about natural people personal data treatment, giving to the interested parties additional rights about their own data, imposing for example a previous consent rule when their data are collected by third parties, the possibility to obtain their data cancellation whenever they want and free revocation of previous consents to their treatment.

This Regulation applies not only to European businesses, but to "*the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not*" if an "*offering of goods or services*" or "*the monitoring*" of the behavior of a customer takes place within the Union[141].

This discipline, or at least a part of it, could thus establish a new barrier and further obstacles to the usage of data mining techniques in the field of personal data.

To understand how much this new regulation can weigh on TDM exploitation we must, in the first place, pause on the meaning that the directive gives to the terms "*personal data*" and "*processing*".

According to article 4 of the GDPR, "*personal data*" means any information relating to an identified or identifiable natural person ("*data subject*"). An "*identifiable natural person*" is one who can be identified, directly or indirectly, in particular by reference to an identifier such as

---

138 S. RODOTÀ, *Repertorio di fine secolo*, Laterza, 1999, 201. According to which: "A definition of privacy as the right to be left alone, as mere confidentiality, has long since lost its general meaning. […] In the information society, functional definitions of the protection of privacy tend to prevail, which, in many ways, refer to the possibility of a subject to know, control, direct, interrupt the flow of information concerning him. Privacy, therefore, can primarily, and more precisely, be defined as the right to maintain control over one's information".

139 PASCUZZI, *Il diritto dell'era digitale,* cit., 80.

140 Regulation (Eu) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

141 Article 3 GDPR "Territorial scope".

a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. "*Processing*", on the other hand, means any operation or set of operations which is performed on personal data or on sets of personal data, by automated or not automated means, such as: collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction[142].

From these definitions one could derive that the discipline imposed by the EU Regulation will be definitely applicable to TDM processing on personal data.

Three main factors that can be discovered within the GDPR regulation, which are the most eligible to interfere with TDM processing are the following:

1. Principles of processing,
2. Legal grounds,
3. Informatory obbligations.

Let's therefore briefly see how and in what extent these factors can limitate TDM.

Starting from the principles of the directive, art. 5 of the GDPR states that personal data shall be:

(a) processed lawfully, fairly and in a transparent manner in relation to the data subject ("*lawfulness, fairness and transparency*");

(b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ("*purpose limitation*");

(c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ("*data minimization*");

(d) accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ("*accuracy*");

(e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organizational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ("*storage limitation*")

(f) processed in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures ("*integrity and confidentiality*").[143]

From these principles, the one that could be an actual obstacle in a TDM process is the (b), "*purpose limitation*", that limits the data processing to the purposes for which data were originally collected. This principle could particularly affect those processes in which data to be mined are not directly collected by the so called "*data controller*" or "*data processor*", but are, on the contrary, obtained from third parties. In this situation, these subjects must take

[142] Article 4 GDPR "Definitions".

[143] Art. 5 GDPR "Principles relating to processing of personal data".

responsibility for evaluating the purposes for which those data were originally collected, moreover assessing if their TDM project purpose is compatible or not with those original purposes. If yes, as we will later see, they should inform the data subjects of their identity and the purposes of the processing, which can be burdersome when dealing with large amount of data subjects[144].

Continuing with the second factor which could affect TDM process, any processing of personal data requires a legal ground in order to be considered legitimate. Several legal grounds that can legitimate and make lawful a processing of personal data activity are listed in article 6 of the GDPR, of which at least one should apply:

> (a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;
> (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;
> (c) processing is necessary for compliance with a legal obligation to which the controller is subject;
> (d) processing is necessary in order to protect the vital interests of the data subject or of another natural person;
> (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
> (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child[145].

With regard to this factor, it is impossible not to notice how the "*consent*" deriving from the individual data subject is probably the legal ground that has the most power to make legitimate a TDM activity conducted with the use of personal data, but at the same time is the one that is most difficult to obtain. This difficulty arises first of all from the conditions set out in Article 7, which are necessary for consent to be considered validly given. Moreover, Article 8 provides for additional conditions in case of consent arising from children. Finally, Article 9 foresees, in addition to the general prohibition to process "*special categories of personal data*", among others, the possibility to derogate from this prohibition by giving explicit consent to the processing for one or more specified purposes of this type of data[146].

On the other hand, one should moreover say that the GDPR has significantly expanded the list of possible legal grounds for conducting personal data processing, given that the previous directive, by not including the legitimate interests, performance of a contract, a legal obligation, the public interest or the vital interests of the data subjects, made this type of activity practically impossible, burdening those who intended to carry out such activity with the need to seek the consent of the subjects to whom those personal data belonged. And it has done the same for the category of data considered "*sensitive*", introducing more exceptions to the general prohibition of processing this type of data[147].

The final factor to consider is the disclosure requirements introduced by the legislation in Articles 13 and 14, which require the controller to disclose a range of information,

144 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 57.
145 Art. 6 GDPR "Lawfulness of processing".
146 Art. 9 GDPR "Processing of special categories of personal data": "Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited".
147 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe,* cit., 57.

including the obligation to provide the data subject at the time when personal data are obtained, with all of the following information: his identity, the purposes of the processing of data, the recipients and categories of personal data, and the existence of the data subject's right of access and right to rectify his data. This obligation could result in a significant effort for the controller of the processing, as in most of the TDM processes there could be lots of data subjects to which give this information.

1.6.1.2. Data protection in the United States

Only a few references will be made, as will be made regarding copyright and TDM, to the American context with specific regard to data protection.

In the US, the right to privacy has known a complicated development throughout the decades and nowadays it is the result of a combination, typical in the common law legal systems, between judge-made law and statutes issued both at federal and national level.

The legal protection of privacy in the USA jurisdiction of the courts develops within the framework of tort law, not becoming, anyway, an important subject of academic speculation until 1960, when W. Prosser published the article "*Privacy*"[148], where the author tries to classify the multitude of ways in which the right of privacy had been declined by courts in the previous time, distinguishing a number of different kinds of torts (among which "*public disclosure of embarrassing private facts about the plaintiff*"), having in common the fact of being injuries to the right to be let alone.

The extent of privacy protection, during the following years, has been widened by the Supreme Court in front of the challenge given by the growth in exploitation of personal data collected and treated within digital environments, under the principles set by the first and fourth amendments of the American Constitution, protecting freedom of speech and right against unreasonable search and seizure[149].

Informational privacy in the United States, as opposed to the European Union, in the absence of an omnibus privacy statute, is also the subject of a number of federal laws that, as a whole, can be seen as thick net of regulations, each one regarding specific contexts of information use[150]. Lawmakers preferred to legislate each sector separately than to give a comprehensive privacy law [151]. Examples of this sectoral data protection legislation are the "*Health Insurance Portability and Accounting Act*", which regulates the collection of health information, or the "*Gramm Leach Bliley Act*", regulating personal information collected by banks and financial institutions[152]. If this policy grants specific answers to the issues raised by the continuous technological discoveries and changing practices, it has been underlined that the sectoral statutory framework, in its rigidity, lacks promptness in adaptation, so that the new behavior can fall within the gap of the statutory regulation for months or years before being restricted[153].

Regarding the regulations enacted at the national level, worth of a citation is the "*California Consumer Privacy Act*" (CCPA), that recently came into force in 2020, which is

---

148 W. L. PROSSER, *Privacy*, 48 Calif. L. Rev. 383 (1960), 383-423.

149 In the well-known case *Katz v. United States*, 389 U.S. 347 (1967) the Supreme Court proclaimed that the fourth amendment protection covers people (that is to say citizens) and not just the premises or places falling under their property, extending therefore privacy protection to a series of new "*reasonable privacy expectations*".

150 P. SCHWARTZ, *The Value of Privacy Federalism,* in *Social Dimension of Privacy: Interdisciplinary Perspectives,* Cambridge University Press, 2015, 326.

151 J. S. COOK, L. L. COOK, *Social, Ethical and Legal Issues of Data Mining*, in *Data Mining: Opportunities and Challenges,* Idea Group Publishing, 2002, 408.

152 HIPAA (Public Law 104-191), GLBA (15 USC § 6802 et seq.).

153 L. J. STRAHILEVITZ, *Toward a Positive Theory of Privacy Law*, 126 Harv. L. Rev. 2010 (2013), 2036.

similar to GDPR in many respects, and for this reason is considered to be the most comprehensive state data privacy legislation at the present time.

CHAPTER 2

COPYRIGHT AND DATABASE RIGHTS AND THEIR RELATIONSHIP WITH TECHNOLOGICAL INNOVATION

2.1 Copyright and its relationship with technological innovation. Copyright wars

We realized in the introduction to the first chapter how our lives are nowadays affected by the current digital technological revolution. This revolution is bringing with it violent clashes and wars, and copyright is for sure currently one of the most relevant theaters of war of this revolution. At the two ends of this war there are, on the one hand, those who strongly claim that the exclusive rights should be strengthened and the time and scope limits that characterize it since its birth even erased, in order to control the current immensely more accessible reproduction capacity and massive uses of intellectual works, and, on the other hand, those who claim that the immense potential brought by this great revolution, just mentioned in Chapter 1 in a list that is by no means intends to be exhaustive, should not be bound by exclusive rights and control, hoping therefore the cancellation of the same[1], or at least their attenuation for the good of the collective interest.

This war is however nothing new. The contemporary debate about copyright does not lay on exclusively contemporary issues, but rather reflects conflicts that have never been fully solved and settled in the past, and thus have reverberated for several centuries throughout copyright's long history[2].

When we will face and analyze, within the limits allowed by the purpose of this work, the history of copyright, we will get aware that "*copyright wars*" have cyclically occurred in the past, for four hundred years, all the times corresponding with technological leaps of historical importance.

In fact, copyright law was born at the dawn of a technological revolution such as the invention of printing in Europe and the resulting mass dissemination of information. Technological change, along with other important factors, led to deep social, economic and cultural changes. In more recent times, other technological breakthroughs, such as the invention of the photocopier, and more recently the advent of digital technology (especially the internet) have been the catalyst for equally profound societal development[3].

Even the current war, in fact, stimulated by the social, economic and technological changes due to the advent of the second phase of the digital era and by the growing possibility of exploiting big data, has reignited the contrast between those who would like to see growing exclusive rights and those who would like to limit them, in the so called "*public interest in free circulation of information*". This latter interest, however, is not new in the copyright arena. Since its inception, copyright law was meant to seek a compromise between conflicting interests: the interest in protecting works and the interest in allowing freer and more liberal access to works[4].

---

[1] R. CASO, *Alle origini del copyright e del droit d'auteur: spunti in chiave di diritto e tecnologia*, in Trento Law and Technology Research Group, 2010, 6. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2254259

[2] T. K. ARMSTRONG, *Two Comparative Perspectives on Copyright's Past and Future in the Digital Age*, 15 *J. Marshall Rev. Intell. Prop. L.* 698 (2015-2016), 701.

[3] O. FISCHMAN AFORI, *The Evolution of Copyright Law and Inductive Speculations as to Its Future*, 19 *J. Intell. Prop. L.* 231 (2012), 233.

[4] *Ibid.*, 241.

If what we have just said is true, the reasons why TDM's activities have been considered, since their origins, to be pacifically included within the scope of those rights that copyright grants to authors of original works, are to be found in the evolution that copyright has had a long time since its birth, and in the several "*victories*" obtained in the different "*battles*" that followed one another during this wider and centuries-long war on copyright.

Hence, after having briefly introduced the basic concepts that nowadays constitute the copyright discipline of almost any legal system in the world, we will analyze, starting from its birth, the evolution of copyright. We hope that a historical overview could allow us to better understand how we have got to the current legislation, which has been widely criticized in the past and is currently criticized too, within the European Union, in spite of a revision that, apparently, has not changed very much its main features.

The most challenged and criticized fundamentals in this regard, are the basic representation of copyright as a property right and, on the other hand, the breadth of its exceptions and limitations, which are giving rise to the idea that we are currently in the midst of the "*copyright wars*"[5].

In the light of the current technological revolution, which has resulted in a real utopia in terms of information accessibility, we will therefore ask ourselves, later in the third chapter, whether the time has come to rethink the foundations of copyright in order to adapt the law to the contemporary needs[6].

2.2 Copyright: an overview

We will consider in the present paragraph the rights conferred by copyright and, at the same time, how they are necessarily the result of a balancing of divergent interests, resulting therefore in the fact that the rights it confers on the creator of a given work are almost always followed by some form of limitation of the same.

In fact, it is usually observed that the legal monopoly of rights over intellectual work is always limited by the presence of free privileged uses which "*create breathing space for socially valuable ends*"[7] and that therefore the two opposing terms together shape the overall copyright system, which should always be a form of balancing.

We will then start introducing the general characteristics of copyright, common to most of the legal systems, focusing specifically on copyright as it is regulated in the European Union and in the United States.

We will then briefly point out the different theories that have developed in civil law and common law systems in order to offer a justification for the rights/duties offered by copyright, an important discourse to understand the divergent solutions adopted by the European Union and several non-European countries in regulating TDM.

Finally, we will see how copyright regulation is based on a difficult compromise and balancing of divergent interests, facing then exceptions and limitations to the rights offered by copyright, focusing in particular on the two different approaches used in their regulation, their main justifications, their respective pros and cons.

2.2.1. Essential concepts of copyright law

[5] *Ibid.,* 233.

[6] *Ibid.*

[7] M.R.F. SENFTLEBEN, *Copyright, Limitations and the Three-step test. An analysis of the Three-step test in International and EC Copyright Law,* Kluwer Law International, The Hague, 2004, 5.

In a relatively short period of time, the world economy, especially that of the more developed countries of the world such as the United States and the European Union, has rapidly evolved from a purely industrial economy to an economy based on information and services. In the post-industrial era we are currently living, characterized by extremely rapid technological changes, the ability to reproduce and receive information is growing exponentially, and is accompanied by a consequent growth and prosperity of the communication and information industries[8], industries that have heavily based their operation and value on the intensive use of intellectual property rights, such as patents, trademarks, industrial designs and copyrights.

To quantify that growth and prosperity, these industries generated 45% of GDP (EUR 6.6 trillion) in the EU annually in 2019 and account for 63 million jobs (29% of all jobs)[9], with similar numbers for the United States, where the IP-intensive industries contributed with $6.6 trillion to the U.S. economy in 2014, up from $5.1 trillion in 2010, an increase of 29.4%. The share of total U.S. GDP attributable to IP-intensive industries also increased to 38.2% in 2014, up from 34.8% in 2010, supporting a total of 45.5 million jobs (27.9 million jobs directly and 17.6 million more jobs through the supply chain), which is about 30% of all employment in the U.S.[10]. With specific reference to copyright, the primary focus of this discussion, in 2019, the value added by the core copyright industries to U.S. GDP reached more than $1.5 trillion dollars ($1,587.16 billion), accounting for 7.41% of the U.S. economy, employing 5.7 million workers in 2019, accounting for 3.79% of the entire U.S. workforce, and 4.46% of total private employment in the U.S.[11]. Also here, similar numbers come from the EU, where the copyright intensive industry is estimated to contribute to the EU GDP for 6.9%, and to employ 7% of the total EU employment[12].

A consequence of the increased value of communicative expression is, of course, also the increased importance and value of the legal structure that governs the rules of its ownership[13]. In fact, products of the mind, informational products, are protected by several areas of intellectual property law. These include patent law, which protects inventions and certain types of discoveries; trademark law, which protects the words and symbols that identify to consumers the goods and services manufactured or provided by particular individuals or firms; trade-secret laws, which protect commercially valuable information that companies attempt to obtain from their competitors; and finally, copyright, which protects various and differing "*original forms of expression*". The economic and cultural importance of this set of rules is therefore rapidly increasing, and the fortunes of many businesses now heavily depend on intellectual property rights, leading to an increased interest from the legal professions, who are more and more frequently specializing in the profitable field of

[8] M. A. LEAFFER, *Understanding Copyright Law*, Fifth Edition, LEXISNEXIS, 2010, 2.

[9] *Intellectual property rights intensive industries and economic performance in the European Union*, Industry-Level Analysis Report, September 2019, Third edition. A joint project between the European Patent Office and the European Union Intellectual Property Office. Available at: https://www.epo.org/service-support/publications.html?pubid=201#tab3. For a summary of the key findings of the report see: "Intellectual property rights strongly benefit the European economy, EPO-EUIPO study finds", available at: https://www.epo.org/news-events/news/2019/20190925.html.

[10] J. ANTONIPILLAI, M. K. LEE, *Report: Intellectual Property and the U.S. Economy, Industries in Focus, 2016 Update*, report co-authored by the US Economics & Statistics Administration and the United States Patent and Trademark Office. Available at: https://www.uspto.gov/learning-and-resources/ip-motion/intellectual-property-and-us-economy.

[11] R. STONER, J. DUTRA, *Copyright Industries in the U.S. Economy, The 2020 Report*, prepared for the International Intellectual Property Alliance (IIPA), 2020, 4. Available at: https://www.iipa.org/files/uploads/2020/12/2020-IIPA-Report-FINAL-web.pdf.

[12] *Intellectual property rights intensive industries and economic performance in the European Union*, cit.

[13] LEAFFER, *Understanding Copyright Law*, cit., 2.

intellectual property disputes, from legislators around the world, who are constantly busily revisiting their intellectual property laws in order to build a competitive environment and to attract the most profitable businesses, and from academics, who, as an obvious consequence, have dramatically regained interest in the field in recent years[14].

Having made this due preliminary observation about the leading role and growing importance of intellectual property rights in general, and copyright in particular, in the modern world, let us now briefly describe copyright in its essential features.

Copyright is a statutory type of ownership that grants specific exclusive rights to the authors of a work, as well as companies that make products that contain the work or control its dissemination. These rights are largely economic in nature, allowing authors and producers to control the process of copyrighted material production and distribution, but are in most of the cases supplemented with moral rights. The latter are personal rights that are not normally assignable, benefitting just the author, so that they usually expire when the author dies or the copyright expires[15].

Then, the rights that copyright traditionally assigns are both economic and moral.

Economic rights are those types of rights that give rights owners the opportunity to derive financial reward from the use of their works by others. Most copyright laws state that authors or other rights holders have the right to authorize or prevent certain acts in connection with a work, specifically by authorizing or prohibiting: the reproduction of the work in various forms, such as printed publications or sound recordings; the distribution of copies of the work; the public performance of the work; the broadcast or other communication of the work to the public; the translation of the work into other languages; and the adaptation of the work, such as turning a novel into a screenplay[16].

On the other hand, especially in civil law countries, which, as we shall see, traditionally see copyright as linked to the author's personality and whose protection takes the form of a property right deriving from the act of creation, there are moral rights, which allow authors and creators to undertake certain actions to preserve and protect their link with their work, through the recognition of the right of attribution, i.e., the right to be recognized as the author of a work, the right of integrity, i.e., the right to prevent alterations of the work that are harmful to its honor or reputation[17].

With particular reference to economic rights, the author/creator may be their owner or those rights may be transferred to other (one or more) copyright owners. As for moral rights, many countries do not allow the transfer of them, as it is the case with France, where not only the inalienability of the moral rights is guaranteed to authors residing in France, but also to exploitations in France of foreign authors[18]. In most common law countries, on the other hand, moral rights are freely alienable, though the law may require the contract to specify which moral rights are transferred for what purposes[19][20].

Copyright essentially protects creative works originating from the human mind. The level of creativity required for a work to receive protection is usually very low, and it usually doesn't require any artistic merit, civic virtue, or commercial value. For example, in the EU

---

[14] W. FISHER, *Theories of Intellectual Property*, 1, in S. MUNZER, *New Essays in the Legal and Political Theory of Property,* Cambridge University Press, 2001.

[15] B. ATKINSON, B. FITZGERALD, *A Short History of Copyright: The Genie of Information*, Springer International Publishing, Berlin, 2014, 3.

[16] Understanding Copyright and Related Rights, WIPO, 2016, 9.

[17] *Ibidem.*

[18] See France, Code of intellectual property (n29), Art L 121-1, cl 2.

[19] See Visual Artists Rights Act of 1990 (VARA) and 17 USC § 106A.

[20] J. C. GINSBURG, *Overview of Copyright Law*, in R. DREYFUSS, J. PILA (eds.), *Oxford Handbook of Intellectual Property Law*, Oxford University Press, 2018, 14.

the protection of the work arises with the demonstration that the work is the "*author's own intellectual creation*"[21] while in the USA a demonstration of "*at least some minimal degree of creativity*"[22] is needed. Even if legislations differ a lot, it is possible to identify a trend: some jurisdictions give just a general description, for example "*works of the mind*", while others offer long lists of the types of works eligible for protection[23].

However, the economic and moral rights assigned by copyright law are not absolute, being limited in various ways, both intrinsically, through the construction of the system of rights granted by copyright itself, and extrinsically, through the introduction within the copyright system of limitations and exceptions to otherwise existing rights, in particular situations deserving protection.

Example of the first type of limitations is the requisite of fixation, which, in some countries, excludes some works from protection if they are not fixed in a tangible form[24]. Another is copyright duration. Copyright protection, indeed, does not continue indefinitely, but for a limited (although usually long) period of time, during which the rights of the copyright owner exist and may be exploited[25]. The copyright term begins with the mere creation of the work, because international copyright law, having embraced the natural law justification for copyright, no longer requires the work to be published, or to comply with notice, registration, or renewal of registration formalities[26] for copyright to attach or persist. The Berne Convention stipulates a minimum protection period of fifty years *postmortem auctoris*. Many nations, like the United States and European Union member states, have chosen for a lengthier term of 70 years *postmortem auctoris*[27].

Another key limitation to copyright, intrinsic in the protection given to authors, is the basic principle of copyright that "*ideas in themselves are not protected by copyright*" but rather copyright tends to protect "*the original expression of ideas*". This principle seeks to prevent monopolization of ideas by the authors of works, and to ensure their free circulation in society[28]. We will see later on how this principle, however, in today's digitized world, has been distorted and is no longer fully obeyed. The process of renewal and adaptation of copyright to the innovations of the last century has indeed led in some countries to the protection of data, such as those contained in databases, under certain conditions.

The second type of limitations, the extrinsic ones, is provided by the so-called "*exceptions*" or "*limitations*" to copyright. We will devote a whole paragraph to a general discussion about exceptions and their basic characteristics, therefore we refer to that paragraph for a more in-depth discussion of this topic. Instead, the next two chapters will deal respectively with the exceptions offered by the European Union and other non-European countries to copyright law for text and data mining activities.

2.2.2. Copyright traditions: author's rights and copyright. Copyright justifications: natural law theory and the utilitarian theory

[21] See Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society [2001] OJ L167/10; or Case C-5/08 Infopaq International A/S v Danske Dagblades Forening [2009] ECR I-6569.

[22] See Feist Publ'ns Inc v Rural Tel Serv Co 499 US 340, 345 (1991).

[23] GINSBURG, *Overview of Copyright Law,* cit., 4.

[24] Understanding Copyright and Related Rights, WIPO, 2016, 15. Available at: https://www.wipo.int/edocs/pubdocs/en/wipo_pub_895_2016.pdf

[25] *Ibid*., 19.

[26] Regarding the registration requirement, copyright differs significantly from the other types of IP rights.

[27] GINSBURG, *Overview of Copyright Law,* cit.

[28] M. CASPERS, L. GUIBAULT, *A right to "read" for machines: assessing a black-box analysis exception for data mining*, in Proceedings of the Association for Information Science and Technology 53(1), 2016, 1. Available at: https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/pra2.2016.14505301017

To fully understand how copyright works and why it is so differently regulated around the world, it is important to take a quick look at how philosophies and social movements have influenced its birth and development over time. This is obviously both a historical and a comparative exercise and it will be of great help in order to understand why certain legislative and jurisprudential choices were made when, later on, we will address the different approaches among different countries in the regulation of Text and Data Mining.

In order to identify these different approaches to the issue with regard to the development of copyright, it is necessary to focus mainly on its history in the United Kingdom, the United States and various civil law countries, especially France[29].

The history of copyright is somewhat comparable to the history of information, and for most of human history and everywhere in the world, state authority has always sought to control the dissemination of information to the population. The emergence of copyright law has represented a shift in social norms in human history, from a prone acceptance of hierarchy and authority to the assertion of individual and private interests, with the demand for ownership over things that are produced[30].

In fact, two approaches have been used throughout history by various communities to regulate information: the obligation-based approach and the entitlement-based approach. Societies based on the first approach are characterized by the individual's acceptance of his or her social status, who tends to fulfill his or her duty and obligation to society within the limited context of his or her class. Examples of such types of societies are the early Chinese and Indian societies. Societies based on the second approach are instead characterized by the revulsion for the idea of a fixed social order and by the belief of the individual's freedom to acquire material wealth through work, which is accompanied by the exclusion of others from this wealth[31].

Within the latter type of society, since the 18th century and the earliest copyright legislation, two different traditions about the ultimate goal of copyright protection and thus the role of authorship, more often referred to as "*justifications of copyright*", have been adduced: on the one hand, the first tradition, which considers the primary goal of copyright to be the protection of the creation and intellectual elaboration of the creator of the work; on the other hand, the second tradition, which views copyright primarily as a means of promoting and advancing knowledge, learning, culture, and art[32].

The first tradition, commonly called "*author's rights*", characteristic of the countries of continental Europe belonging to the civil law tradition, is based on the assumption that property rights on intangible goods are "*a moral right to reap the fruits of his/her labor*", an idea based on the philosophy of natural law. The second one, called "*copyright*", typical of common law countries such as United States or the United Kingdom, is rather based on utilitarian considerations and sees copyright as a system of incentives aimed at increasing the production of works of authorship, and therefore benefiting the public welfare[33].

It is worth premising that the study of these theories, especially in the last historical period, in which criticism against the traditional copyright system has led to a renewed interest in the origins and intrinsic motivations for the recognition of copyright, is not a purely academic matter but can instead be practically helpful in determining the scope of

[29] ATKINSON, FITZGERALD, *A Short History of Copyright: The Genie of Information*, cit., 6.
[30] *Ibid., 7.*
[31] *Ibid.*
[32] J. J. HUA, *Toward A More Balanced Approach: Rethinking and Readjusting Copyright Systems in the Digital Network Era*, Springer-Verlag, Berlin Heidelberg, 2014, 53.
[33] LEAFFER, *Understanding Copyright Law*, cit., 18.

rights granted to right owners. In fact, in case of uncertainty in the normative element, lawmakers will tend to go back to the rationales behind specific provisions to interpret, apply or modify them. Whether a particular copyright regime is based on utilitarian principles or natural law principles can indeed provide a useful indication of where the balance between the conflicting interests of rights holders and the public lies[34].

As we will see when we will deal with the history of copyright in more detail, the relationship and closeness between these two different visions has varied considerably over the last three centuries. In the eighteenth century they were more or less assimilable, since they shared most of the points of view; then they followed different paths in the nineteenth century, dividing themselves even more in the first half of the last century, when, shortly after the Second World War, they reached their maximum clash. The gap that had been created suddenly closed in the last two decades of the twentieth century, thanks to the more recent international treaties and to the incredible change of front of the United States (which from being a strenuous opponent of the treaties at the beginning of the last century, became, around the turn of the century, one of their greatest promoters), only to reappear again with the advent of the digital revolution at the beginning of this century[35].

The first justification, traditionally linked to continental European civil law countries, holds that human beings who create intellectual works should be able to enjoy a specific right that would include protection of their moral and economic interests and cover all uses of their works[36]. This right of legal monopolization of intellectual works is mainly rooted in natural law, in an ideal vision that tends to establish a strong link between the author and his work. According to this theory, the law would therefore be called upon just to "*formally recognize what is established by the very nature of things*". The figure of the author has thereby a central role in the protection of copyright, according to this theory[37], in line with the process of accentuating individualism that took place between the Renaissance and the Enlightenment, culminated in the French Revolution of 1789[38].

The purely individualistic approach to copyright protection, a characteristic of this theory, can be further divided into two different declinations: one based on a "*reward*" argument, which focuses primarily on the material interests of the author (e.g. exploitation rights), the other based instead on a "*personality rights*" argument, more focused on the intangible interests of the author (e.g. moral rights)[39].

The first declination is identified with the so-called "*Labour model*", whose main author of reference is John Locke, an English philosopher of the eighteenth century. He argued that people have a natural right of ownership over their bodies, a right of ownership that necessarily extends to the labor of their bodies and, by extension, to the fruits derived from such labor[40]. If such labor, indeed, takes place on resources that are either unowned or "*held in common*" he has a natural property right to the fruits of his efforts, and the state is duty-bound to respect and enforce that natural right[41].

---

[34] L. GUIBAULT, *Copyright Limitations and Contracts, An Analysis of the Contractual Overridability of Limitations on Copyright*, Kluwer Law International, The Hague, 2002, 7.
[35] ARMSTRONG, *Two Comparative Perspectives on Copyright's Past and Future in the Digital Age*, cit., 703.
[36] GUIBAULT, *Copyright Limitations and Contracts, An Analysis of the Contractual Overridability of Limitations on Copyright*, cit., 8.
[37] SENFTLEBEN, *Copyright, Limitations and the Three-step test. An analysis of the Three-step test in International and EC Copyright Law*, cit., 6.
[38] GUIBAULT, *Copyright Limitations and Contracts, An Analysis of the Contractual Overridability of Limitations on Copyright*, cit., 8.
[39] *Ibid.*
[40] LEAFFER, *Understanding Copyright Law*, cit., 19.
[41] W. FISHER, *Theories of Intellectual Property*, cit., 2.

In fact, he wrote:

"[*t]he 'labour' of his body and the "work" of his hands, we may say, are properly his. Whatsoever, then, he removes out of the stat that Nature hath provided and left it in, he hath mixed his labor with it and joined to it something tht is his own and thereby makes it his property. It being by him removed from the common state Nature placed it in, it hath by this labour something annexed to it that excludes the common right of other men. For this 'labour' being the unquestionable property of the labourer, no man but he can have a right to what that is once joined to, at least where there is enough and as good left in common for others'*"[42].

Although initially meant as ownership over tangible goods, this theory soon extended to ownership over intangible goods and thus to intellectual property, where the raw materials subject to labor (facts and concepts) seem to be somewhat "*held in common*" and labor seems to be a rather important contributor to the value of the finished products[43].

This theory had little influence in the campaign that led to the establishment of the first English copyright law of 1710. However, the emphasis it certainly placed on the nascent concept of "*authorship*" and "*authors' rights*" spread quickly and had some success in continental Europe, where countries such as France and Germany were, in the 18th century, searching for a justificatory theory to put as the basis for the recognition and assignment of new legal interests in European intellectual property law, due to the emerging commercial market for works of authorship[44].

The second declination of the natural law theory, alternative to the Labour Model, is the so-called "*Personality model*", which is based instead on the "*personality*" argument, associated with the German philosophers Kant and Hegel. This claims that private property rights are indispensable for the satisfaction of certain fundamental human needs[45] and thus can provide a useful means for self-actualization, personal expression, and individual dignity[46]. The role of the lawmaker would then be to create and allocate rights to resources in a way that best permits people to meet those needs. From this perspective, intellectual property rights can be justified either by the fact that they protect against appropriation or modification the artifacts through which authors and artists have expressed their "*wills*" (an activity that is considered central to personhood) or by the fact that they create social and economic conditions conducive to creative intellectual activity, which is important to human prosperity[47].

This second declination of the theory of natural law has, for example, been taken as a reference for the recognition of those personality-based rights called "*moral rights*" of integrity and attribution, which can be found in the Berne Convention in article 6bis[48].

The natural law theory, in both its forms, has been very successful, and it has certainly been used as a basic theory to justify the extension of the scope and intensity of copyright protection, for example in the subsequent revisions of the Berne Convention[49].

The second tradition, on the other hand, which has always been associated with common law countries, believes that the legal theory behind the copyright system serves an entirely different purpose, being it based not on natural law but on utilitarian considerations. It does not identify an intrinsic right of the author over his creations, but rather a promise

42 J. P. LOCKE, *Two Treatises on Government,* 1689.
43 FISHER, *Theories of Intellectual Property*, cit., 3.
44 LEAFFER, *Understanding Copyright Law*, cit., 19.
45 FISHER, *Theories of Intellectual Property*, cit., 3.
46 LEAFFER, *Understanding Copyright Law*, cit., 21.
47 FISHER, *Theories of Intellectual Property*, cit., 3-4.
48 LEAFFER, *Understanding Copyright Law*, cit., 21.
49 *Ibid.,* 22.

of economic rewards that is offered to the creators of literary or artistic works to encourage their productivity. Copyright is therefore conceived as a real engine of free expression, whose purpose is to encourage the dissemination of knowledge[50].

Therefore, according to the utilitarian theory, the legislator, in shaping property rights in general, should have as its objective the maximisation of net social welfare. With specific regard to intellectual property, this should be realized by striving to find that optimal balance between, on the one hand, the ability of exclusive rights to stimulate the creation of inventions and works of art and, on the other hand, the compensatory tendency of such rights to limit the general public enjoyment of such creations[51]. Finding this balance is quite a difficult task, because if the law provides too little incentive through reward, presumably people will stop creating. On the other hand, if the law provides too much reward or protection, presumably the public will stop paying and thus stop benefiting from the newly discovered or created work, and progress will decelerate. Essentially, copyright law should strike a balance between providing enough protection to encourage people to create, and not providing so much that it discourages those who would like to create new works from doing so[52].

This theory is thought to be based on the "*principle of utility*", or the ability of an action to satisfy individuals and more specifically its ability to satisfy as many individuals as possible, thus achieving "*the greatest good for the greatest number*". The English philosophers and economists of the late eighteenth and nineteenth centuries, Jeremy Bentham and John Stuart Mill, established that human behavior is limited to the extent to which the greatest "*pain*" is avoided, and the greatest "*pleasure*" is sought through action. The utility of any action, therefore, would depend on minimizing "*pain*" and maximizing the resulting "*pleasure*" in the largest possible group of people. The role of government is, consequently, to achieve utility through this simple "*pleasure-pain scale*" for any action, with the most favorable action giving the most pleasure to the greatest number of individuals possible. The way the legislature usually achieves the social good is by rewarding and punishing individual actions in order to induce society to follow a desired course of action[53].

In order to provide some concrete examples of legislation based on this theory, we must refer to the two most important common law systems: English law and American law.

As far as English law is concerned, its inspiration from the principles of utilitarian theory is already evident in the Statute of Anne of 1710, the objective of which is already stated in the title:

> *An Act for the Encouragement of Learning, by Vesting the Copies of Printed Books in the Authors, or Purchasers, of such Copies, during the Times therein mentioned.*

The same can be said for Article I, Section 8, Clause 8 of the United States Constitution, which authorizes Congress:

> *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*

---

[50] SENFTLEBEN, *Copyright, Limitations and the Three-step test. An analysis of the Three-step test in International and EC Copyright Law*, cit., 6-7.

[51] FISHER, *Theories of Intellectual Property*, cit., 1.

[52] D. KEMP, *Copyright on Steroids: In Search of an End to over Protection*, 41 McGeorge L. Rev. (2017), 799.

[53] GUIBAULT, *Copyright Limitations and Contracts, An Analysis of the Contractual Overridability of Limitations on Copyright*, cit., 10.

authorization that will take the form of the Copyright Act of 1790, whose title is very similar to the title of the Statute of Anne:

> *An Act for the encouragement of learning, by securing the copies of maps, Charts, and books, to the authors and proprietors of such copies, during the times therein mentioned.*

Not only the law but also American jurisprudence has repeatedly affirmed how this conception of copyright is the basis of American copyright law. For example, the Supreme Court of the United States in *Sony Corporation of America v. Universal City Studios*, Inc. 464 U.S.417 (1984), in the person of Justice Stevens says:

"*The monopoly privileges that Congress may authorize are neither unlimited nor primarily designed to provide a special private benefit. Rather, the limited grant is a means by which an important public purpose may be achieved. It is intended to motivate the creative activity of authors and inventors by the provision of a special reward, and to allow the public access to the products of their genius after the limited period of exclusive control has expired*"[54].

Although in the past, especially in the academic field, there has always been a tendency to delineate a clear division between the two different approaches to copyright, nonetheless it has been pointed out that there has never been a historical incompatibility between them, so much so that at the international level it has been possible to develop conventions, such as the Berne Convention, the TRIPS agreement and the WIPO treaties, inspired by both philosophies and implemented, at an early stage, by both common law and civil law nations[55].

Also from a historical point of view, the Statute of Anne (1709), which, as we have just said, is based on essentially utilitarian notions, matured in a climate strongly influenced by the philosophy of John Locke and his conception of a natural right to property which, by extension, was taken as the basis of the natural link between the author and his work, comparable to the concept of property. Locke's thought, moreover, influenced both the jurisprudence and the first U.S. laws on copyright prior to the constitution of the United States, which unequivocally referred to the principles of natural law[56].

Conversely, in civil law countries, for example in Germany, it has been pointed out that Locke's theories on the natural rights of authors have been used in a distorted way, aiming to protect the interests of booksellers. France's early enactments also reflected instrumentalist objectives, under the rhetoric of natural law, responding to a broader scheme of promoting public education as conceived by the legislators of the French Revolution. It can conclusively be said that utilitarian considerations in the early stage of copyright development were also present in civil law countries[57].

2.2.3. Balancing private and public interests: copyright limitations and exceptions. Open clauses and statutory exceptions

Copyright, as already mentioned, is never absolute, but is rather a complex legal situation in which a series of different interests meet. These interests are represented by

[54] Sony Corporation of America v. Universal City Studios, Inc. 464 U.S.417 (1984), 429.
[55] SENFTLEBEN, *Copyright, Limitations and the Three-step test. An analysis of the Three-step test in International and EC Copyright Law*, cit., 7.
[56] *Ibid.,* 8.
[57] *Ibid.,* 9.

different players, for example those of the authors/creators, those of the public and those of commercial operators (exploiters such as producers or distributors).

Balancing this intricate set of interests has never been easy, especially since the interests of these players vary considerably and conflict with each other.

Just to give an idea of how a complicated situation it can be, let's take a brief example: it can happen that the authors have an interest in benefitting from the fruits of their work, claiming payment for its exploitation. However, the authors themselves, at some point in the creation process, may have an interest in accessing existing works in order to create from them or use them as a source of inspiration, as in the scientific field, where access to existing works grants a higher level of professionalism to the whole work. At the same time, exploiters will want to recoup the investment they put in the production of the work and, at the same time, when they produce work that incorporates elements already protected by copyright, it will be in their interest not to be overly obstructed by existing monopolies. Finally, the public will want to be able to have easy and affordable access to works for entertainment and information purposes. At the same time, it will be in their interest too that the creators of such works were paid for their work, so that they have incentives to produce new and innovative works[58].

Copyright must then balance this intricate set of usually competing interests in order to promote both creative production and social progress through its dissemination. The law gives monopoly protection to the creator to inspire him or her to produce and to make the creations available to society, thereby contributing to the constitutional goal of social progress through promotion of science and the useful arts. An imbalance in the copyright law, either too much or too little protection, may result in reduced progress[59].

It is clear, then, that in copyright the basic conflict that permeates the discipline of the generally meant intellectual property, i.e., the tension between "*exclusion*" and "*access*" (of non-contractually bound third parties), is considerably more felt and complicated. This is effectively explained by Ghidini[60], who notes that in patent law, the power of exclusion is typically exercised against competing subjects, i.e. other inventors interested in accessing the protected material for the purpose of "*exploiting*" rather than for "*knowing*" them. The clash of competitive innovation is shown in the width of the patent scope, in the notion of equivalence, etc.; issues that relate to economic interests and whose constitutional reference is essentially the principle of economic initiative.

Even in copyright there are certainly purely economic conflicts, but what is different from patents is that in copyright the rights of exclusion can be exercised, and in fact are exercised in most cases, even against non-competitive "*social stakeholders*", representatives of various categories of users with varied interests, including subjects belonging to the world of entertainment, study, research, teaching, debate, who seek to produce, diffuse and receive information etc., and claim access to protected material "*for knowing*", not for economic exploitation. These interests can be linked to constitutional principles different from the principle of free economic initiative, which are, however, equally constitutionally protected.

---

[58] C. GEIGER, *The Future of Copyright in Europe: Striking a Fair Balance between Protection and Access to Information*, Report for the Committee on Culture, Science and Education – Parliamentary Assembly, Council of Europe, 3.

[59] KEMP, *Copyright on Steroids: In Search of an End to over Protection*, cit., 797.

[60] G. GHIDINI, *Exclusion and Access in Copyright Law: The unbalanced features of the European Directive "on Information Society"*, in *Rivista di Diritto Industriale*, 2013, 1-2.

It is in this very difficult balancing of important and difficult-to-sacrifice interests that the complexity of copyright is manifested, not only with respect to other areas of law, but also with respect to other types of intellectual property[61].

Since its inception, copyright has thus struggled to reconcile these competing interests, involved in political and judicial wars, and in doing so it has always used a structure of limits and restrictions. In an ideal copyright system, these limits and restrictions are essential balancing tools, calibrated to allow users of copyrighted works sufficient freedom to interact with those works without unduly undermining copyright's multiple rationales. While general copyright limitations define the subject matter, scope of protection, and duration of exclusive rights, statutory limitations (or "*limitations and exceptions*" as they are often called) are introduced into legal systems to accommodate a variety of cultural, social, informational, economic, and political issues and needs[62]. Along time, therefore, various types of limitations have been introduced into the legislation of individual countries, and it is quite normal, browsing through the legislations of various countries, to come across catalogs of specific provisions on copyright exceptions, such as those for quotation, private copying, libraries, educational institutions, incidental inclusion, freedom of panorama and so on[63].

As for all rulemaking activities, also in regulating copyright and its exceptions, the legislator has to perform a difficult activity of mediation and compromise between the search for the maximum legal certainty, favored by the presence of precisely defined provisions, able to offer predictability, and on the other hand fairness, easier to obtain through open and flexible legal concepts that allow a wide margin for ad hoc judicial adjustment and adaptation[64].

Over time, two types of approaches to regulating exceptions have developed around and from these two concepts.

The first approach is to provide a small number of generally worded exceptions. The second is to provide a larger number of specific exceptions, including carefully defined activities[65].

Although it must be said that there are no countries that rigidly adhere to one or the other approach, it is possible to identify a mere "*trend*" between the approaches of the various countries. In fact, despite the fact that international law making and harmonization have led, in recent decades, to a considerable convergence of the U.S. copyright with the continental *droit d'auteur*, the difference in approaches [66] is clearly discernible between these two traditions. In particular, the U.S. fair use approach is generally considered the key example of a generally worded exception. In Europe, on the contrary, long lists of specific exceptions prevail[67].

The difference in the approaches to regulate copyright limitations between these two traditions depends on the different theoretical substrate, which has been analyzed above. The

---

[61] For an overview of the tools used by jurists to perform the balancing of interests from time to time see GHIDINI, *Exclusion and Access in Copyright Law: The unbalanced features of the European Directive "on Information Society",* cit., 3.

[62] B. HUGENHOLTZ, M. R.F SENFTLEBEN, *Fair Use in Europe: In Search of Flexibilities*, in Amsterdam Law School Legal Studies Research Paper No. 2012-39 Institute for Information Law Research Paper No. 2012-33, 2011, 6.

[63] T. UENO, *The Flexible Copyright Exception for "Non-Enjoyment" Purposes – Recent Amendment in Japan and Its Implication*, in GRUR International, 70(2), 2021, 145.

[64] HUGENHOLTZ, SENFTLEBEN, *Fair Use in Europe: In Search of Flexibilities*, cit., 6.

[65] R. BURRELL, A. COLEMAN, *Copyright Exceptions: The Digital Impact*, Cambridge University Press, 2005, 4.

[66] For an Overview on the differences of the approaches to copyright law see: P. SAMUELSON, *Regulating Technology Through Copyright Law: A Comparative Perspective*, 42 European Intellectual Property Review (2020).

[67] C. GEIGER, F. SCHÖNHERR, *The Information Society Directive*, in I. STAMATOUDI, P. TORREMANS, EU Copyright Law, 11.64.

fair use approach can in fact be ascribed to the utilitarian foundations of the Anglo-American copyright tradition, which views copyright as a prerogative granted to authors to improve the overall welfare of society, in particular by ensuring a sufficient supply of knowledge and information. For the goal of this theory to be achieved, it is necessary to provide rights that are strong enough to lead to the desired production of intellectual works, while those works that do not need to be assigned to the right owner to provide the necessary incentive would remain free. Thus, exclusive rights are precisely delineated, while limitations to those rights can be flexibly regulated, through open standards; hyper-simplifying this theoretical model: freedom of use is the rule, whereas rights are the exception[68].

The opposite approach, i.e., the one that considers rights the rule and freedom the exception, stems instead from the natural law substratum in which the *droit d'auteur* of continental Europe has developed, for which the author occupies a central position and his/her intellectual work is seen as the materialization of the author's personality. If these are the premises, the author receives protection by the mere act of creation, and nothing should be left to the law, other than the simple recognition of what is already "*in the nature of things*". This necessarily leads the legislator to safeguard rights in a broad way, so as to grant the author the opportunity to profit from the use of his expression, and to exclude factors that may interfere with this exploitation. Therefore, the copyright system belonging to the civil law tradition, recognizes wide and flexible exclusive rights, while the exceptions are often rather narrow and interpreted restrictively [69].

Of course, for both approaches, scholars have identified both disadvantages and advantages. Proponents of the fair-use approach used in the U.S. framework argue that only this approach can provide the flexibility that copyright needs, especially in times of rapid technological and social change[70] and innovation. Only an open-textured standard would enable courts to accommodate new and beneficial uses of protected works and balance, on a case-by-case basis, the interests of authors against the many other interests deserving protection brought about by technological innovation, which could hardly be predicted by the legislature at the time it composed the exhaustive list of designated exceptions. It would furthermore enhance legal stability by making it unnecessary to constantly update the list of limitations[71].

Conversely, those who argue for an exhaustive catalog of exceptions believe that the fair use approach does not provide enough legal certainty. In fact, according to them, users of copyrighted works would need specific criteria to guide their behavior. This would only be easily achievable through a list of defined limitations, subject to strict interpretation, that would allow users to predict in advance which uses fall outside the rightsholder's sphere of control with greater certainty, and in case of legal disputes, would allow them to predict more effectively the possible outcomes of the conflict[72].

---

[68] M. SENFTLEBEN, *Bridging the Differences between Copyright's Legal Traditions – The Emerging EC Fair Use Doctrine*, in Journal of the Copyright Society of the U.S.A., 57(3), Spring 2010, 524.

[69] *Ibid.*

[70] See for example P. B. HUGENHOLTZ, W*hy the copyright directive is unimportant and possibly invalid,* in EIPR, 2000, 501. Critically commenting the specific list of exceptions provided by the EU InfoSoc Directive, states: "Of course, the whole idea of drawing up a finite set of limitations was ill-conceived in the first place. The last thing the information industry needs in these dynamic times are rigid rules that are cast in concrete for the years to come. How can a legislature in his right mind even contemplate an exhaustive list of limitations, many of which are drafted in inflexible, technology- specific language, when the Internet produces new business models and novel uses almost each day?".

[71] T. RENDAS, *Destereotiping the Copyright Wars: The "fair use v. closed list" debate in the EU*, 8. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2657482.

[72] *Ibid.*, 9.

The choice for a specific exceptions approach, therefore, typical of countries belonging to the civil law tradition, has been criticized by many for its alleged lack of flexibility and responsiveness in adapting to developments in the digital environment[73].

More and more representatives of the European doctrine have called for a reform of the system of exceptions in the European Union, arguing that the system as it is currently outlined obstructs the process of technological and scientific innovation in the European Union, thus running the risk of being left behind and giving way to other countries whose copyright system is more flexible, first of all the USA. Since the legislator was unable to foresee the technological innovations that would develop in the future with the necessary precision, new uses not covered by the existing exceptions would inevitably fall within the exclusive rights attributed to copyright owners, making them, if not duly authorized, infringing[74].

This is what to a certain extent has happened with Text and Data Mining. In fact, we will see at the beginning of the third chapter how the exceptions resulting from the (until not many months ago) prevailing InfoSoc Directive, were not sufficiently suitable to accommodate such activities within them, and how this resulted in the pressing need, manifested since the middle of the last decade, to reform the catalog of exceptions, leading, not in a short time and without considerable discussions, to the inclusion of the two new exceptions in Articles 3 and 4 in the Copyright in the Digital Single Market Directive (CDSM Directive).

We will analyze in the fourth chapter in more detail what these broad clauses consist of and under what conditions they may apply, focusing rather in the next chapter mainly on the European approach and in particular on that embodied in the exceptions arising from the last two European directives, the InfoSoc Directive and.

2.3. Copyright in historical perspective: the origins, its evolution and more recent developments

"*The roots of the present lie, of course, in the past, the knowledge of which often provides the keys to understanding a possible future*"[75].

The last decade has seen an increasing interest in historical research on intellectual property[76]. This is justified by the fact that we have begun to understand how the historical narratives may help us face the new challenges with greater confidence and awareness.

Indeed, the history of copyright law teaches us that there have always been a wide range of motivations inherent in copyright law, some of which relate to the personal and economic interests of authors, to the economic interests of entrepreneurs, to the social and economic interests of the public, and to the social and economic political interest of government. However, over the years the evolutionary dynamic has created a trend in which

---

[73] See, for example: L. M.C.R. GUIBAULT, *Why Cherry-Picking Never Leads to Harmonisation: The Case of the Limitations on Copyright under Directive 2001/29/EC,* in 1 (2010) JIPITEC 55, para 1. Available at: https://www.jipitec.eu/issues/jipitec-1-2-2010/2603; M. SENFTLEBEN, *The International Three-Step Test: A Model Provision for EC Fair Use Legislation*, in 1 (2010) JIPITEC 67, para. 1. Available at: https://www.jipitec.eu/issues/jipitec-1-2-2010/2605.

[74] RENDAS, *Destereotiping the Copyright Wars: The "fair use v. closed list" debate in the EU* cit., 10.

[75] G. SPEDICATO, *Principi di diritto d'autore*, Il Mulino, Bologna, 2020, 15.

[76] This is demonstrated by the many initiatives born in Europe aimed at the reconstruction and dissemination of primary resources in copyright. See, for example, "Copyright History" -Primary Sources of Copyright, CREATe, UK Copyright and Creative Economy Centre,
University of Glasgow. Available at: http://www.copyrighthistory.org/cam/index.php

the public interest, from being perhaps the primary objective of copyright regulation, has been progressively weakened while the economic interests of the entrepreneurial copyright industry have rather acquired supremacy[77].

Before proceeding to analyze in chapter three the contemporary rights granted by the European copyright legislation and the main issue of this work, namely the complicated relationship between these rights and the innovative techniques of TDM, new disruptive product of the technological revolution that has challenged the copyright system, it is therefore necessary to address the question of how copyright has evolved and how we have actually reached the current situation in which copyright has come to cover and hinder an activity (TDM) that, according to many scholars, at first glance does not seem to be considered infringing on the basis of the principles on which copyright was originally established. Studying the past of copyright can therefore prove to be extremely valuable in suggesting alternative solutions to the new challenges that have recently emerged, with a greater awareness of what is about to be regulated.

Since the birth of the first copyright laws, such as the aforementioned English one of 1710 or the American one of 1790, copyright has in fact witnessed several and rather profound changes, due to the growing emergence of new technologies, the main driver, to date, of copyright reforms. The same has happened recently, in the last of the technological leaps, the one that began at the end of the last century, which in Europe led to adoption of the InfoSoc Directive and continued at the beginning of this century with the advent of big data, which was in part the reason for reform of the previous directive, just mentioned, by Directive 790/2019, both of which are the subject of this treatise.

In fact, several authors claim, having analyzed its history, that there has been an uncontrolled expansion of the boundaries of copyright, resulting in the current crisis of the copyright system[78]. Initially and in the course of various early stages of its evolution, the copyright system had succeeded in the difficult balancing act between the protection of the interests of right holders on the one hand and the interests of the general public on the other, without submitting to the lobbying of the cultural industry, which invoked an absolute-property paradigm. However, with the advent of the internet and digital technologies, copyright seems to have failed in the difficult task of adapting to digital reality[79].

In the remainder of this chapter, we will therefore try to understand which were the most influential events that led to the situation in which copyright finds itself today, by analyzing its subsequent developments, especially at the level of international law.

We will come to affirm that up to the middle of the 19th century, copyright was, in my view correctly, perceived as a legal concept based on a multiplicity of foundations and objectives, albeit always aimed at establishing a compromise formula that did not endorse any particular interest or ideology, being the public interest in access to works variously combined with the conflicting economic interests of authors and entrepreneurs. However, in the second half of the 19th century, the balance between the competing interests began to change when the two main economic stakeholders of the time-authors and entrepreneurs-joined forces to establish an international intellectual property regime. This move was put into practice with the adoption of the first international copyright treaty, which still serves as

---

77 FISCHMAN AFORI, *The Evolution of Copyright Law and Inductive Speculations as to Its Future*, cit., 242.

78 For an overview of the expansion of copyright with specific reference to the US, deriving both from jurisprudential and legislative intervention, see KEMP, *Copyright on Steroids: In Search of an End to over Protection*, cit., 805 ss.

79 P. DE FILIPPI, K. GRACZ, *Resolving the Crisis of Copyright Law in the Digital Environment: Reforming the "Copy-Right" into a "Reuse-Right"*, in Proceedings of the 7th International Conference on the Interaction of Knowledge Rights, Data Protection and Communication, KnowRight 20, 2013, 2.
Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2318016

the foundation of intellectual property law today: the Berne Convention for the Protection of Literary and Artistic Works of 1886 (Berne Convention). This convention, the first international convention, together with the Paris Convention of 1883 on the protection of industrial property, brought about an epoch-making change in the development of intellectual property rights, allowing a new and different balance with respect to the traditional balance, in a way that ended up favoring private economic interests to a greater extent, to the detriment of public ones[80].

Having made this premise, which is meant to be a key to the understanding of what the rest of this chapter will deal with, let us now briefly see how copyright was born, what its initial form and characteristics were, and the interests it initially intended to protect.

2.3.1. The origins of copyright and the initial scope of protection: the privilege systems

It must be premised that copyright law, as it will be demonstrated below, is closely related to technological advances in the production, duplication and dissemination of the relevant works[81].

In fact, the origins of copyright, as a concept similar to the one we know today, are rather recent, if compared with other legal institutes such as property rights or non-contractual liability[82], dating back to the first copyright legislation in the United Kingdom, the Statute of Anne of 1710. However, these origins are, in turn, the ultimate result of the evolution of a huge technological revolution, namely the introduction of printing with movable types, in 1400, which led, thanks to the lowering of costs and time of reproduction of copies, to the replacement of the previous manual reproduction of writings, giving way to the creation of the economic preconditions, business models and legal structures on which to build the market for books[83].

Before then, just timid attempts of protection began to appear, however absolutely far from the current concept of copyright. The first literary works of ancient Greece were in fact mostly oral and anonymous. Moreover, the authors themselves preferred to consider their works more as collective results and belonging to the community, rather than personal works on which to claim their authorship in an exclusive manner. Such a trend will emerge only in the sixth century B.C.[84]. It is only with the ancient Hebrews that the need to record the prevailing oral culture, due to the introduction of new civil and religious laws, leads to the attempt to attribute the authorship of these rules and to assess in this way their authority and accuracy[85]. However, we are still very far from the modern concept of copyright.

Even in ancient Rome, up until the Middle Ages, intellectual creations were considered a collective and not an individual good. However, in this context, began the exchange of the first books and therefore the copying of writings for various types of remuneration. This occurred primarily under the control of the Roman Catholic Church, which had control over both the production and distribution of writings. Monasteries held the libraries and exchanged manuscripts for land, cattle or money. It is only with the disintegration of the

---

80 FISCHMAN AFORI, *The Evolution of Copyright Law and Inductive Speculations as to Its Future*, cit., 250.

81 T. EGER, M. SCHEUFEN, *The Past and The Future of Copyright Law: Technological Change and Beyond*, 5, in J. DE MOT, Liber Amicorum Boudewijin Bouckaert, die Keure, 2012.

82 SPEDICATO, *Principi di diritto d'autore*, cit., 15.

83 CASO, *Alle origini del copyright e del droit d'auteur: spunti in chiave di diritto e tecnologia*, cit., 7.

84 HUA, *Toward A More Balanced Approach: Rethinking and Readjusting Copyright Systems* in *the Digital Network Era,* cit., 48. Citing: A. HAUSER, *The social history of art,* New York, Kopft, 1952, 55.

85 R. V. BETTIG, *Copyrighting Culture: The Political Economy of Intellectual Property*, Taylor & Francis, New York, 1996, 12.

religious and cultural unity that took place in the Middle Ages that the concepts of authorship and intellectual property came to the forefront[86]. In fact, in the late Middle Ages in Europe will begin the secular trade of books, by the stationers, those individuals who, at the request of buyers, organized the reproduction of texts and provided copies of the same[87]. The advent of the printing technology is that sensational event that will lead to the birth, although obviously still in embryonic form, of the modern copyright. While the previous method of producing literary copies was characterized by few fixed costs and many variable costs, the new printing technology allowed for higher fixed costs and lower variable costs, leading to a decrease in the cost needed to produce one copy[88]. The stationers themselves therefore invested their previous earnings in the development of the new printing technology in order to be able to implement a mass production of copies.

Precursor of all this, which we will return to immediately, is what happened shortly before in Venice, where the Venetian Collegio gave John of Speyer, the first person to import the printing press into the Venetian territory, an exclusive printing privilege for a period of 5 years, a privilege that had the initial purpose of further encouraging the import and development of this technology. Not only that, the privilege system was implemented in such a way as to promote other social and economic objectives, such as a sufficient supply of books, good quality printing and the control of the market price of books. Venice had in fact become a territory where printing was an important source of income, and the unregulated printing of books had led to a surplus of the same, with the consequence that many printers had gone bankrupt and were subservient to their creditors. For these reasons, the first privilege system, born indeed in Venice, was founded to protect the printing industry and stabilize the market[89]. This model of printing privileges therefore constituted a form of protection more similar to patents for invention than to modern copyright, conferring on the individual booksellers and printers benefitted by the sovereign (therefore not yet on all authors of intellectual works) a monopoly on purely industrial and commercial activities, the printing and sale of books, in the context of a given territory[90]. This printing privilege was restricted to particular categories of books and did not confer the status of a proprietary asset that could be inherited, but rather were seen as non-transferable personal permissions. As the printing industry became more important over time, it became necessary to implement additional regulation to continue to fulfill the original purposes. Privileges were only granted for new books and were withdrawn if the book was not printed during the year following the granting of the privilege[91].

Moreover, the growth of the printing industry and its profitability was inevitably accompanied by free-riding counterfeiting behaviors. Copying activities of this nature were seen as a violation of public order and were therefore sanctioned with heavy fines to be paid to the authorities. Thus, from its inception, the regulation of the early printing industry reflected that attempt to create a legal order that took into account various dynamic and evolving interests, translating them into a balanced legal formula designed to promote public benefit while protecting private economic interests[92].

A further role of this system of privileges, which should not be underestimated, is that besides being a control on the printing machines, it became over time, also thanks to the

---

86 HUA, *Toward A More Balanced Approach: Rethinking and Readjusting Copyright Systems in the Digital Network Era*, cit., 48.
87 *Ibid.,* 49.
88 EGER, SCHEUFEN, *The Past and The Future of Copyright Law: Technological Change and Beyond*, cit., 6.
89 FISCHMAN AFORI, *The Evolution of Copyright Law and Inductive Speculations as to Its Future*, cit., 243.
90 SPEDICATO, *Principi di diritto d'autore*, cit., 16.
91 FISCHMAN AFORI, *The Evolution of Copyright Law and Inductive Speculations as to Its Future*, cit., 243-244.
92 *Ibid.,* 244.

pressure of the Church and the universities, a system of control on the same books produced by the machines. In fact, the state power realized in this period that printed matter could not only be extremely profitable, but also very dangerous. The privileges granted to individuals or guilds were therefore transformed into exceptional instruments of censorship at the service of the powerful in charge[93].

The privilege system will soon extend to other parts of Europe, such as Germany and the Netherlands, where the privilege granted by the state will be accompanied by non-interference agreements, aimed at protecting businesses from competition and piracy[94]. Finally, in 1504 the system of privileges also made its way to England, where the first royal privilege granted for the practice of the book trade was the creation of the title of "*King's Printer*", which was granted to William Facques, thereby giving him the exclusive right to print royal proclamations, statutes, and other official documents[95].

It was here that this system of privileges was centralised by powerful entities, which eventually led to the emergence of a private monopoly controlling all privileges. In 1557, in fact, in England, the guild of printers, known as the "*Stationers' Company*", was given the exclusive privilege of printing books. This privilege was exercised exclusively by members of the Stationers' Company who had obtained a permit to print a book from the Company. As time went on, the various members of the guild began to exchange these permits. The commercial nature of the stationery document had transformed its perception into a "*right*" to copy a book, rather than an "*administrative permit*". These rights were now transferable, and did not depend on the discretion or leniency of the Company of Stationers, and the holders of such entries considered themselves as having the right to copy books, i.e., possessing a "*copyright*". Towards the end of the seventeenth century, a group of four guild members managed to gain a monopoly in the book printing industry in England by purchasing all the entries, resulting in a significant increase in the price of books, accompanied by a decline in supply and quality[96].

The privilege system thus focused on regulating book printers, who were mostly entrepreneurs investing in book production. The author was not yet part of this system, having no protection or special status, since they usually sold their works for a lump sum or were forced to look for a patron. This situation changed, however, toward the end of the seventeenth century, when there was a dramatic surge in the demand for secular writing[97].

2.3.2. The first national copyright legislations as a properly made balance of interests

In fact, with the development of printing technology, the proliferation of the book trade began to take on significant proportions, also due to the advent of the Industrial Revolution, which, with the introduction of the steam engine and, later, of electricity, led to a mechanization of the printing process and to a further decrease in the cost of producing a book[98].

In 1700, cultural life in Europe suddenly underwent a dramatic and spectacular transformation. There was in fact in this period a sudden shift from intensive to extensive reading, as the rise of a new middle-class reading public led to an explosion of the print trade

[93] CASO, *Alle origini del copyright e del droit d'auteur: spunti in chiave di diritto e tecnologia*, cit., 9.
[94] HUA, *Toward A More Balanced Approach: Rethinking and Readjusting Copyright Systems in the Digital Network Era*, cit., 49.
[95] C. HESSE, *The rise of Intellectual Property, 700 B.C. – A.D. 2000: An Idea in the Balance*, Daedalus Vol. 131, No. 2, On Intellectual Property, Spring 2002, 30.
[96] FISCHMAN AFORI, *The Evolution of Copyright Law and Inductive Speculations as to Its Future*, cit., 244.
[97] *Ibid.,* 245.
[98] EGER, SCHEUFEN, *The Past and The Future of Copyright Law: Technological Change and Beyond*, cit., 8.

in the 18th century. In England, in fact, the annual production of books is estimated to have quadrupled during the 18th century. But France, too, saw a marked increase in literacy rates and a dramatic rise in demand for modern secular literature[99]. The increased productivity of printers and this new and increased demand for printed books were the two elements that finally led to the recognition of writing as a profession, and writers gradually gained recognition for both their social importance and their justified right to earn a living. This moment in history has been identified as the origin of the birth of the modern author, who worked independently and was motivated by his ambition to be able to express himself freely[100].

These two events put great strain on the old conception of authorship, which held that the publication system was based on the assumption that there was a definite amount of divine or ancient knowledge to be understood, conveyed, and interpreted. A growing number of men and women were enticed to become writers by the renewed interest in printed matter, particularly novellas, plays, and self-help manuals. However, these writers were different from their predecessors in that they were more concerned with the commercial potential of their work for contemporary readers than with eternal glory. Instead of selling a manuscript to a publisher, authors were increasingly hoping to sell the "*rights*" to a single edition. Secular authors began to assert that they were the makers of their works rather than merely the bearers of God's eternal truths with increasing frequency. They began to claim that their creations were their property, subject to legal protection and inheritable or sellable like any other type of property because they regarded themselves the authors of their works[101].

Not only that, but during this time period, a growing phenomenon of literary piracy emerged. Astute printers and booksellers all over Europe, sensing a new public demand for printed material and realizing an unmet market demand caused by artificial inflation of book prices due to publishers' perpetual privileges, began to produce cheap reprints across national borders or in small provincial towns on the best-selling and thus most profitable works, largely disregarding the existence of claims of exclusive perpetual privileges, justifying these actions in the light of the "*public interest*" and positioning themselves as adversaries to the dominant elements of the publishing industry[102].

Returning to the discussion of authors, one might logically assume that printers fought strenuously against the emergence of the modern author, as they were holder of a competing interest that would erode their profits. Printers, however, did not fight these claims, adopting a much more sophisticated strategy. The printers, taking advantage of the emerging concept of the modern author, immediately realized that requiring recognition of an author's ownership rights in his or her writings would actually promote their business. While the right of ownership might originally have rested quietly with the authors, it would inevitably and quickly be transferred to the printers, allowing them to increase their level of control over the market and boost their profits. By the end of the seventeenth century, the privilege system was thus bitterly criticized by all the parties: authors demanded a right of ownership over their writings, a growing number of provincial printers demanded the right to participate freely in the printing industry (which meant the abolition of the London guild), and the general public demanded a reduction in the price of books, an increase in their supply, and the creation of a market not governed by censorship. Some of these demands also found support in John Locke's aforementioned labor theory, which quickly made inroads throughout Europe. The theoretical justification for attributing ownership of books to their

[99] HESSE, *The rise of Intellectual Property, 700 B.C. – A.D. 2000: An Idea in the Balance*, cit., 31.
[100] FISCHMAN AFORI, *The Evolution of Copyright Law and Inductive Speculations as to Its Future*, cit., 243.
[101] HESSE, *The rise of Intellectual Property, 700 B.C. – A.D. 2000: An Idea in the Balance*, cit., 32.
[102] *Ibid.,* 32.

creators significantly strengthened the civil protest against the system of privilege. The combination of all these stresses culminated in England in the abolition of the privilege system in 1695. The stage was thus set for a new and modern copyright law, based on a new understanding of "*rights*"[103].

By the middle of the eighteenth century, not only in England, but also in France and Germany, the traditional system of publishing was in total disarray. Here, too, the same demands for reform of the regulation of the book trade were coming from all sides. Not only in fact, as we mentioned earlier, the emergence of a new audience of avid readers demanded cheaper books, but also the government was seeking to increase the book trade in such a way as to further encourage the more educated population in their kingdoms. Foreign and provincial publishers-especially in Switzerland and secondary French cities like Lyon-were loudly opposing the perpetual monopolies of the Paris book guilds on the most profitable books. Here, too, authors were clamoring for their property rights in their compositions to be recognized as absolute and perpetual. And privileged guild publishers, especially in Hamburg, Leipzig, Frankfurt am Main, and Paris, also hoped to see their traditional privileges recognized as perpetual property rights that could be defended against pirates in the courts[104].

It was in this context that, following the abolition of the privilege system, the first true national copyright legislation took shape in the United Kingdom, the Statute of Anne, of 1710.

The British, in fact, were the first to address the issue after the lapse of the *Licensing Act* in 1695, which had regulated the book trade and censorship. In seeking to end pre-publication censorship by removing the requirement to submit to prior licensing prior to publication, Parliament inadvertently called the entire system of privileges into question as well. If a work had not been registered prior to publication, no mechanism would have existed to protect literary privileges against pirated editions[105]. It is already in this context of initial reform that one can recognize the presence of that metaphor of war introduced at the beginning of the chapter, and which one will be able to identify in the copyright revisions of every part of the world in the centuries to come. In fact, already at that moment, the legislator had to deal with two opposing fronts: on the first, the London stationers, who were pushing for the privilege to be replaced by a formally perpetual right; on the second, the new figures of the provincial intermediary printers, who were pushing for a temporal limitation of the right[106].

The *Statute of Anne* definitively separated the issue of censorship from that of literary ownership. The statute stated that authors, and those who purchased a manuscript from an author, would have the exclusive right to publish the work for fourteen years. This right could be renewed for another fourteen years. But after this period (of fourteen or twenty-eight years), the work became part of the public domain, and anyone was free to publish it. As a result, all monopolies held by the Stationers' Company on classic texts were completely abolished[107]. However, at this point, the right in question concerned only books and only certain activities, mainly related to the economic exploitation of the same, such as the publication, reproduction and distribution for commercial purposes of the work[108].

This is where copyright was born, in a form quite similar to the one we know today: an exclusive right for the protection of intellectual works, potentially belonging originally to

---

[103] FISCHMAN AFORI, *The Evolution of Copyright Law and Inductive Speculations as to Its Future*, cit., 243.
[104] HESSE, *The rise of Intellectual Property, 700 B.C. – A.D. 2000: An Idea in the Balance*, cit., 33.
[105] *Ibid.,* 36.
[106] CASO, *Alle origini del copyright e del droit d'auteur: spunti in chiave di diritto e tecnologia*, cit., 13.
[107] HESSE, *The rise of Intellectual Property, 700 B.C. – A.D. 2000: An Idea in the Balance*, cit., 36.
[108] CASO, *Il diritto d'autore nell'era digitale*, in G. PASCUZZI, *Il diritto dell'era digitale*, cit., 197.

any author and transferable to third parties through a contract, whose heart lies in the exclusive right, a right that gives only its owner the power to economically exploit the work. In the traditional technological scenario, the author does not have the economic and entrepreneurial capabilities to commercialize his own work and therefore he turns to an intermediary, to whom he transmits by contract the exclusive right to the intangible asset (intellectual work), who in turn, through intermediaries (bookshops, for example) sells paper supports (tangible copies) of the work to readers[109].

Then came the turn of the United States, which in 1790 introduced the main aspects of the English Statute of Anne into their constitution, integrating them with the introduction of protection not limited to books but extended to maps and charts as well[110].

In the same years, in 1791 and 1793, also France enacted copyright laws, building, however, the rights on the author and not on the work and adding during the 19th century to the economic rights of exploitation of the work of the author's personal rights, moral rights, with unlimited duration[111].

By the end of the eighteenth century, therefore, several countries had decided to adopt copyright legislation[112]. These laws initially differed widely, both in terms of the works protected, the duration of protection, and the specific rights granted to authors. However, there was a sort of harmony among these different legislations, which were based on fairly shared premises: in all countries the legislators had tried to limit the privileges of publishers and rather give the rights over works to their authors; all believed that works were property motivated by natural rights arising from the work done by authors; everyone also agreed that authors were entitled to the benefits of copyright law only when they sold their works to publishers[113]; not only, the legislations of the time shared another important characteristic: not only did they introduce ownership rights in order to promote and stimulate literary productions, but they also tried to create a system of rapid and efficient transfer of works into the public domain, through the formulation of terms of duration of rights rather short compared to current standards, which were around 14 years in the USA and Britain and 5-10 years from the death of the author in France[114].

To summarize this first phase of evolution of the concept of copyright: it can be argued that up to this moment there was a broad consensus that copyright had to balance the interests between the intellectual property owner and the wider public good: authors and inventors could capitalize on their works and ideas, but only for a certain period of time. It seems no coincidence then that the term "*intellectual property*" should not, according to the Oxford English Dictionary, have appeared before 1845[115].

What we have just stated, however, is limited to the early part of the nineteenth century. As Hesse states, because modern laws governing intellectual property are based on a largely unexamined set of premises stemming from divergent philosophical views, these laws have proven particularly vulnerable to challenges; challenges largely driven by the continuing rise of new methods of distributing ideas and information across national borders. As a result, the philosophical tensions at the heart of modern concepts of intellectual property have been played out on an increasingly global scale, reframing the balance between private rights and public interest, often in new and dramatic ways[116].

---

109 *Ibid.,* 196.
110 EGER, SCHEUFEN, *The Past and The Future of Copyright Law: Technological Change and Beyond*, cit., 9.
111 *Ibid.*
112 We refer in particular, in chronological order, to: Britain, United States, France and Germany.
113 P. BALDWIN, *The Copyright Wars: Three Centuries of Trans—Atlantic Battle*, Princeton University Press, 2014, 53.
114 ARMSTRONG, *Two Comparative Perspectives on Copyright's Past and Future in the Digital Age*, cit., 708.
115 HESSE, *The rise of Intellectual Property, 700 B.C. – A.D. 2000: An Idea in the Balance*, cit., 40.
116 *Ibid.*

2.3.3. From national to international copyright law. The Berne Convention: the establishment of copyright as a property right and the initial enduring focus on balancing interests

Thus, copyright is now regulated by an intricate set of rules that originated at the national level, primarily in the 18th century, as we have seen, but which has been subject to attempts of harmonization over time, through various and diverse legislation. The industrial revolution was thus creating an international market for literary works and mechanical inventions, and the need was now developing for a new international regime of intellectual property rights[117].

European national copyright laws enacted in the nineteenth century generally focused on national authors and did not protect foreign authors or publishers[118]. The protection offered by national laws at the time was in fact limited in geographic scope, which meant that the law applied only to nationals of the country concerned, and this was becoming a problem because of the growing importance of international trade in the works of science and art[119]. There were special cases in which such laws protected foreign nationals, but always subjecting foreign nationals to conditions to qualify for protection. For example, German law protected the works of German citizens wherever they were published but protected the works of foreign citizens only if they were published by a German publisher in Germany[120].

Since the exploitation of works was no longer limited to the national territory, the idea arose that it was necessary to develop an international instrument to protect intellectual works outside the national borders where they were created. Until this moment, in fact, most European countries did not find particularly immoral or unfair the indiscriminate exploitation of works from other countries without prior permission, but this practice was rather seen positively, as an act that could lead to an advancement and distribution of knowledge among the local population[121]. But the lack of protection of foreign works or conditional protection became detrimental not only to foreign authors and publishers, but also to domestic authors and publishers. Indeed, as a result, more expensive domestic works began to be replaced with less expensive, unauthorized copies produced abroad. Similarly, unauthorized copies of locally produced foreign works not only led to reduced revenues for foreign authors and/or publishers, but also decreased revenues for domestic authors who then had to compete with cheaper foreign copies in other countries[122]. Inter-state piracy spread through Europe, and caused damage mainly to countries such as England, France and Germany, which soon discovered that the works produced by their citizens were reproduced at much lower prices in countries that spoke the same language, such as the United States, Belgium or other German-speaking countries, respectively. Thus, the interest of these nations in stopping interstate piracy rapidly led to the development of the first international copyright relations. The aforementioned profound differences between national laws, with the uncertainty arising from them, led to the emergence of the opinion that it was necessary to create an international copyright law, in order to make the substantive aspects of the various national laws more homogeneous and standardized[123].

---

117 *Ibid.*

118 R. GARZA BARBOSA, *Revisiting International Copyright Law*, 8 Barry L Rev 43 (2007), 45.

119 EGER, SCHEUFEN, *The Past and The Future of Copyright Law: Technological Change and Beyond*, cit., 10.

120 GARZA BARBOSA, *Revisiting International Copyright Law*, cit., 44.

121 FISCHMAN AFORI, *The Evolution of Copyright Law and Inductive Speculations as to Its Future*, cit., 250.

122 GARZA BARBOSA, *Revisiting International Copyright Law*, cit., 44.

123 FISCHMAN AFORI, *The Evolution of Copyright Law and Inductive Speculations as to Its Future*, cit., 250.

It is therefore in the context of international law that this harmonization was achieved and that major innovations in copyright law have taken place over time. Within the international copyright treaty movement, three different phases can be identified. The first phase, characteristic of the beginning of the nineteenth century, is that of bilateral agreements between states, based on the principle of reciprocity. The second phase is the development of multilateral copyright treaties. The third phase, the most recent one, is characterized by the birth of trade agreements that include substantial copyright provisions[124].

Leaving aside here the bilateral treaties between states because they are not pertinent, we will focus instead on the second and third phases, which have brought significant innovations to the initial copyright and have contributed to shaping it in the way it now appears.

Numerous international conventions and treaties have been stipulated on copyright, but the most important and relevant for this discussion are certainly:

- The Berne Convention[125], which established basic standards and minimum protection for works
- The TRIPS Agreement[126], which first associated copyright to international trade.
- The WIPO Internet Treaties[127], which have instead addressed the problems of copyright arising from digital technological innovations.

The first fundamental step towards worldwide recognition of copyright was certainly *The Berne Convention for the Protection of Literary and Artistic Works* (Berne Convention), 1886. This convention is considered to be the key agreement in copyright legislation, from which subsequent treaties were inspired. In particular, it is attributed with the merit of having responded to the just mentioned need expressed in a shared way by many countries, at that time the most culturally mature and developed, by progressively making global the previous bilateral agreements between nations developed to combat interstate piracy (for example the *International Copyright Act of the United Kingdom* in 1838, which served as the basis for the treaties with France and Germany)[128]. The Convention was initially signed by ten countries: Germany, Belgium, Spain, France, the United Kingdom, Haiti, Italy, Liberia, Switzerland, and Tunisia, but, in any case, the colonial empires of the member countries covered most of the globe[129].

These countries were at the time the major exporters of cultural and scientific production, counting for almost half of the book production industry. Obviously, on the other hand, importing states such as those belonging to the Latin American area or Russia, did not want to join the Convention. Among these importing states, mainly developing countries, there was still the United States, which, however, from a cultural importing country soon became the main exporter, surpassing in 1910 the British book production (also due to unauthorized reprints of British texts). As we have already seen, the US copyright system considered the right to reproduce and distribute a work as a right granted by the State and which required formal registration. This peculiarity of the U.S. system further prevented their adherence to the Convention for a long time, as the first revision conference of the Convention in 1908 abolished the requirement of central registration, previously necessary

124 GARZA BARBOSA, *Revisiting International Copyright Law*, cit., 45.
125 Berne Convention for the Protection of Literary and Artistic Works (1886). For a brief summary on the contents provided by the Convention see: https://www.wipo.int/treaties/en/ip/berne/summary_berne.html
126 The Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) (1994). Full text available at: https://www.wto.org/english/docs_e/legal_e/trips_e.htm#art1
127 WIPO Copyright Treaty (WCT) and WIPO Performances and Phonograms Treaty (WPPT) (1996). Full texts available at: https://wipolex.wipo.int/en/text/295166 and https://wipolex.wipo.int/en/text/295578
128 F. CANTATORE, *Authors, Copyright, and Publishing in the Digital Era*, Information Science Reference, 2014, 13.
129 GARZA BARBOSA, *Revisiting International Copyright Law*, cit., 47.

for a work to be considered protected, embracing entirely the European view, according to which legal protection derives from the author's "*act of creation*"[130].

The Convention is credited with having established the so-called "*principle of national treatment*", which, like the previous bilateral treaties, was based on the principle of reciprocity, and therefore led to a mechanism of mutual assurance in guaranteeing the same level of copyright protection offered to its own citizens to the citizens of the signatory countries[131], as well as establishing basic principles and a minimum standard of protection, thus profoundly influencing the treaties that came into force subsequently and therefore the national copyright laws of the signatories of these treaties[132].

For the purposes of this analysis, it is very important to underline how the Berne Convention played a crucial role as a catalyst in the initial phase of that process which shifted that initially well-practiced balancing of interests to the side of the private interest of the rightsholders to the expense of the public interest. Although, as we will soon see, this concept of balancing of interests was still clearly present and even affirmed in its first version of 1886, its subsequent revisions, inspired by the rapid technological development that took place throughout the last century, rapidly lost its original features.

The first key element, which initiated the uncontrolled expansion of private interests in copyright, is the property right character given to copyright by the Berne Convention. In fact, the Convention immediately embraced and was built around the continental justification of copyright, namely that of natural law, which saw copyright as an author's property, thus overriding any positive national law and effectively institutionalizing the concept of copyright as a property right worldwide. In fact, this concept naturally influenced the laws of the contracting countries, which, as signatories of the convention, had to comply with it[133].

This decision had its origin in the fact that the seed of the Berne Convention was sown, in 1878, by the birth in Paris of the "*Association littéraire international"* (ALI), the predecessor of today's "*Association littéraire et artistique internationale*" (ALAI). Its first president was the famous French author Victor Hugo, perhaps the most well-known supporter of the Romantic Movement, so closely associated with the foundation of natural rights of copyright. Indeed, the Romantics considered creative works to be extensions of their authors[134]. The ALAI created the preliminary guidelines that served as the basic principles for the enactment of the Berne Convention. Among these principles was, indeed, the recognition that copyright is a property right, which must be protected by law[135]. In the speech "*Discours d'ouverture du Congrès littéraire international*"[136] it was stated that the main purpose of the Association was:

"Vous allez faire une fondation, la propriété littéraire. Elle est dans le droit, vous allez l'introduire dans le code. Car, je l'affirme, il sera tenu compte de vos solutions et de vos conseils"[137].

---

130 EGER, SCHEUFEN, *The Past and The Future of Copyright Law: Technological Change and Beyond*, cit., 11.
131 *Ibid.,* 10.
132 HUA, *Toward a More Balanced Approach: Rethinking and Readjusting Copyright Systems in the Digital Network Era*, cit., 5.
133 FISCHMAN AFORI, *The Evolution of Copyright Law and Inductive Speculations as to Its Future*, cit., 251.
134 D. GERVAIS, *The 1909 Copyright Act in International Context*, 26 Santa Clara High Tech. L.J. 185 (2012), 187.
135 FISCHMAN AFORI, T*he Evolution of Copyright Law and Inductive Speculations as to Its Future,* cit., note 129, citing S. RICKETSON, J. GINSBURG, *International Copyright and Neighbouring Rights*, *The Berne Convention and Beyond*, Oxford University Press, 2006, 49-50.
136 Discours d'ouverture du Congrès littéraire international Victor Hugo 7 juin 1878. Available at: https://fr.wikisource.org/wiki/Discours_d%27ouverture_du_Congrès_littéraire_international
137 Discours d'ouverture du Congrès littéraire international Victor Hugo 7 juin 1878, § 85. Unofficial translation: "*You are going to make a foundation, literary property. It is in the law, you are going to introduce it into the code. Because, I affirm, your solutions and advice will be taken into account*".

In fact, a few years later, in Article 9 of the Berne Convention we find the following sentence:

> *"Authors of literary and artistic works protected by this Convention shall have the exclusive right of authorizing the reproduction of these works, in any manner or form"*[138].

In spite of this first element of the process that would lead to an increasing proprietization of copyright, the first version of the Berne Convention firmly stated that just as the principle that copyright is a property right had to be established at that time, so too the protection of copyright had to take into due consideration the public interest. In particular, Hugo states later in the same speech:

"Messieurs, rentrons dans le principe: le respect de la propriété. Constatons la propriété littéraire, mais, en même temps, fondons le domaine public. Allons plus loin. Agrandissons-le. Que la loi donne à tous les éditeurs le droit de publier tous les livres après la mort des auteurs, à la seule condition de payer aux héritiers directs une redevance très faible, qui ne dépasse en aucun cas cinq ou dix pour cent du bénéfice net. Ce système très simple, qui concilie la propriété incontestable de l'écrivain avec le droit non moins incontestable du domaine public […]"[139].

The following sentence,

"Le principe est double, ne l'oublions pas. Le livre, comme livre, appartient à l'auteur, mais comme pensée, il appartient — le mot n'est pas trop vaste — au genre humain. Toutes les intelligences y ont droit. Si l'un des deux droits, le droit de l'écrivain et le droit de l'esprit humain, devait être sacrifié, ce serait, certes, le droit de l'écrivain, car l'intérêt public est notre préoccupation unique"[140].

will establish what is still known as one of the key principles of copyright, and one of the main inherent limitations of copyright, namely the aforementioned principle of idea-expression dichotomy, according to which it is not the idea that is protected but the particular expression of it. The principle of exclusion of ideas from the scope of copyright protection is included in both the major legal systems and in the TRIPS agreement[141]. Just to cite a couple of examples: first of all, § 102(b) of the US Copyright Act excludes from the scope of protection ideas, procedures, processes, systems, methods of operation, concepts, principles, and discoveries. Another example can be the French copyright legislation, which provides that "*les iddes sont de libre parcours*" (ideas should circulate freely) is a fundamental principle of intellectual property protection[142].

---

[138] Art. 9 Berne Convention 1886.

[139] Discours d'ouverture du Congrès littéraire international Victor Hugo 7 juin 1878. § 86 – 87. Unofficial translation: "*Gentlemen, we must get back to the principle: the respect for property. We must recognise literary property, but at the same time we must establish the public domain. We must go further. We must expand it. Let the law give all publishers the right to publish all books after the death of the authors, on the sole condition that they pay the direct heirs a very low royalty, which in no case should exceed five or ten percent of the net profit. This very simple system, which reconciles the unquestionable ownership of the writer with the no less unquestionable right of the public domain […]".*

[140] Discours d'ouverture du Congrès littéraire international Victor Hugo 7 juin 1878. § 87. Unofficial translation: "*The principle is twofold, do not forget. The book, as a book, belongs to the author, but as a thought, it belongs - the word is not too broad - to the human race. All the intelligences have a right to it. If one of the two rights, the right of the writer and the right of the human mind, were to be sacrificed, it would certainly be the right of the writer, because the public interest is our only concern*".

[141] Article 9(2) of TRIPs agreement provides: "Copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such".

[142] GERVAIS, *The 1909 Copyright Act in International Context*, cit., 188.

But the public interest is not merely mentioned by Hugo. On the other hand, it still seems to be at the heart of copyright protection. This is particularly evident at the point in the speech where he argues that if a conflict should ever arise, such that it should be necessary to sacrifice one of the two elements, the element to sacrifice should surely be the individual right of the author[143]. This, therefore, confirming what we have said about the mutual contamination of the two traditions on copyright, was not at all distant from the dictates of utilitarian theories, which maintained that the protection of copyright should stop when it had maximized the public interest. The initial understanding of what Hugo thought the Berne Convention was supposed to be, therefore, not only included the public interest as an element of the framework, but this public interest was the very framework of the Convention. In short, the translation of the role of this public interest was to protect authors for the personal contribution they made to humanity and the development of human "*intelligence*", while at the same time therefore imposing limits on that protection when required by the public interest, i.e., when the public interest did not need to protect the rights of the author[144].

And the primary and original text of the Convention actually seemed to have achieved this objective, in particular through a rather minimalist normative content, if compared to the wording of its most recent version. The main objective of the Convention was indeed to ensure that authors from countries that were signatories to the Convention would be protected in the other countries of the Union without discrimination[145].

For this reason, the Convention, in its original formulation, in addition to providing for the principle of national treatment[146], provided for only a few of the rights currently present in its most recent version, in particular: translation[147] and public representation for dramatic and dramatico-musical works[148].

The fundamental and perhaps most well-known reproduction right was initially "*taken for granted*" by the convention, because it was fully incorporated into the convention only in the Stockholm Convention of 1967, although it was already present from the earliest time in national legislation (e.g., the Statute of Anne)[149]. All the signatory states to the Berne Convention provided in one way or another for a copying or reproduction right in their national laws, but at that stage they could not agree on the scope of the right to be included in the treaty. For this reason, the treaty did not initially provide for an explicit reproduction

---

143 Discours d'ouverture du Congrès littéraire international Victor Hugo 7 juin 1878: "Si l'un des deux droits, le droit de l'écrivain et le droit de l'esprit humain, devait être sacrifié, ce serait, certes, le droit de l'écrivain, car l'intérêt public est notre préoccupation unique". Unofficial translation: "If one of the two rights, the right of the writer and the right of the human spirit, were to be sacrificed, it would certainly be the right of the writer, for the public interest is our only concern".

144 D. J. GERVAIS, *Making Copyright Whole: A Principled Approach to Copyright Exceptions and Limitations*, 5 University of Ottawa Law & Technology Journal. 1 (2008), 5-6.

145 GERVAIS, Ma*king Copyright Whole: A Principled Approach to Copyright Exceptions and Limitations*, cit., 6.

146 Art. 2 Berne Convention 1886: "Authors belonging to any country of the Union, or their lawful representatives, shall enjoy in the other countries for their works, whether unpublished or published for the first time in one of those countries, the right which the respective laws do now or may hereafter grant to natives".

147 Art 5(1) Berne Convention 1886: "Authors belonging to any country of the Union, or their lawful representatives, shall enjoy in the other countries the exclusive right of making or authorizing translations of their works during the whole duration of the right in the original work. […].

148 Art 9(1) Berne Convention 1886: "The stipulations of Article 2 apply to the public performance of dramatic or dramatico-musical works, whether such works be published or not".

Art. 9(2) "Authors of dramatic or dramatico-musical works, or their lawful representatives, are, during the existence of their exclusive right of translation, in like manner protected against the unauthorized public representation of translations of their works".

149 GERVAIS, *Making Copyright Whole: A Principled Approach to Copyright Exceptions and Limitations,* cit., 6.

right[150]. However, it already contained an implicit reference to the right of reproduction[151], which however seemed to be still rather limited, referring to "*infringing copies*", which, as such were "*liable to seizure on importation*"[152]. It also contained an explicit, albeit conditional, reproduction right for newspapers and periodicals[153], which applied only if explicitly made clear by the author[154]. Not only that, but the original text also contained a useful partial definition of "*unlawful reproductions to which this Convention applies*", which included "*unauthorized indirect appropriations of a literary or artistic work, of various kinds, such as adaptations, musical arrangements, etc., when they are only the reproduction of a particular work, in the same form, or in another form, without essential alterations, additions, or abridgements, so as not to present the character of a new original work*"[155][156].

Therefore, apart from the initial insertion of these (still few) rights, the provisions of this first version of the Berne Convention mainly concerned and focused on national treatment, and at a first glance the Convention seemed to have remained rather loyal and to have taken well into consideration that concept of copyright as a balancing of divergent interests deriving from the initial legislations of national systems.

However, it is not only the preliminary works leading up to its birth and its basic structure that lead to this opinion. In fact, browsing through the provisions, one can glimpse exceptions that undeniably reflect considerations of public interest, in particular the right to information and the press[157]. This is the case, for example, of that exception, the only mandatory one, under which the reproduction of "*articles of political discussion, news of the day or miscellaneous facts*", could not be prohibited[158], and that, in this case however not mandatory, for "*use in publication for teaching or scientific purposes, or for chrestomathies*"[159].

2.3.4. The expansion of the scope and terms of copyright as a consequence of the emergence of new technologies: Berne Convention revisions, TRIPS Agreement and the WIPO Treaties

In the previous paragraph, we examined the Berne Convention, the first truly multilateral international treaty concerning copyright. We mentioned how it, while continuing to give due consideration to the balancing of interests of copyright, laid the groundwork for what has led to the current imbalance of interests, as it effectively went on to institutionalize, globalize, and over time consolidate the idea that copyright was a property

---

150 E. SCHWARTZ, *An Overview of the International Treatment of Exceptions*, PIJIP Research Paper no. 2014-02, American University Washington College of Law, Washington, D.C, 7.

151 Art. 7(1) Berne Convention 1886: "Serial novels, including short stories, published in the newspapers or magazines of any country of the Union may not be reproduced, in original or in translation, in the other countries, without the authorization of the authors or their lawful representatives".

152 Art. 12(1) Berne Convention 1886.

153 Art. 7(2) Berne Convention 1886: "This applies equally to other articles in newspapers or magazines, whenever the authors or publishers shall have expressly declared in the newspaper or magazine in which they have published such articles that they forbid the reproduction of these. For magazines it is sufficient if the prohibition is made in a general way at the beginning of each member".

154 GERVAIS, *The 1909 Copyright Act in International Context*, cit., 189.

155 Art. 10 Berne Convention 1886.

156 GERVAIS, *The 1909 Copyright Act in International Context*, cit., 189.

157 *Ibid.*, 190.

158 Art. 7(4) Berne Convention 1886: "No prohibition can in any case apply to articles of political discussion, news of the day, or miscellaneous items (notes and jottings)".

159 Art. 8 Berne Convention 1886: "As regards the liberty of lawfully making extracts from literary or artistic works for use in publications destined for education, or having a scientific character, or for chrestomathies, this matter is reserved to the law of the countries of the Union and to particular arrangements existing or to be concluded between them".

right, through the explicit choice of the justificatory copyright theory of natural law as the basis on which to establish international and national copyright law in much of the world.

A second element that has certainly led to the current imbalance of interests and that is therefore useful to analyze here is the continuous revision to which copyright has been subjected in the two centuries to come, due to the increasingly massive and frequent emergence of new technologies occurred from the late nineteenth century onwards, both through successive revisions of existing international conventions, and through the enactment of new treaties between states.

In fact, we have already pointed out several times in this chapter how copyright is closely linked to the process of technological evolution and how its very origins can be attributed to a major technological revolution. Since its birth, therefore, and in the course of the next three centuries, its existence and justification has been challenged and extremely controversial, thus placing the legislator in the difficult position of having to take into account the opposing needs expressed by the various stakeholders in the many occasions of its reform. In fact, the circumstance in which the emergence of a new technology of reproduction and/or transmission of information induces the legislator to intervene in order to protect the economic interests of the creators of original works is to be considered a real constant in the history of copyright. The most significant changes in the juridical discipline of intellectual works have in fact occurred, over the course of time, in response to the main innovations in the technological horizon of reference: this happened with the diffusion of new technologies such as radio, television or, on the other hand, the video-recorder, the latter being the object, in 1984, of the historical Betamax decision[160], in which the Supreme Court of the United States pointed out how "*from its beginning, the law of copyright has developed in response to significant changes in technology*"[161].

In discussing technological revolutions that have challenged the copyright system, Yoshiyuki Tamura effectively identifies three "*waves*" of different threats to the legitimate interests of copyright rightholders and subsequent adjustments and responses in reaction to them, which can be distinguished in relation to the type of technology and the actual use of copyrighted material[162].

The first is identified in the aforementioned spread of printing technology, which led to the emergence of the threat of piracy, with the consequent need to protect publishers against cheaper editions from provincial publishers, met through the creation of the modern copyright institution in England. That institution functioned quite well until the mid-20th century, mainly because of the fact that the cost of making a copy of copyrighted material at that time was quite prohibitive, requiring significant investment. The small group of people who could afford such investments at the time were certainly not consumers, but entities that made copies for commercial purposes. Consumers of copyrighted works could not afford to reproduce books, nor could they afford to record high-quality sound or images. The main function of copyright at that time was limited to protecting against competition for commercial reproduction of works by entities that did not have those rights. The copyright system functioned effectively precisely because of the limited number of actors whose compliance was easier to control[163].

Things changed radically when, between the late nineteenth and late twentieth centuries, the emergence of new communication and transmission technologies, such as

---

160 *Sony Corp. of America v. Universal City Studios, Inc.,* 464 U.S. 417 (1984).

161 SPEDICATO, *Principi di diritto d'autore*, cit., 17.

162 Y. TAMURA, *Rethinking Copyright Institution*, in World Intellectual Property Organization Journal (W.I.P.O.J.) No. 1, 2009, 67.

163 TAMURA, *Rethinking Copyright Institution*, cit., 67.

photography, sound recording, radio broadcasting, cinema, television and satellite imagery, exponentially increased the number of ways in which information could be transmitted, thus enriching the variety of works produced, and stimulating the emergence of a number of new copyright-related industries. Expressions of literary works were no longer simply confined to the printed text, but rather expanded to images, sound, and film. The rapid development of the industries that relied on these new types of works consequently saw a growing concern on the part of the legislature, which progressively tightened up the legislation, increasing the protection afforded by copyright, both through an inexorable expansion of subject matter matters and exclusive rights, and by increasing the length of the terms of protection[164].

Not only that, the same technological revolution that led to the birth of new technologies and new goods to be protected has in fact led to the emergence, on the other hand, of more advanced reproduction possibilities at much lower costs. This is identified by Tamura as the Second Wave of Copyright development, characterized by the massive use of analog reproduction technologies by individuals, not necessarily for commercial purposes. As new reproduction technologies entered the lives of consumers of copyrighted works, the contours of copyright radically changed, starting to regulate and intervene extensively in the activities of private individuals. This attempt to limit the increased freedom of consumers was not without difficulties, however, mainly due to the fact that the number of actors to be controlled had now increased considerably, thus jeopardizing the effectiveness of copyright[165]. Of particular concern, for example, was the introduction of the Xerox 914 photocopier in 1959, which made printed documents easily and quickly photocopiable, leading to a significant increase in the production of copies of copyrighted works, especially in schools, universities and administrations, thus putting a considerable strain on the exclusive rights of right holders to their own works[166]. But the same can be said about what happened with the introduction of tape recorders and video cassette recorders (VCRs) into the market around the same period.

This is, therefore, the historical period that marks the very beginning of the concrete process of strengthening of copyright, which will inevitably lead to the inclusion of activities such as text and data mining to be pacifically identifiable as activities in conflict with the rights conferred to right-holders.

So let us return again briefly to the topic of the last paragraph, the Berne Convention, which we left at its origins, and let us look at the novelties progressively brought to the international system of copyright law by the impulse of this Second Wave of technological innovations, and how these novelties have contributed to shift the already unstable balance of international and national copyright law towards the private interest.

The Berne Convention, even at the time of its birth, was designed with an eye to the future, knowing that new technologies would soon be born and that new works would have to be protected. It was therefore designed to be changed and adapted over time[167]. The result is that the Berne Convention has evolved through several versions during subsequent revisions[168]. In particular, new minimum rights were included in successive Revision Conferences, the first of which was held in Berlin in 1908, in part to follow the evolution of

---

164 HUA, *Toward a More Balanced Approach: Rethinking and Readjusting Copyright Systems in the Digital Network Era*, cit., 4.

165 TAMURA, *Rethinking Copyright Institution*, cit., 67.

166 EGER, SCHEUFEN, *The Past and The Future of Copyright Law: Technological Change and Beyond*, cit., 11.

167 See Art 17 Berne Convention 1886: [1] "The present Convention may be submitted to revision with a view to the introduction of amendments designed to improve the system of the Union". [2] "Questions of this kind, as well as those which in other respects are of interest to the Union, shall be considered in conferences to be held successively in the countries of the Union among the delegates of the said countries".

168 GARZA BARBOSA, *Revisiting International Copyright Law*, cit., 47.

the new forms of exploitation of work available after 1886. Among the various rights, some exceptions were also included, but as efforts were concentrated more on the extension of rights, the area of exceptions and limitations remained for a long time an area practically unregulated at international level[169]. For reasons of practicality and in order not to bore the reader too much with a long list of rights and exceptions, a schematic table containing the main changes to the first version of the Convention that have occurred in the seven subsequent revisions of the Convention over time will be inserted below.

Table 1 The evolution of rights, limitations and exceptions in the Berne Convention[170]

Revision or protocol (year)	**New rights (article)**	**New limitations or exceptions (article)**
Paris (1886)	• Extension of reproduction right to serial novels (must be asserted, IV) • Right of adaptation applied specifically to transformation of a novel into a theatrical play and viceversa	
Berlin (1908)	• Terms of protection of life + 50 years (7) • Broader translation right (8) • Removal of need to assert reproduction right in serial novels and short stories • New right of adaptation for mechanical reproduction and public performance using such reproductions (13) • Extension of right to obtain seizure to such adaptations (13(4)) • New right of reproduction and public performance by cinematography (14)	• Possible conditions and restrictions on mechanical reproduction right (13)

[169] GERVAIS, *The 1909 Copyright Act in International Context*, cit., 190.
[170] The present table is taken from: GERVAIS, *Making Copyright Whole: A Principled Approach to Copyright Exceptions and Limitations*, cit., 8-9.

Rome (1928)	• Moral right (6bis, 9(2), 11bis(2)) • New exclusive right of communication by broadcasting (11bis(1))	• Possible exclusion from protection of political speeches and speeches in legal proceedings (2bis(1)) • Possible limit on right of reproduction of lectures, addresses and sermons (2bis(2)) • Possible limit on the right of communication by broadcasting, including compulsory licenses (11bis(2))
Brussels (1948)	• Broader right of translation (8) • Broader moral right (in quotations, 10(3)) • Extension of public performance right to communications to the public of the performance (11(1)) • Extension of communication right to broadcasting or communication by any other means of wireless diffusion of signs, sounds and images; any communication to the public by wire (cable) or broadcasting; and public communication by loudspeaker (11bis(1)) • New right of public recitation (11ter) • Broader right of adaptation, arrangement and other alteration (elimination of reference to new original work as being excluded, 12) • Broader right in cinematographic adaptations (now includes distribution as well as public performance, 14) • New droit de suite (resale right, 14bis(1))	• Mandatory right of quotation (10(1)) • Possible exception to use of excerpts in educational and scientific publications (10(2)); replaces previous possibility of maintaining existing exceptions • Possible exception for the recording, reproduction and public communication of short extracts for the purpose of reporting current events (10bis) • Possible conditions (incl. compulsory license) on broader communication right (11bis(2)) • Possible exception for ephemeral recoding and official archiving (11bis(3)) • Possible limit on resale right (14bis(2))
Stockholm (1967)	• New/broader right of reproduction (all categories of works, 9(1))	• Three-step test (9(2))

	• News reporting reproduction exception may be excluded by rightsholder (10bis(1)) • Broader right of public performance and communication (reservations no longer mentioned, 11) • New right of public communication of a recitation; right extended to translations (11ter) • Right of performance for cinematographic works extended to communication by wire (14bis)	• Possible limits on protection of official texts (2(4)) • Right of quotation extended to all works but must be compatible with fair practice and the extent of use must be justified by purpose (10(1)) • Modification of educational exception, limited to "by way of illustration" and compatibility with "fair practice" but applied also to broadcasts and recordings (not just publications, 10(2)) • Newspaper/periodicals reproduction exception now applies to broadcasting and communication and to publications on "economic, political or religious topics" (10bis(1)) • Exception for reporting current events by photography, cinematography, broadcasting or communication to the public limited to "the extent justified by the information purpose" (10bis(2))
Paris (1971)		• New Appendix (providing developing countries with the possibility of issuing compulsory reproduction and translation licences, subject to a complex administrative machinery)

The table shows a significant expansion of rights and somewhat less attention paid to exceptions. New rights, in particular, have often been included because some works, especially theatrical, musical and film works, derive much of their commercial value from their public performance (live) or communication (distance). On the occasions when exceptions and limitations have been inserted alongside the introduction of new rights, they have often been in the form of unspecified possibilities offered to national legislators. In

some cases, they have been implemented through compulsory licensing systems, while when they have been included "*in principle*", they have mostly been made optional[171]. Therefore, while on the one hand the evolution of the rights provided by the Convention has followed a rather clear path, the same cannot be said of the exceptions, which have remained almost unregulated, and when they were actually included in the system, they were permissive and optional, as they did not provide for an obligation of implementation by member states[172].

The end result of the above trend in Berne Convention amendments culminated in the adoption of what governs the introduction of new exceptions and limitations to this day, the three-step test. It was introduced in 1967, when the Berne Convention was amended to introduce that initially non-existent explicit reproduction right. The introduction of the reproduction right was counterbalanced by the introduction of an exception in Article 9. The reproduction right provided for "*the exclusive right of authorizing the reproduction of [literary and artistic] works, in any manner or form*". On the same occasion, the first explicit exception to this exclusive right was inserted in article 9 (2), which did not apply at that time to the other exclusive rights (public performance, broadcasting, public recitation, adaptation, or the rights granted to producers of cinematographic works), which basically means that there were no exceptions granted by the Berne Convention, other than that provided for the right of reproduction[173].

This article provided that exceptions or limitations to exclusive rights had to meet three conditions, hence the name "*three-step test*". The three conditions are as follows: (1) the exception had to be limited to certain special cases; (2) it should not conflict with the normal exploitation of the work; and (3) it should not unreasonably prejudice the legitimate interests of the right holder. This "*three step test*" has thus standardized national limitations, which were previously characterized by considerable discrepancies, being very broad in some countries and narrow in others. It is important to bear in mind the role that the three-step test has played in conditioning the subsequent possibility of introducing new exceptions. In fact, it does not specify the minimum standard, but rather the maximum, thus not obliging the contracting countries to adopt limitations on copyright. Should they wish to adopt limitations, they must not exceed the threshold of the "*three-step test*"[174].

To date, looking at its most recent version of 1971[175], the Berne Convention appears to be very different from the concise initial version and considerably more extended than its origins. Those basic principles and minimum standards that the Convention intended to grant at the outset have been significantly expanded. The expansion has concerned not only the list of works and, consequently, the rights granted, but also the duration of protection. As regards the former, the Berne Convention, in Article 2, still provides for the protection of "*literary and artistic works*", in particular with regard to "*every production in the literary, scientific and artistic domain, whatever may be the mode or forms of its expression*"[176]. The exclusive rights

[171] Examples of the first case are article 11a(2) or article 13; examples of the second case the droit de suite.
[172] GERVAIS, *Making Copyright Whole: A Principled Approach to Copyright Exceptions and Limitations*, cit., 9.
[173] SCHWARTZ, *An Overview of the International Treatment of Exceptions*, cit., 9.
[174] FISCHMAN AFORI, *The Evolution of Copyright Law and Inductive Speculations as to Its Future*, cit., 251.
[175] Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979). Available at: https://wipolex.wipo.int/en/text/283698.
[176] Art. 2 Berne Convention 1971, which now provides as Protected Works: The expression "literary and artistic works" shall include every production in the literary, scientific and artistic domain, whatever may be the mode or form of its expression, such as books, pamphlets and other writings; lectures, addresses, sermons and other works of the same nature; dramatic or dramatico-musical works; choreographic works and entertainments in dumb show; musical compositions with or without words; cinematographic works to which are assimilated works expressed by a process analogous to cinematography; works of drawing, painting, architecture, sculpture, engraving and lithography; photographic works to which are assimilated works expressed by a process

protected are both moral rights and economic rights, including the right to translate[177]; make adaptations and arrangements to the work[178]; perform and recite the works in public[179]; communicate the work to the public through drama, musical, or cinematographic performance[180]; broadcast[181]; reproduce in any manner or form[182]; use the work as the basis for audiovisual works[183] and reproduce, distribute, and perform in public or communicate the audiovisual works to the public[184], as well as the right to claim authorship of the work and the right to object any mutilation or deformation or the modification of, or other derogatory action in relation to, the work to which would be prejudicial to the author's honor or reputation[185]. The term of protection of these exclusive rights covers a period of 50 years after the author's death[186][187]. However, the Berne Convention still did not provide at this point the most crucial of rights (and exceptions) in the digital age - the rights of distribution and/or communication to the public, including the making available right-, all of which will be considered, and added to, in the later treaties we are now going to address (notably WIPO's "*digital*" treaties in 1996)[188].

We are now in the nineties and the changes occurred in those years by the recent innovations introduced by the technological boom had made the previous international copyright law quite outdated, as the Berne Convention had not been revised since the 70s.

An important treaty following the Berne Convention is the *The Agreement on Trade Related Aspects of Intellectual Property Rights* (TRIPS Agreement)[189], the first international treaty to regulate copyright in relation to international trade. The year is 1994, and it is now that the Berne Convention is definitively adopted as the mandatory starting point for all national copyright laws[190]. The basic concept of copyright as a property right and the narrow "*three-step test*" as the upper threshold for the introduction of exceptions have, therefore, been further consolidated and established as "*inherent*" copyright concepts by the TRIPS Agreement. These concepts are currently at the heart of the copyright wars, as they promote the supremacy of private economic interests over public interests. Moreover, the restrictive "*three-step test*" for copyright limitations has been specifically reinforced by an ad hoc

analogous to photography; works of applied art; illustrations, maps, plans, sketches and three-dimensional works relative to geography, topography, architecture or science".

177 Art. 8 Berne Convention 1971: Right of Translation.

178 Art. 12 Berne Convention 1971: Right of Adaptation, Arrangement and Other Alteration.

179 Art. 11ter Berne Convention 1971: Certain Rights in Literary Works: Right of public recitation and of communication to the public of a recitation.

180 Art. 11 Berne Convention 1971: Certain Rights in Dramatic and Musical Works.

181 Art. 11bis Berne Convention 1971: Broadcasting and Related Rights.

182 Art. 9 Berne Convention 1971: Right of Reproduction.

183 Art. 14(2) Berne Convention 1971: Cinematographic and Related Rights: Adaptation of cinematographic productions.

184 Art. 14 (1) Berne Convention 1971: Cinematographic and Related Rights: Cinematographic adaptation and reproduction; distribution.

185 Art. 6bis Berne Convention 1971: Moral Rights.

186 Art. 7 Berne Convention 1971: Term of Protection.

187 HUA, *Toward a More Balanced Approach: Rethinking and Readjusting Copyright Systems in the Digital Network Era*, cit., 6.

188 SCHWARTZ, *An Overview of the International Treatment of Exceptions*, cit., 9.

189 Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) 1994. Available at: https://www.wto.org/english/docs_e/legal_e/27-trips_01_e.htm

190 Art. 9(1) TRIPS Agreement: Relation to the Berne Convention: "Members shall comply with Articles 1 through 21 of the Berne Convention (1971) and the Appendix thereto. However, Members shall not have rights or obligations under this Agreement in respect of the rights conferred under Article 6*bis* of that Convention or of the rights derived therefrom"

provision of the TRIPS Agreement[191]. Thus, the now prevailing tendency of international agreements to gradually strengthen copyright through ownership and originality measures has been compounded by the fact that there is no longer even the possibility of developing a parallel system of copyright limitations because of the restrictive contours of the "*three-step test*" as a standardized ceiling[192].

The TRIPS Agreement is also aimed at harmonization and has the fundamental characteristic of having led to a further and remarkably important increase in the list of subjects protected by copyright, in particular through the introduction of the protection[193] offered to computer programs and databases[194], a subject we will address later in this work.

Since the copyright system was still practically untouched since 1971, having been only partially innovated by the TRIPS Agreement, mainly through the introduction of the protection granted to software and databases, and therefore even before an adequate solution was found to respond to the threats brought to the rights of right-holders by the advent of the Second Wave, there came the last and most disruptive wave, the Third Wave, identifiable with the invention and dissemination of digital technologies and the Internet in the last decade of the twentieth century. The special feature of this wave was that it allowed any individual to produce perfect copies identical to the original digitised work. Not only that, but the advent of the internet allowed for the almost immediate distribution of such copies from one part of the world to another and potentially to millions of individuals without any difficulty or prohibitive cost. Before the Internet, copyright holders had the right to prohibit the reproduction of copyrighted works for commercial purposes and the right to prevent certain public uses of those works. With the Internet, things change dramatically, because copyright begins to affect many activities by private individuals that were previously considered legal in the analog era. In fact, despite the increase in the number of players, and therefore the increased presence of potential infringers, digital technologies have nevertheless made it possible to more effectively and efficiently monitor their compliance with copyright law[195].

This dynamic in fact leads to the birth of the phenomenon that has been described as the "*digital dilemma*"[196], according to which "*while digital technology can be used to facilitate and improve copying, it can also be used to restrict access to content in ways not possible with the technologies of Gutenberg*"[197]. The heart of the digital revolution can pretty much all be summarized in the ability to reduce information to binary digits. Unlike traditional analog, print, or video, where music, writings, and images were respectively captured and conveyed as physical representations of what was being recorded, digital technology reproduces those same

---

191 Art 13 TRIPS Agreement: Limitations and Exceptions: "Members shall confine limitations or exceptions to exclusive rights to certain special cases which do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the right holder."

192 FISCHMAN AFORI, *The Evolution of Copyright Law and Inductive Speculations as to Its Future*, cit., 251.

193 Art. 10 TRIPS Agreement: Computer Programs and Compilations of Data:

(1) "Computer programs, whether in source or object code, shall be protected as literary works under the Berne Convention (1971)".

(2) "Compilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations shall be protected as such. Such protection, which shall not extend to the data or material itself, shall be without prejudice to any copyright subsisting in the data or material itself".

194 HUA, *Toward a More Balanced Approach: Rethinking and Readjusting Copyright Systems in the Digital Network Era*, cit., 6.

195 TAMURA, *Rethinking Copyright Institution*, cit., 67.

196 National Research Council, The Digital Dilemma: Intellectual Property in the Information Age, 2000.

197 R.S.R. KU, *The Creative Destruction of Copyright: Napster and the New Economics of Digital Technology*, 69 U Chi L Rev 263 (2002), 270. Available at: https://chicagounbound.uchicago.edu/uclrev/vol69/iss1/7/.

images, sounds, and words in the form of numbers. By reducing information to 0s and 1s, digital representation has revolutionized the basic characteristics of content[198]. First, digital information can be delivered to the public without the need for any physical medium such as paper or audio and videotapes. Second, as noted above, unlike analog technologies, digital technologies make it possible to make copies virtually identical to the original, without any loss of quality. Third, digital technology, together with the Internet, has made copying and dissemination of all kinds of digital content not only extremely easy, but also not at all expensive[199].

These features and the economics of digital reproduction have definitively eliminated what were thought to be the major obstacles to copying: the investment required to reproduce and distribute content in physical form, the physical form itself, and the poor quality of reproductions[200].

However, digital technology, as mentioned, is a double-edged sword, in that it now enables what was previously unachievable. Indeed, although it has facilitated the copying and distribution of digital information, it has equally enabled greater control over the use and distribution of information[201]. With the development of digital technology, in fact, "*authors and publishers can have more, not less, control over their work*"[202]. In fact, since digital technology bases its operations on computer code, the same computer code can also be used to regulate user behavior at the right-holder's whim. For example, through the use of trusted systems[203], copyright holders can now use technology to control how works are accessed and used. Trusted systems use encryption[204] to prevent unauthorized access to digital content and use rights management[205] to determine what "*rights*" each user has with respect to a given piece of content[206], dramatically shifting the balance of interests on the side of right-holders and publishers and giving both a considerable amount of power[207].

This scenario, in which a greater potential access is immediately "*balanced*" by the introduction of a stronger level of control, will materialize two years after the TRIPS Agreement came into force, in 1996, when, in response to this disruptive Third Wave of technological revolution, two international treaties were adopted by the WIPO International Diplomatic Conference of Geneva: the *WIPO Copyright Treaty* (WCT) and the *WIPO Performances and Phonograms Treaty* (WPPT). These two treaties were intended to update the general principles of copyright and in particular to modernize them to reflect the birth of the Internet. The one relevant to this work is the WCT. The WCT, as well as incorporating, as

198 *Ibid.*

199 EGER, SCHEUFEN, *The Past and The Future of Copyright Law: Technological Change and Beyond*, cit., 15.

200 KU, *The Creative Destruction of Copyright: Napster and the New Economics of Digital Technology*, cit., 272.

201 *Ibid.,* 274.

202 M. STEFIK, *Shifting the Possible: How Trusted Systems and Digital Property Rights Challenge Us to Rethink DigitalPublishing,*12 Berkeley Tech L J 137 (1997),138. See also: M. GIMBEL, *Some Thoughts on the Implications of Trusted Systems for Intellectual Property Law*, in 50 Stan. L. Rev. (1998), pp. 1671–1687. *JSTOR*, www.jstor.org/stable/1229309.

203 "A trusted system is a system that can be relied upon to follow certain rules. In the context of digital works, a trusted system follows rules governing the terms, conditions and fees for using digital works", STEFIK, *Shifting the Possible: How Trusted Systems and Digital Property Rights Challenge Us to Rethink DigitalPublishing,* cit., 138.

204 "Encryption is the technology used to prevent unauthorized access to computer files typically by scrambling the data with what is known as a public key. The data can then only be unscrambled with the corresponding private key". KU, T*he Creative Destruction of Copyright: Napster and the New Economics of Digital Technology*, cit., 274.

205 "Rights management is the ability of a publisher of a work to define what rights subsequent users of her work will have to use, copy, or edit the work". KU, *The Creative Destruction of Copyright: Napster and the New Economics of Digital Technology*, cit., 274.

206 KU, *The Creative Destruction of Copyright: Napster and the New Economics of Digital Technology*, cit., 274-275.

207 STEFIK, *Shifting the Possible: How Trusted Systems and Digital Property Rights Challenge Us to Rethink DigitalPublishing, cit.*, 155.

did the Berne Convention and the TRIPS Agreement, the three-step test, i.e., the subordination of the introduction of exceptions and limitations to exclusive rights to the condition that this takes place "*in certain special cases that do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the authors*", also added to the previous list of subjects receiving copyright protection two further elements, already present in the copyright system thanks to their introduction in the TRIPS Agreement: computer programs and compilations of data or other materials (databases)[208].

The most important innovation, however, confirming the fact that the advent of digital technology has given right-holders an unprecedented possibility of control, is undoubtedly the introduction of the protection granted to technical protection measures (TPM), which required contracting parties to provide remedies against the circumvention of technological measures used by authors for the protection and exercise of their rights[209].

In particular, these last extensions of copyright law - the widening of the subject matter to software and databases and the protection granted to technical protection measures, later implemented in Europe through the *2001 European Copyright Directive*, and the *1998 Digital Millennium Copyright Act* (DCMA) in the US copyright law - have definitely pushed the balance of interests decisively in favor of the rights holders, i.e. publishers and authors. In particular, the introduction of technological measures to control access and use of electronic content has created many obstacles not only to copying but also to the mere enjoyment of information goods by consumers, which had never been affected before. These legislative choices, implemented to strengthen copyright law, were mainly a reaction of the legislation to cope with the vast changes in the cost structure caused by new copying technologies[210].

2.3.5. A brief introduction to European Copyright Law

So far, we have been dealing with the birth of copyright and its evolution through international law. However, international conventions and treaties are not the only supranational sources of copyright. In fact, there are other regional sources, belonging to the law of the European Union, which, following their ratification by the domestic legislator, contribute to conform the internal law of the member states belonging to the same. Since the 90s of the 20th century, in fact, the European Community, now the European Union, has issued numerous directives aimed at harmonizing aspects, both general and detailed, of the copyright disciplines of the individual member states[211].

In the successive parts of this work we will deal in particular with two directives: *The InfoSoc Directive*[212] and the *Database Directive*[213].

The first directive, in chronological order, since it was introduced in 1996, is the *Database Directive*, which sought to harmonize national laws with regard to the copyright

---

[208] HUA, *Toward a More Balanced Approach: Rethinking and Readjusting Copyright Systems in the Digital Network Era,* cit., 7.
[209] *Ibid.,* 8.
[210] EGER, SCHEUFEN, *The Past and The Future of Copyright Law: Technological Change and Beyond*, cit., 15.
[211] SPEDICATO, *Principi di diritto d'autore*, cit., 29.
[212] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society. Full text available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32001L0029&from=EN.
[213] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. Full text available at: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML

protection of databases. In addition, it required all Member States to introduce new sui generis right for protection of the contents of databases[214].

On the other hand, the more relevant of the two directives, namely the *Information Society Directive* (InfoSoc Directive), in 2001, sought to harmonize certain exclusive rights, provided for one mandatory exception, and created a white list of permitted exceptions. Although, at first glance, the Directive has a rather limited harmonization effect, it is becoming more and more clear that the CJEU is interpreting the Directive in a way that results in a broader harmonization effect[215]. It was the first European directive to harmonize copyright and related rights issues in an extensive and horizontal manner. The InfoSoc Directive implemented into EU law the WCT and the WPPT, signed in 1996. However, some of the rules introduced by this directive, especially those concerning the three-step test, went beyond the provisions of these treaties and have therefore been the subject of controversy[216].

The Infosoc Directive and the Database Directive are part of a set of Directives that were envisaged back in 1988 by the European Commission's Green Paper on "*Copyright and the Challenge of Technology*"[217], whose intent was to balance the rights of authors and right holders with the new challenges arising from the emerging technologies and the Internet. In this context, two sets of directives have been enacted. The first set ("*first generation of directives*") contained six directives focusing on particular areas of Copyright, such as the *Computer Programs Directive*, the *Rental and Lending Directive*, the *Satellite and Cable Directive*, the *Term Directive*, the *Database Directive*, the *Resale Right Directive*. The second set of directives instead ("*second generation Directives*"), was rather horizontal in its scope and intended to harmonize the copyright and the related rights in general. Examples of this set of directives are the *InfoSoc Directive* and the *Enforcement Directive*, whose provisions were not supposed to affect the previously issued Directives, so that when one interprets the provisions of the directives belonging to the strand of the first generation of directives must also take into account the provisions of the directives contained in the second set of directives for the sake of uniformity and consistency[218].

Within the European Union, it must be pointed out that an extremely important role has been played by the jurisprudence of the Court of Justice of the European Union, which in certain cases, as we shall see, has dictated essential principles to define the legal framework on copyright and, consequently, text and data mining. In fact, the Court has not limited itself to precluding the individual initiative of member states to modulate exclusive rights and exceptions and limitations but has played an important interpretative role that has led to further harmonization of concepts not touched by legislation[219].

We will address both directives in detail in the next chapter, particularly the rights and exceptions they provide in relation to text and data mining, to which we refer.

2.4. The outcomes of the copyright evolution: current imbalance of interest of the copyright system. Copyright overprotection: the InfoSoc Directive

214 Sir R. ARNOLD, *An Overview of European Harmonization Measures in Intellectual Property Law*, in A. OHLY, J. PILA, *The Europeanization of Intellectual Property Law*, Oxford University Press, 2013, 33.
215 *Ibid.*
216 GEIGER, SCHÖNHERR, *The Information Society Directive*, cit., 11.01.
217 Green Paper on Copyright and the Challenge of Technology - Copyright Issues Requiring Immediate Action. COM (88) 172 final, 7 June 1988. Available at: http://aei.pitt.edu/1209/
218 GEIGER, SCHÖNHERR, *The Information Society Directive*, cit., 11.01.
219 E. ROSATI, *The Construction of Economic Rights in the Infosoc Directive*, in Copyright and the Court of Justice of the European Union, Oxford University Press, 2019, 86.

From this brief and by no means exhaustive reconstruction of the history of copyright, one conclusion can be drawn: the introduction of technologies that allow consumers to easily access fast and cheap reproduction of cultural goods and knowledge, and more recently the advent of the Internet, which has allowed users to exchange unlimited numbers of files containing protected content, have seriously challenged copyright law. The legal system's response to the alleged ineffectiveness of the measures applied so far is the recent trend of exorbitant copyright expansion, which has manifested itself in a variety of forms, including: the prolongation of the terms of protection; the extension of the catalog of protected goods, resulting in inclusion of goods whose nature is foreign to the copyright regime (e.g., computer programs); the introduction of technological means of protection simultaneously with legal restrictions and a legal prohibition on circumvention of codifications; the reduction of the catalog of limitations and exceptions to copyright; the increasing tendency to transfer copyright rules from the private realm to the realm of criminal law, etc. In general, the trend could be described as the strengthening of rights holder protection through the introduction of the absolute and unlimited property paradigm at the expense of other conflicting interests[220].

This strengthening of copyright, which protects in a greater way the economic interests of rights holders, perceived by part of the doctrine since some time, gave rise to the phenomenon that has been called "*overprotection of copyright*". This perception is certainly not without foundation, especially if we critically look at the evolution of the discipline of authorship over the last three decades[221].

In fact, if we think about the essential and primitive function of copyright, i.e., the legal protection against the risk of losses deriving from free riding activities, precisely those activities that can frustrate the author's interest in recovering the investment and obtaining a fair compensation (fair, appropriate, not maximized)[222], in particular if we quickly consider the legislation that will be largely the subject of the next chapter, i.e., the InfoSoc Directive[223] of 2001, it is difficult not to notice a certain imbalance.

Let us therefore briefly look at some of the elements of the directive to understand whether or not the difficult balancing activity between exclusion and access, outlined in its essential features in the first paragraph, has been correctly implemented[224].

Even looking at the structure and systematics of the directive, one can draw interesting considerations. First of all, it is possible to immediately identify a preferential treatment of the different interested parties. In particular, the protection of right-holders "*of copyright and related rights in the framework of the internal market*"[225] is considered to be a "*mandatory*" task of the member states, as evidenced by the term "*shall*", present in articles 2(1), 3(1) and 4(1); the provisions regarding free access, qualified as limitations (paying access) and exceptions (free access), provided in favor of various categories of users such as researchers and teachers, journalists, critics etc., in article 5, is on the other hand foreseen as "*discretionary*" for member states, as made clear by the use of the term "*may provide*" in article 5(2) and 5(3).

---

[220] K. GRACZ, Brid*ging the Gaps Between Social and Legal Norms Concerning Protection of Intellectual and Artistic Creations: On the Crisis of Copyright Law in the Digital Era*, in The Journal of World Intellectual Property, 2013, 16(1-2), 39.
[221] SPEDICATO, *Principi di diritto d'autore*, cit., 21.
[222] GHIDINI, *Exclusion and Access in Copyright Law: The unbalanced features of the European Directive "on Information Society",* cit., 5.
[223] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society.
[224] The following critical review of the InfoSoc Directive is made by G. GHIDINI in *Exclusion and Access in Copyright Law: The unbalanced features of the European Directive "on Information Society", cit.*
[225] Art 1 (1) InfoSoc Directive.

Furthermore, the directive, on the one hand ensuring the widest possible scope of copyright protection, clearly states that the list of exceptions "*permitted*" (recital 14)[226] is "*exhaustive*" (recital 32)[227]. Thus, since the "*permission*" of each exception is voluntary, allowing more is excluded, while allowing less is allowed and left to the discretion of member states. This ties the hands of the courts, which cannot recognize a legitimate right of access to users, even when this is easily recognizable as fair use, unless there is a specific exception promulgated by national legislation at the time of implementation of the directive.

Even the qualification of the users' right of free access in terms of "*exceptions*" (unlike the Berne Convention which qualifies them as "*free uses*" in Articles 10 and 10bis), denotes a cultural and jurisprudential imprint contrary to the idea that the right of access and use of data and information for reasons of public interest, which is constitutionally guaranteed in various ways, should not be affirmed as an exception but as a full right, to be balanced in a fair and constitutional way with the interest of the copyright holder, a balance to which the directive itself refers in recital 31[228].

This rather closed and unbalanced approach on the side of the private party is further reinforced by another element, widely criticized, which has moved the directive away from the original normative model of the Berne Convention as a balancing of interests. Indeed, under InfoSoc, the actual enjoyment of an "*exception*" granted by a national law seems to be further subject to possible judicial review to establish that, in specific ("*special*") cases, the exceptions and limitations whose application is sought "*do not conflict with a normal exploitation of the work [...] and do not unreasonably prejudice the legitimate interests of the right-holder*" (Article 5(5)).

Therefore, despite the identical wording, this provision does not represent a simple application/transposition of Article 9 of the Berne Convention (or Article 13 of the TRIPS agreement), which were addressed only to national legislators ("*Member States* […]"), providing for a three-step general level of principles/guidelines (the "*three-step test*") "*which the legislators themselves must follow in balancing the rights of rights holders and users*" (recital 31). On the contrary, the Directive, after having precisely and "*exhaustively*" listed the various exceptions and limitations "*allowed*", retrieves the "*three-step test*" that the Berne Convention addressed to national legislators, in order to introduce a specific screen also at the level of judicial "*application*" of such exceptions, without considering that the exhaustive insertion of the provisions of the various exceptions and limitations introduced in the directive was already based on the same test, and thus introducing an additional level of protection to the interests

[226] Recital 14 InfoSoc Directive: "This Directive should seek to promote learning and culture by protecting works and other subject-matter while permitting exceptions or limitations in the public interest for the purpose of education and teaching".

[227] Recital 32 InfoSoc Directive: "This Directive provides for an exhaustive enumeration of exceptions and limitations to the reproduction right and the right of communication to the public. Some exceptions or limitations only apply to the reproduction right, where appropriate. This list takes due account of the different legal traditions in Member States, while, at the same time, aiming to ensure a functioning internal market. Member States should arrive at a coherent application of these exceptions and limitations, which will be assessed when reviewing implementing legislation in the future".

[228] Recital 31 InfoSoc Directive: "A fair balance of rights and interests between the different categories of right-holders, as well as between the different categories of right-holders and users of protected subject-matter must be safeguarded. The existing exceptions and limitations to the rights as set out by the Member States have to be reassessed in the light of the new electronic environment. Existing differences in the exceptions and limitations to certain restricted acts have direct negative effects on the functioning of the internal market of copyright and related rights. Such differences could well become more pronounced in view of the further development of transborder exploitation of works and cross-border activities. In order to ensure the proper functioning of the internal market, such exceptions and limitations should be defined more harmoniously. The degree of their harmonisation should be based on their impact on the smooth functioning of the internal market".

of right-holders, which may, on a case-by-case basis, potentially restrict the exercise of the users' access rights even when these were provided for by the legislator[229].

All this without considering the regime provided by the Directive for the already mentioned "*Technological Protection Measures*" (TPMs), introduced in the European legislation through this directive, which was supposed to implement the WIPO Treaties. The ways and terms in which InfoSoc introduces protection to TPMs seem far from reflecting a reasoned and balanced consideration of the legitimate interests of copyright holders against unauthorized reproduction with regard to the exercise of the equally constitutionally protected rights of users to access and share information and science. For example, not only does the Directive take little care to ensure that TPMs do not apply to non-copyrighted material, although a systematic reading of various provisions of InfoSoc appear to confirm that the application of such measures can only apply to works protected by copyright or related rights[230]; but the Directive also does not provide any effective means to prevent and punish, with appropriate sanctions and procedures, the application of TPM to non-copyrighted or non-copyrighted data and information, a finding that is confirmed by the fact that most Member States have "*ignored*" Article 6(4), jeopardizing the enjoyment of the exceptions, which, let it be reminded, remain mainly optional and left to the discretion of the member states.

[229] See for example: Tribunal de Grande Instance, Paris, 3me ch, 30 April 2004, confirmed by Cour de Cassation, Paris, 1rech, 28February 2006 (comment by C. GEIGER, *Three Steps Test, a Threat to a Balanced Copyright Law?* in IIC, 6/2006, 683), and Tribunale di Milano, Section for intellectual property, 1 July 2009, no. 8787 (G. BONELLI, *Diritto di riproduzione, misure tecniche di protezione e copia privata*, in *Dir. Ind.* 2010, 183 ss).

[230] See, for example, Articles 6(3) and 6(4) (Member States shall take appropriate measures to ensure that rights holders make available to the beneficiary of an exception or limitation...the means of benefiting from that exception or limitation...'), and Recitals 12 and 21, which prohibit the interpretation of the Directive as permitting unlimited blocking power to the owner of copyright - or related rights. Accordingly, national laws implementing the Directive should be interpreted to deny TPM protection if they cover data, information, and works that are not copyrightable, unprotectable, or otherwise subject to lawful unfettered access.

CHAPTER 3

THE EUROPEAN LEGAL LANDSCAPE ON TEXT AND DATA MINING

The first paragraph of this chapter will start with the analysis of the rights included in the European 2001 *InfoSoc Directive* and of the sui generis rights included in the 1996 *Database Directive*, particularly focusing on their specific compatibility with text and data mining activities.

We will first of all analyse the regulation that was in place prior to the new *Digital Single Market Directive* 790/2019, concerning exceptions to copyright and sui generis rights; furthermore we will try to evaluate whether these exceptions to the rights conferred by the two directives could and still can be considered sufficiently broad and suitable to legally carry out the wide range of different text and data mining activities.

We will then look at the very first exceptions introduced by various European countries that could be considered pioneers in text and data mining and describe their main characteristics.

We will introduce the different initiatives undertaken by the European Union during the last decade in the field of text and data mining, focusing our attention particularly on the EU legislative process that led to the introduction of Directive 790/2019, which eventually introduced two exceptions for text and data mining, then briefly discussing the criticism made to it and to these exceptions and its current implementation at national level.

Finally, we will present some alternative solutions to the issue, proposed by the European legal doctrine.

3.1. Relevant European Directives and exclusive rights in Copyright and Database Protection Law in Europe and their possible infringement throughout TDM process

We will initially focus our attention on the two directives that have most affected and (probably) still affect the possibility of carrying out certain actions in the TDM process and that therefore constituted and, in part, still constitute (due to the various references made to them in the new Directive 790/2019) its main source of regulation, namely: Directive 2001/29/EC *on the harmonisation of certain aspects of copyright and related rights in the information society*[231] (henceforth *InfoSoc*) and Directive 96/9/EC of 11 March 1996 *on the legal protection of databases*[232] (*Database Directive*).

Let us now describe, for both of the aforementioned directives, the subject matters covered by their protection and subsequently the exclusive rights granted therein, in particular analysing them in the light of the activities carried out in the common text and data mining process in order to understand how they may be hindered by these.

In the first chapter of this Thesis, we pointed out that in the common process of TDM there are actions of reproduction and copying which, when carried out through the use of material protected by copyright or sui generis rights, may imply one or more infringements of these rights.

[231] Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society.

[232] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

As above extensively explained, text and data mining consists of a set of research techniques that, by using automated digital tools, extracts useful information from large amounts of data. This is done, first of all, by finding material to be analysed, which is usually organised in databases; copying substantial amounts of it, which is then pre-processed to make it machine-readable and, although not always, uploaded into a platform so that the data can be extracted and, eventually, useful and previously unknown patterns can be extracted[233]. Therefore, TDM often, though not always, involves some activities of copying and adaptation of material which, even if it occurs on a limited extent of material, can be considered an infringement, in the light of the exclusive rights granted by the two directives just mentioned. These activities may be carried out with the use of material in the form of text or data, both of which may be covered by the protection offered by intellectual property, either by copyright or by the database sui generis right, or, on the contrary, may be outside the scope of protection (e.g. because lacking the requirement of originality or public domain material)[234].

The exclusive rights granted to the copyright holder that may be infringed, and that we are going to analyse hereinafter, are the reproduction right (Article 2 of the InfoSoc Directive) and, to a lesser extent, the communication and distribution rights (Article 3 of the InfoSoc Directive and article 5 of the Database Directive). On the other hand, with regard to the sui generis database right, we shall focus on Article 7 of the Database Directive.

3.1.1. European Copyright Law. The Infosoc Directive: subject matter and the originality requirement

We have just mentioned in the general introduction to copyright that an essential feature common to most copyright laws around the world is that authors of particular intellectual works are granted special exclusive rights to exploit their works because of their originality.

What rights are granted to authors and what types of works are eligible for protection is usually a matter of national law. However, there have been several harmonisation initiatives at international level over time. A first attempt to harmonise the subject of copyright protection stems from the above-mentioned *Berne Convention*[235]. Article 1 of the Directive provides that subjects of protection are: "*literary and artistic works*"[236]. This expression includes a wide range of different types of works, varying from books, pamphlets and other writings; to the lectures, addresses, sermons and other works of the same nature; dramatic or dramatic-musical works; choreographic works; musical compositions; cinematographic works; works of drawing, painting, architecture, sculpture, engraving and lithography; photographic works; works of applied art; illustrations, maps, plans, sketches and three-dimensional works relative to geography, topography, architecture or science[237], to the point of even protecting collections of literary and artistic works, that "*by reason of the selection and arrangement of their contents, constitute intellectual creations*"[238], being the latter protected also by other international treaties such as the already mentioned *TRIPS agreements* or the *WIPO treaties*.

With regard to the process of legislative harmonisation of copyright matters in the context of the European Union, it must be said that it is generally agreed that it has

233 C. GEIGER, G. FROSIO, O. BULAYENKO, *Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?,* in IIC 2018, Vol. 49, No. 7, 817.
234 *Ibid.*
235 The Berne Convention for the Protection of Literary and Artistic Works.
236 Article 1 Berne Convention: Establishment of a Union.
237 Article 2 (1) Berne Convention: Protected Works: *"Literary and artistic works".*
238 Article 2 (5) Berne Convention: Protected Works, *Collections.*

deliberately ignored the issue of categorisation of general subject matters. Indeed, the Commission has been reluctant to intervene in this area over time, noting that "*[m]any of issues of copyright law, do not need to be subject of action at Community level. Since all Member States adhere to the Berne Convention […], a certain fundamental convergence of their laws has already been achieved*"[239].

On the other hand, as regards the criteria or threshold of originality necessary for a work to be considered protected by copyright, this too has not been harmonised at European legislative level but rather achieved through case law, with works proving to be "*the author's own intellectual creation*" being considered original and therefore protectable by copyright[240]. Thus, we can say that, regarding the originality requirement, there is a uniform standard which was provided by the CJUE that has not been completely implemented by the members states. As a result, the definition of "*intellectual creation*", the minimal level of originality, and the criteria, standards, and components that might fill-in the substance of the notion of intellectual creations are left to national legislators and courts[241].

The only areas where it was deemed necessary to regulate such criteria, leading to a "*vertical*" harmonisation (i.e. through legislative interventions regulating only specifically identified subject matter), were the protection of computer programs in 1991 (through the *Software Directive*), photographs (*Term Directive*) and databases (*Database Directive* of 1996), directives from which the requirement of "*author's own intellectual creation*" was first derived[242]. Similar provisions concerning the criteria necessary for a work to obtain protection[243] are therefore not present in any other European legislative initiative, and the harmonisation of other subject matters has therefore taken place just through the pronouncement of the CJEU ("*horizontal harmonization*"), which has applied the criterion of the "*author's own intellectual creation*" in various judgments and in various different fields[244][245].

As regards newspaper articles, for example, a famous ruling is extremely useful for us in order to ascertain whether or not an infringement of copyright can be deemed to have occurred along a TDM activity. We refer to the so called *Infopaq case*[246], in which the CJEU for instance stated that the extraction of part of the text contained in a newspaper, consisting of 11 words, falls within the scope of the author's exclusive right to authorise or prohibit the making of reproductions, if that part constitutes the author's own intellectual creation. With this ruling, it was argued, the CJEU has de facto harmonised the originality requirement for works falling within the scope of the InfoSoc Directive, thus imposing a very high threshold of protection on such works[247].

---

239 Communication from the Commission of the European Communities, "Green Paper on copyright and the challenge of technology – Copyright issues requiring immediate action", COM(88) 172 final, 4.

240 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 18.

241 T. MARGONI, *The Harmonisation of EU Copyright Law: The Originality Standard*, June 2016, 5. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2802327

242 *Ibid.*

243 Respectively: Article 1(3) of Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs (Software Directive), Article 6 of Directive 2006/116/EC of the European Parliament and of the Council of 12 December 2006 on the term of protection of copyright and certain related rights (Term Directive), and Article 3(1) of Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases (Database Directive).

244 This happened, in particular, between 2009 and 2012, in five key CJUE decisions. See: *Infopaq International v. Danske Dagblades Forening* [2009]; C- 393/09 *Bezpečnostní softwarová asociace v. Ministerstvo kultury* [2010] E.C.R. I-13971; C-403/08 and C-429/08 *Football Association Premier League and Others v. QC Leisure and Others* and *Karen Murphy v. Media Protection Services* [2011] E.C.R. I-09083; C-145/10 *Eva-Maria Painer v. Standard VerlagsGmbH and Others* [2011] E.C.R I-12533; C-604/10 *Football Dataco v. Yahoo! UK and Others* [2012] EU:C:2012:115.

245 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 18.

246 ECJ, 16 July 2009, Case C-5/08, Infopaq International A/S v Danske Dagblades Foreing, §36.

247 E. ROSATI, *The Construction of Economic Rights in the InfoSoc Directive*, in Copyright and the Court of Justice of the European Union, Oxford University Press, 2019, 91.

What the CJEU means by the concept of author's own intellectual creation is not precisely defined by the CJEU itself, but it is clear that what is essential is that the author put his personal imprint to the work and that the concepts to be used to assess its originality are the notions of "*creativity*", "*creative freedom*" and "*free and creative choices*"[248]. According to the CJEU, this is to be found for example in the "*choice, sequence and combination*[249]" of words in newspaper articles, or the "*specific arrangement or configuration*" of a GUI's components[250], or in the case of photographic works in the choice of the background, subject's pose, framing, angle of view or developing technique[251][252]. The Court did not rule on the amount of creativity required for a subject-matter to become an author's own intellectual creation, but it has defined the boundaries of the originality requisite in a negative way, holding that a work should not be regarded as original where choices are dictated by technical function[253], rules or constraints[254], e.g. when there is only one way to express an idea or when the expression is predetermined by the specific objective or dictated by rules as a consequence of which the author is not able to *"[express] his creative ability in an original manner by making free and creative choices*"[255].

3.1.1.1. Article 2 of the InfoSoc Directive: the reproduction right

Having made the necessary preliminary remarks about the types of works that are protected by copyright in the European Union and the characteristics that these works must have in order to obtain this protection, let us see the exclusive rights that copyright confers and that most constitute barriers to the activities of TDM.

Among the exclusive rights recognised by copyright, the most important one is certainly the right of reproduction, which has always constituted the central and essential core of the broader right of the work economic exploitation[256]. Since the days of the printing privileges, it has always been the first right to face the challenges of technological development and the expansion of the subject matter of copyright[257].

Despite the considerable relevance of this right in the digital environment, the first international conventions that regulated copyright and related rights at the beginning of the digital era, i.e. the two WIPO treaties of 1996, did not intervene on it. They referred instead to the old definition of reproduction right provided by the Berne Convention, justifying this decision by claiming that the international acquis contained provisions that could be easily adapted to the digital context. It is in this context that the European Union, faced with a considerable risk of fragmentation of national laws, addressed the issue by harmonising - starting with the 1991 Software Directive, and subsequently with other directives and thanks to the intervention of the CJEU - the different approaches to the regulation of exclusive

[248] Football Dataco v. Yahoo! UK and Others [2012] EU:C:2012:115,), §38; *Infopaq International v. Danske Dagblades Forening* [2009], §45; Bezpečnostní softwarová asociace v. Ministerstvo kultury [2010], §50; Eva-Maria Painer v. Standard VerlagsGmbH [2011], §§ 89, 92.
[249] ECJ, 16 July 2009, Case C-5/08, Infopaq International A/S v Danske Dagblades Foreign, §§ 34-36.
[250] CJEU Bezpečnostní softwarová asociace v. Ministerstvo kultury [2010], §48.
[251] CJEU 1 December 2011, C-145/10 (Painer), §91.
[252] CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 18.
[253] CJEU Bezpečnostní softwarová asociace v. Ministerstvo kultury [2010], §48-49.
[254] CJEU 1 March 2012, C-604/10 (Football Dataco), §39.
[255] CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 18.
[256] SPEDICATO, *Principi di diritto d'autore*, cit., 96.
[257] C. SGANGA, *The right of reproduction*, in E. ROSATI (ed.), *The Routledge Handbook of EU Copyright Law*, Routledge, 2021, 1 (also available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3803999).

rights adopted by member states, some of which were broad categories and others detailed lists[258].

Let us therefore see how this right is defined by the directive concerned and by the CJEU, then going on to highlight how this may limit the copying activities carried out in the TDM process.

In European Union law, the reproduction right is provided for in Article 2 of the InfoSoc Directive. This article goes beyond Article 9 (1) of the Berne Convention, which states that:

> *Authors of literary and artistic works shall have the exclusive right of authorizing the reproduction of these works, in any manner and form*[259]*,*

giving, instead of a definition, a broad description of the reproduction right, providing[260]:

> *Member States shall provide for the exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part:* (a) *for authors, of their works*[261]

It is evident at first glance that the rule was deliberately formulated in a very broad manner, encompassing the exclusive right of reproduction not only of the creation of "*permanent*" copies, but also of "*temporary*" copies (now almost inevitable with the advent of digital devices), direct or indirect, by any means and in any form, in whole or in part. This is undoubtedly the result of the unstoppable diffusion of digital technologies that we have talked so much about, which have made extremely easier and cheaper to make copies almost identical to the original. The reassuring presence of a material support incorporating the work, whose circulation was relatively easy to control, has been replaced by the proliferation of digital copies, which are much more difficult to control[262]. Thus, to this diminished power of control over the creation and circulation of digital copies, the legislator responded by extending the perimeter of the author's power of control[263].

And in fact, it is evident that, in addition to formulating the rule very broadly, the directive itself states that the term reproduction must be interpreted broadly. Indeed, one could draw this conclusion from recital 21 of the Directive, which states that:

> *This Directive should define the scope of the acts covered by the reproduction right with regard to the different beneficiaries. This should be done in conformity with the acquis communautaire. A broad definition of these acts is needed to ensure legal certainty within the internal market*[264].

Not to mention that the European Court of Justice, in the *Infopaq case* of 2009[265], gave a rather broad interpretation of the rule, stating that the reproduction right is to be considered

---

[258] *Ibidem*, 2.
[259] Article 9 (1) Berne Convention.
[260] ROSATI, *The Construction of Economic Rights in the InfoSoc Directive*, cit., 86.
[261] Art. 2 InfoSoc Directive.
[262] SPEDICATO, *Principi di diritto d'autore*, cit., 98.
[263] *Ibid.*
[264] Recital 21 InfoSoc Directive.
[265] ECJ, 16 July 2009, Case C-5/08, Infopaq International A/S v Danske Dagblades Foreign. The Infopaq case, which we have already mentioned at the beginning of this chapter when discussing the requirement of harmonisation at European level of the originality requirement and which we will certainly continue to mention

as infringed in the presence of an unauthorised copy of even just 11 words, highlighting how the margins for manoeuvre are rather limited. In fact, the judgment reads:

"An act occurring during a data capture process, which consists of storing an extract of a protected work comprising 11 words and printing out that extract, is such as to come within the concept of reproduction in part within the meaning of Article 2 of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, if the elements thus reproduced are the expression of the intellectual creation of their author"[266].

Therefore, looking at the reproduction right as outlined by both the Directive and the CJEU, it is necessary to acknowledge that those activities involving the copying of material, necessary in the TDM process, and carried out on copyright protected material, are certainly to be considered acts of reproduction falling within the scope of protection of the reproduction right and are limited by it.

This conclusion, however, may not be equally true for certain exceptional circumstances, given by the specific type of TDM process chosen for the project. There are in fact cases in which the software used for the project of TDM only "*crawls*" through the texts and data and processes them "*one by one*", without proceeding to copy all the text or all the data but instead copying only one data or word or only a few of them, which is sufficient in order not to lead to an infringement. In this case, when the software only "*swallows*" one or two words or data at a time without keeping a copy of them but only for instance "*counting*" the number of times the "*malaria*" word appears, it should not be considered as copying in terms of copyright. Therefore, the rightsholder's permission should not be necessary to carry out such a kind of activity and, consequently, it is not even necessary to have an exception that would make this otherwise unlawful activity lawful[267].

3.1.1.2. Article 3 of the InfoSoc Directive: communication and distribution right

Other rights included in the InfoSoc Directive to be taken into consideration, although to a lesser extent, are the rights of communication and distribution provided for in Article 3 of the Directive, on which we'll dwell just briefly, since they are not to be considered so relevant for the purposes of our study.

Article 3 of the Infosoc Directive provides for the right of communication, making available and distribution, establishing that the Member States shall provide for the insertion in their national legislation of provisions enabling the:

> *Authors with the exclusive right to authorise or prohibit any communication to the public of their works, by wire or wireless means, including the making available to the public of their*

later, is a very important case. In fact, it has addressed and resolved several crucial issues of copyright, including the concept of originality, the methodology for assessing infringement and the interpretation of exceptions and limitations, effectively harmonising the same aspects. For a general overview of the facts of the case and its repercussions on EU copyright law, see: J. GRIFFITHS, Infopaq, BSA and the 'Europeanisation' of United Kingdom Copyright Law, in Media & Arts Law Review, Vol. 16, 2011. Available at SSRN: https://ssrn.com/abstract=1777027

[266] ECJ, 16 July 2009, Case C-5/08, Infopaq International A/S v Danske Dagblades Foreing.

[267] TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* cit., 31.

works in such a way that members of the public may access them from a place and at a time individually chosen by them[268].

As regards the relationship of these rights with TDM, it is useful to distinguish two situations, in which one could potentially think of an act of communication of copyright protected material taking place: communications that take place during the TDM process and communications that occur at the end of the process, i.e. concerning the output of the process. As regards the first case, it is very difficult for a distribution to take place at this stage, since the only person in possession of the material to be mined is the analyst and there are no moments in which the material to be analysed is distributed to a "*public*". As far as the output of the process is concerned, in most cases there are hardly any moments when the output resulting from the analysis needs to be distributed, communicated or made available to a public, but instead it remains rather confidential or available to a few people. However, it may happen that in some cases the results of the process, very often in the form of reports, graphs or tables, are made public. Even if this is the case, it is highly unlikely that the copied material will be visible, as it is highly processed at the end of the TDM process and presented at most in the form of statistics, occurrences, patterns, relationships etc. It may be, if the analysis concerns parts of text, that a few words belonging to copyrighted material are recognisable, but it will hardly be possible to conclude that the right of communication has been infringed[269].

3.1.2. The Database Directive: subject matter e requirements for protection

As mentioned earlier, both international and EU law provide that compilations of works can also obtain copyright protection, irrespective of whether the works they contain are protected or not[270].

According to the definition given in the Database Directive, the following can be considered a database:

a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means[271].

We have also mentioned that within the European Union, the Database Directive is one of the three legislative initiatives that has harmonised the criteria necessary for a certain type of work to be considered as protected by copyright. According to the Directive, in fact, the databases receive copyright protection when:

by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation […][272].

Therefore, copyright protects the peculiar expressive form of the database, the originality of its systematic organisation, which is realised through "*free and creative choices*" that show the author's "*personal touch*", while the requirement of originality cannot be considered

[268] Art. 3 InfoSoc Directive.
[269] TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* cit., 33.
[270] CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 19.
[271] Art. 1 of the Database Directive.
[272] Art. 3 Object of Protection.

present "*when the setting up of the database is dictated by technical considerations, rules or constraints which leave no room for creative freedom*"[273].

But the Database Directive does not merely cover this area. In fact, an important aspect of the Database Directive is that it offers a double-tier protection system. Namely, one part of it deals with copyright and confers some typical rights attributed by copyright law to the author of a database, which will be briefly analysed hereinafter. The second part instead, in addition to the typical rights conferred by copyright, aims to protect further and in a different way the creators of databases, giving them a sui generis database right on the database itself due to the fact that:

> *there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents […]*[274].

Thus, the main reason for introducing the so-called sui generis right was the lack of protection granted not to the database itself, but to its content, which is said to often be the most substantial part of the investment in the creation of a database[275]. In fact, the copyright rights conferred by the Database Directive itself do not protect the contents of the database but only its structure, so that extracting the contents without the structure could not, without the introduction of an adequate and additional protection of the same, lead to an infringement. A second reason for the introduction of such additional rights to the protection provided to the database by copyright, as made clear by recitals 11 and 12 of the Directive[276], was the will to make database industry competitive with that of other countries, first and foremost the United States[277].

This additional protection, as provided for in Article 7 (4) of the Directive[278], is therefore granted irrespective of the existence or absence of protection for the database or its content, the rights conferred to the author of the database being separate from those deriving from the protection offered by copyright[279]. The two forms of protection thus operate independently of each other and can therefore both be present on the same database if the requirements for the rights to be considered applicable are met. The crucial aspect is that these rights therefore tend to protect two totally different goods: originality in the selection or arrangement of the database in the case of copyright on the database; the

---

273 R. DUCATO, A. STROWEL, *Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to "Machine Legibility*", CRIDES Working Paper Series, 2018, 6.

274 Art. 7 (1) of the Database Directive.

275 See also BHB v. William Hill, §46, which, quoting and emphasising recital 48 of the directive, says: "According to the 48th recital of the preamble to the directive, the sui generis right has an economic justification, which is to afford protection to the maker of the database and guarantee a return on his investment in the creation and maintenance of the database".

276 Recital 11: "Whereas there is at present a very great imbalance in the level of investment in the database sector both as between the Member States and between the Community and the world's largest database-producing third countries"; Recital 12: "Whereas such an investment in modern information storage and processing systems will not take place within the Community unless a stable and uniform legal protection regime is introduced for the protection of the rights of makers of databases".

277 E. DERCLAY, *The Database Directive*, in I. STAMATOUDI, P. TORREMANS, *EU Copyright Law*, 9.31.

278 Art 7(4) Database directive: "The right provided for in paragraph 1 shall apply irrespective of the eligibility of that database for protection by copyright or by other rights. Moreover, it shall apply irrespective of eligibility of the contents of that database for protection by copyright or by other rights. Protection of databases under the right provided for in paragraph 1 shall be without prejudice to rights existing in respect of their contents".

279 TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* cit., 37.

substantial investment in the obtaining, verification and presentation (but not creation) of the data in the case of the sui generis right[280].

The CJEU has also ruled on the qualification of the different types of investment that can make the database susceptible to protection by the sui generis right. It has held in several judgments that the investment in the obtaining can be the use of resources to "*seek out existing independent materials and collect them in the database*"[281]. It also specified, however, how such an investment must be made independently of the resources used to create the materials[282][283]. With regard to the investment for verification purposes, it argued that the investment in ensuring that the information in the database is reliable and monitoring the accuracy of the material collected when the database was created, but also during its operation, can be considered as such[284].

3.1.2.1. Copyright in the Database Directive: Article 5

Article 5 of the Database Directive, which is included in the part of the directive dedicated to the protection offered by copyright to databases, provides, not differing much from the corresponding rights provided for in the InfoSoc Directive already analysed:

> *In respect of the expression of the database which is protectable by copyright, the author of a database shall have the exclusive right to carry out or to authorize:*
> *(a) temporary or permanent reproduction by any means and in any form, in whole or in part;*
> *(b) translation, adaptation, arrangement and any other alteration*
> *(c) any form of distribution to the public of the database or of copies thereof. The first sale in the Community of a copy of the database by the rightholder or with his consent shall exhaust the right to control resale of that copy within the Community;*
> *(d) any communication, display or performance to the public;*
> *(e) any reproduction, distribution, communication, display or performance to the public of the results of the acts referred to in (b)*[285].

Very similar considerations will necessarily be made here to those already made with regard to the corresponding rights provided by the InfoSoc Directive.

The first restricted act (reproduction) implies that no act of reproduction may be performed on the whole or a substantial part of a database by means of a partial copy of it, provided that this portion of the database presents the requirement of originality and that it is recognisable. The right at issue certainly mainly concerns the initial steps of the TDM process, those in which the material to be mined need to be found, and applies in particular when the selection or the arrangement of the database is copied; for instance, this may be the case when all the data contained in a database are copied: in that case, the selection is

---

[280] T. MARGONI, *The Harmonisation of EU Copyright Law: The Originality Standard*, June 2016, 12. Available at SSRN: https://ssrn.com/abstract=2802327,

[281] ECJ 9 November 2004, C-444/02 (Fixtures Marketing v. OPAP), §40; ECJ 9 November 2004, C-46/02 (Fixtures Marketing v. Veikkaus Oy), §34; ECJ 9 November 2004, C-338/02 (Fixtures Marketing v. Svenska Spel), §37.

[282] BHB v. William Hill, §35; Fixtures Marketing v. OPAP, §40; Fixtures Marketing v. Veikkaus Oy, §34; Fixtures Marketing v. Svenska Spel, §37.

[283] For a more detailed and thorough discussion on this point see: CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 20.

[284] ECJ 9 November 2004, C-203/02 (BHB v. William Hill), §34.

[285] Art. 5 Restricted acts of the Database Directive.

copied; in other cases, it may also be the case that parts of the arrangement are copied during the process; however, this will not always be the case and will depend, as already mentioned for the reproduction right contained in the InfoSoc Directive, on the circumstances and techniques used[286]. With regard to the concrete delimitation of this right of reproduction, the CJEU has ruled on the matter in the already mentioned *Infopaq case*, stating that when a reproduction takes place on a database, the reproduction right only cover those parts that are original[287]. This, it was noted, has implications, since in the phases following the retrieval of the material, in which the copied content is skimmed and variously combined to form a dataset to be mined, the original selection and the structure of the database may be lost, with the result that "*elements on which copyright could have a grasp have become undetectable*". It has been pointed out, however, that it is not entirely certain that CJEU interprets the reproduction right provided by the Database Directive in the same way as its InfoSoc counterpart[288].

As regards instead the final phases of the TDM process, i.e. those concerning the outputs of the process, this is generally not considered a problem since it is considered really unlikely that the outputs can contain whole or part of a protected database which are also recognisable, because of the modifications that the process makes on the original content, usually composed of hundreds or even thousands of data or works that often come from different databases that are joined together, which are then variously analysed, indexed, compared, linked, aggregated, merged, disassembled, etc., in such a way that it is really very difficult to prove that the data come from a particular database and, more importantly, that they infringe upon the selection or arrangement of the database (which is what copyright protects in a database)[289].

As regards the other rights, in the typical TDM process we have described, both the right of distribution and the right of communication to the public can only come into consideration in the final phase of publication of TDM process results[290]. During the analysis phase of the data mining process, in fact, a communication "*to the public*" cannot be considered if a database is being mined by a group of researchers or a private company, since a communication to the public require that the work be communicated to a significant number of persons[291]. In the final phase, the publication of the results of the TDM process, it may indeed happen that elements of the selection and arrangement that are considered original are contained among these results, which is, however, very difficult in the case of small part (of one of the contents) of the databases is cited or otherwise included in the results[292]. Even if, however, a content can be considered original, it is really unlikely that this could configure an act of communication to the public of the protected elements of the database, i.e. the selection and arrangement of the same, considering that the output is in most cases a new text, a graph, a table etc., therefore an independent creation in which the original data that also make up the new work are not communicated to the public, being the

286 TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* cit., 34.
287 ECJ, 16 July 2009, Case C-5/08, Infopaq International A/S v Danske Dagblades Foreing.
288 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 24.
289 TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* cit., 35.
290 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 24.
291 TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* cit., 36.
292 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 24.

output of the analysis usually communicated through statistics, reports and patterns in which by definition the works from which they are taken are not visible[293].

The same can be said for the adaptation right, for which the Database Directive provides for a right of "tra*nslation, adaptation, arrangement and any other alteration*" in Article 5(b). It may happen that an original database is adapted for the creation of a dataset to be mined, for example by selecting only some of the contents or changing their arrangement. It can also happen that all the data and therefore the selection made by the database maker are translated or transformed in their format. Here, therefore, there was not only a reproduction but also an adaptation[294].

3.1.2.2. The sui generis right in the Database Directive. Article 7

Two rights are conferred to the database author by the Database Directive in the sui generis right part: the extraction right and the reutilization right.

As defined by the directive and the interpretations of the CJEU, the two rights of extraction and reutilization are very closely related to, respectively, the right of reproduction and the right of communication offered by copyright[295].

With regard to the right of extraction, where by "*extraction*" the Directive means:

> *the permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form*[296]

and which, according to the CJEU, takes the form of "a*ny unauthorised act of appropriation*"[297] of a whole database or a part of its contents, as a result of which the appropriated material is "*placed on a medium other than that of the original database*"[298], it is almost unanimously agreed that the term "*extraction*" itself should be interpreted broadly. This is demonstrated not only by the wording of the rule, which is clearly formulated in such a way as to require authorisation for permanent or temporary transfers, in whole or in part, of a database, by any means and in any form, but also by the interpretation of the CJEU, which has expressed a ruling on the subject in the case *The British Horseracing Board LTd and Others v William Hill Organization Ltd.* The Court has here held that:

"*The use of expressions such as 'by any means or in any form' and 'any form of making available to the public' indicates that the Community legislature intended to give the concepts of extraction and re-utilisation a wide definition. In the light of the objective pursued by the directive, those terms must therefore be interpreted as referring to any act of appropriating and making available to the public, without the consent of the maker of the database, the results of his investment, thus depriving him of revenue which should have enabled him to redeem the cost of the investment*[299]".

In order that the protection offered by the sui generis right may arise, the acts of extraction must be carried out on material which forms a substantial part of the database, to be ascertained, as previously stated, having regard to the investment by the rightholders

---

293 TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* cit., 37.
294 *Ibid.*, 36.
295 E. DERCLAY, *The Database Directive*, cit., 9.43.
296 Art. 7 (2)(a) Database Directive.
297 ECJ 9 November 2004, C-203/02, BHB v. William Hill, §67.
298 ECJ Apis-Hristovich v. Lakorda, §45.
299 ECJ 9 November 2004, C-203/02, BHB v. William Hill, §51

addressed, which can be verified both with the use of a quantitative criterion, and with a qualitative one. As regards the former, it is interpreted by the CJEU as "*the volume of data extracted from the database and/or re-utilised, [...] assessed in relation to the volume of the contents of the whole [database]*"[300]. For the qualitative part instead, according to the CJEU, the substantiality of the portion of the database extracted must be verified having regard to the "*scale of the investment in the obtaining, verification or presentation of the contents*" that are re-utilised, independently of the fact that it can be considered a substantial part quantitatively[301].

Analysing this right specifically in the light of TDM activities, which, as we have said several times, require preventive actions, it can be said that different extractions can be made when data are retrieved from different sources. For example, often the entire database is used or the same database is used to then choose the data that will be used for the analysis, thus resulting in a high possibility that a substantial part of it is taken. Another case may be where the miner selects data from a third-party database and then retrieves only the selected data in a new dataset[302]. It is easy to argue that in these cases, which are not uncommon, the acts of extraction are such that it must be considered necessary to seek permission from the creator of the database when this concerns a substantial part of it[303].

This seems to be confirmed by the CJEU in the same judgment mentioned above, in the part in which it states:

"*Although William Hill is a lawful user of the database made accessible to the public, […] it appears from the order for reference that it carries out acts of extraction and re-utilisation within the meaning of Article 7(2) of the directive. First, it extracts data originating in the BHB database by transferring them from one medium to another. It integrates those data into its own electronic system. Second, it re-utilises those data by then making them available to the public on its internet site in order to allow its clients to bet on horse races*[304]".

In the mining phase in the strict sense, then, the final dataset thus composed is analysed by the mining software. In this phase numerous extractions are made of the contents of the database, since parts of the database are continuously copied into the working memory of the computer. Although they are only temporary, this does not count, as the right of extraction is very broad and also covers temporary copies[305].

Again, as above mentioned when we talked about the right of reproduction in copyright, not in all TDM processes an action must be performed that is relevant to the exclusive right of extraction from third party databases. Indeed, the specific TDM technique chosen for a TDM project might only use a software that copies or extracts a few elements of the database, in such a small number that the extraction cannot be considered "*substantial*"[306].There may also be a situation where, as above, permanent or temporary transfers of data do not occur at all, such as when the analysis software only counts the number of occurrences or registers a link between this data and another data, without making a copy of them, and thus without ground at all for the creator of the database to invoke his sui generis right[307].

---

300 *Ibidem*, §70.
301 *Ibidem*, §71.
302 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 26.
303 *ibidem.*
304 ECJ 9 November 2004, C-203/02 BHB v. William Hill, §65.
305 TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* cit., 38.
306 *Ibidem.*
307 *Ibidem.*

Regarding the right of reutilisation, art. 7(2)(b) provides:

> "*re-utilization" shall mean any form of making available to the public all or a substantial part of the contents of a database by the distribution of copies, by renting, by on-line or other forms of transmission. The first sale of a copy of a database within the Community by the rightholder or with his consent shall exhaust the right to control resale of that copy within the Community*;[308]

The right of re-utilisation has also been interpreted several times by the CJEU. It has held that this right can apply to situations where there is a distribution, in whole or in part (albeit substantial), of its contents[309], to a number of persons who must be considered to be undetermined[310]. As above, when we talked about the right of communication contained in the InfoSoc directive, we must also distinguish here two situations in which such an act of reutilisation can potentially take place within a TDM process: acts of reutilisation of the content of the database that take place during the phase of analysis in the strict sense; and acts of reutilisation that take place at the end of the analysis, which concern the outputs of the process itself. With regard to the first situation, as already mentioned, it is very unlikely that such an act of distribution can take place during the actual analysis phase, considering the fact that, since the contents retrieved by the database are exclusively analysed by a mining software, not even the researchers are given a real disclosure of these contents. It can therefore be stated with some certainty that acts of reutilisation cannot take place in this phase[311].

With regard instead to the second scenario, the phase of publication of the results of the TDM process, it is possible, albeit in a minority of cases, that such contents are published, and that such results include such contents, e.g. in cases of quotation or examples of data used. Here, it was noted, the only question to be asked is whether the material thus disclosed and redistriubed can be considered a substantial part, as required by the norm. This question must be answered by considering the two criteria, already mentioned, to be used to verify when a part of the database can be considered substantial: the quantitative and qualitative criteria. As regards the first, interpreted by the CJEU, we have seen, as "*the volume of data extracted from the database and/or re-utilised, [...] assessed in relation to the volume of the contents of the whole [database]*"[312], one can safely say that the inclusion of certain elements of the database in the publication can hardly be considered substantial for this purpose. With regard to the second instead, the qualitative criterion, which according to the CJEU must be verified having regard to the "*scale of the investment in the obtaining, verification or presentation of the contents*" that are re-utilised, regardless of whether it can be considered a substantial part quantitatively[313], (for the interpretation about the individual cases see above), it is however very difficult to affirm even here that the inclusion of some data in the publication of the results can be ever considered a substantial part in qualitative terms[314].

---

308 Art. 7(2)(b) Database Directive.
309 CJEU BHB v. William Hill, §67; ECJ 18 October 2012, C-173/11 (Football Dataco v. Sportradar), §20; CJEU Innoweb v. Wegener, §37.
310 CJEU Innoweb v. Wegener, §51.
311 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 26.
312 CJEU BHB v. William Hill, §70.
313 *Ibidem,* §71.
314 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 27.

3.2. The European framework of the exceptions to copyright before the Directive 790/2019

Having dealt in the previous section with the rights conferred by the InfoSoc and Database Directives and their difficult compatibility with TDM activities, we will now introduce the panorama of exceptions provided by the two directives considered, to investigate the possible subsumption of text and data mining activities within the scope of these exceptions and the difficulties of their application to TDM. We will conclude that the framework of exceptions stemming from the two directives already analysed was not broad enough to accommodate most of the text and data mining activities and did not provide adequate certainty to users. Indeed, when the European Commission was asked as to whether European law already authorized TDM, the Commission itself answered by saying that although it had to be acknowledged that Union law already provided certain exceptions that "*may apply*" to TDM, it also acknowledged that they were "*optional and not fully adapted*" to it[315].

In the second part of this paragraph, we will then review the various laudable initiatives meant to introduce specific exceptions for TDM, undertaken by some member states over the last decade.

Let us therefore briefly introduce here the framework of copyright exceptions in the EU. The first legislative initiative that needs to be mentioned when talking about exceptions to copyright in the European Union is the InfoSoc Directive. It should be noted that the attempt to harmonise the EU with this directive was a very controversial issue, which led to considerable delays in the adoption and implementation of the directive itself, also due to the fact that the legislator did not have an easy time choosing and delimiting the scope of exceptions in a way that was acceptable to all member states. One only has to consider how in the time between the proposal of the Directive in 1997 and its introduction in 2001, the total number of permissible exceptions increased from seven to twenty-three[316].

This directive, in fact, includes, in Article 5, the main source of regulation of exceptions and limitations at EU level. The provisions included in article 5 apply to all copyright subject matter, except software and databases. Its first paragraph provides the only mandatory exception for member states, namely the exception for transient or incidental copying. Paragraphs 2 to 4 instead contain an exhaustive list of twenty optional exceptions for member states, which apply to the otherwise exclusive rights of reproduction, communication to the public and distribution[317].

Outside the InfoSoc Directive, other EU initiatives that provide for exceptions and limitations include the following: The Software and Database Directives, which provide for a number of different optional and mandatory exceptions; the Orphan Works Directive, which also includes a mandatory limitation on certain uses of orphan works (potentially subject to fair compensation); and the Marrakesh Treaty Directive, which contains a mandatory limitation on certain uses of accessible format copies for the benefit of persons

---

315 N. JONDET, *The text and data mining exception in the proposal for a directive on copyright: why the European Union needs to go further than the laws of member states*, in Propriétés Intellectuelles, no. 67 (April 2018), 32, citing Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market - COM(2016) 593 final, European Commission, 14 September 2016 [COM(2016) 593 final], recital 9.

316 L. GUIBAULT, G. WESTKAMP, T. RIEBER-MOHN, *Study on the implementation and effect in Member States' laws of Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the Information Society*, Report commissioned by the European Commission's Internal Market Directorate- General, in response to the invitation to tender MARKT/2005/07/D, Amsterdam Law School Legal Studies Research Paper No. 2012-28, Institute for Information Law Research Paper No. 2012-23, 2007, 39.

317 J. QUINTAIS, *The New Copyright in the Digital Single Market Directive: A Critical Look*, in European Intellectual Property Review 2020(1), 5.

who are blind, visually impaired or otherwise print impaired (with limited possibilities for applying compensation schemes)[318].

With regard to the InfoSoc Directive, which we said was the most important source for copyright exceptions, before reviewing some of the exceptions it introduced, it must be said that it has been widely criticised from the outset, not only in general as "*a badly drafted, compromise-ridden, ambiguous piece of legislation*"[319], but also specifically with regard to the exceptions introduced in Article 5. Indeed, regarding this part it has been criticised for its rather unclear rules and the uncertainty to which it has led among the legislators and the courts. The harmonising intent in the field of exceptions initially pursued by it has in fact been largely undermined for several reasons. First, by the fact that most of the provisions of the Directive are formulated in a broad and categorical manner, in particular because of the introduction of one set of broadly worded limitations, within the boundaries of which Member States could have chosen to legislate, and another set of general categories of situations for which Member States could have adopted limitations, thus leaving a great deal of discretion to the Member States, which have thus chosen to interpret the limitations contained in the Directive in the light of their own traditions. Second, the directive failed to arrive at a sufficient number of mandatory limitations, leaving member states completely free to choose from the list of optional exceptions the ones they liked best. The result of all this is a mosaic of exceptions and limitations that differs widely from member state to member state[320]. Also, Directive 2001/29 was criticised because its provisions did not take into account the technological developments and the potential of the Internet[321].

3.2.1. Assessing the potential applicability of the already existing exceptions and limitations in EU Intellectual Property law: exceptions to copyright and exceptions to the sui generis right in the database directive

The InfoSoc Directive, as mentioned above, provided for the inclusion of numerous exceptions to copyright, most of which optional. We are going to analyse below the compatibility of these exceptions with text and data mining activities in order to see whether they could, before the entry into force of the new Directive 790/2019, be considered applicable to TDM.

3.2.1.1. Temporary acts of reproduction

This exception, provided for in Article 5(1) of the InfoSoc Directive, is the only compulsory one among those introduced by the same directive, and has therefore been implemented by all EU member states. No similar exception has been provided for in the Database Directive, neither as regards copyright nor as regards the sui generis right, which means that it does not apply to copyright in databases[322].

---

318 *ibidem*.

319 E. ROSATI, *Copyright in the EU: In Search of (In)Flexibilities*, in Journal of Intellectual Property Law & Practice, Volume 9, Issue 7, July 2014, 1, citing HUGENHOLTZ, *Why the Copyright Directive is unimportant, and possibly invalid* (2000) 22(11) EIPR 499, 500.

320 L. GUIBAULT, G. WESTKAMP, T. RIEBER-MOHN, *Study on the implementation and effect in Member States' laws of Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the Information Society*, cit., 166.

321 A. TŪBAITĖ-STALAUSKIENĖ, *EU Copyright Law: Developing Exceptions and Limitations Systematically – An Analysis of Recent Legislative Proposals*, in Baltic Journal of Law & Politics 11:2 (2018), 162. Available at: https://sciendo.com/article/10.2478/bjlp-2018-0014

322 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 28.

It provides that:

> *Temporary acts of reproduction referred to in Article 2, which are transient or incidental [and] an integral and essential part of a technological process and whose sole purpose is to enable:*
> *(a) a transmission in a network between third parties by an intermediary, or*
> *(b) a lawful use*
> *of a work or other subject-matter to be made, and which have no independent economic significance, shall be exempted from the reproduction right provided for in Article 2*[323].

According to the CJEU, this exception was specifically designed to allow and guarantee the development of new technologies and, moreover, to strike a fair balance between the rights and interests of right holders, on the one hand, and of users of protected works wishing to make use of those works, on the other hand[324].

The rationale of this exception was therefore probably to mitigate the consequences of the broad definition provided by Article 2 of the InfoSoc Directive. We have seen that it has a very broad scope including "*temporary reproduction by any means and in any form*", making it necessary for the European legislator to allow certain copies that are part of a technological process, such as copies needed to allow browsing or caching[325].

Recital 9 of the proposed DSM directive also provides that this exception may be relied upon in limited cases of text and data mining that meet the numerous requirements of Article 5[326], namely the temporary reproductions transient or incidental to an integral and essential part of a technological process which enables a lawful use with no independent economic significance. In addition to the presence of all these different conditions for the application of this exception, which in the CJEU's view seem to have to be present cumulatively[327], the CJEU itself had previously held that, since this was an exception to the rights conferred by the directive, it would have to be interpreted restrictively[328]. Doubts therefore currently remain as to whether the acts of reproduction typical of TDM activities can be classified as transitory acts and as an integral and essential part of a technological process, falling within the category of lawful uses and without independent economic significance[329]. To find if this could be the case, it is necessary to analyse the individual requirements of the exception.

The making of transient copies is considered an essential part of a technological process when it is carried out entirely in the context of the making of a technological process, which therefore *a priori* excludes reproductions made (partially) outside of that process. Furthermore, it is also required that the technological process "*could not function correctly and*

---

323 Art. 5(1) InfoSoc Directive.
324 ECJ, 4 October 2011, joined Cases C-403/08 and C-429/08, Premier League, §164.
325 See Recital 33 InfoSoc Directive.
326 Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market, 14 September 2016, COM(2016) 593 final, 2016/0280 (Text with EEA relevance), recital 9.
327 CJEU, C-302/10, Infopaq International A/S v. Danske Dagblades Forening (17 January 2012) (Hereinafter Infopaq II) ECLI:EU:C:2012:16, §§ 26. Stating: "First, it must be borne in mind that those conditions are cumulative in the sense that non-compliance with any one of them will lead to the act of reproduction not being exempted, pursuant to Article 5(1) of Directive 2001/29, from the reproduction right provided for in Article 2 of that directive (Infopaq International, paragraph 55)".
328 ECJ 16 July 2009, Case C-5/08, Infopaq I, § 56. "For the interpretation of each of those conditions in turn, it should be borne in mind that, according to settled case-law, the provisions of a directive which derogate from a general principle established by that directive must be interpreted strictly (Case C-476/01 Kapper [2004] ECR I-5205, paragraph 72, and Case C-36/05 Commission v Spain [2006] ECR I-10313, paragraph 31)".
329 C. GEIGER, G. FROSIO, O. BULAYENKO, *Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU*, in C. SAIZ GARCÍA, R. EVANGELIO LLORCA, *Propiedad intelectual y mercado único digital europeo*, Valencia, Tirant lo blanch, 2019, Centre for Intellectual Property Studies (CEIPI) Research Paper no. 2019-08, 33.

efficiently without that act"[330]. According to the CJEU's interpretation, it is not necessary that no human activity is involved in the process[331], but those reproductions must be deleted automatically[332], since a manual deletion carries the risk that the copy will be kept longer[333], depriving it of its transitory nature. In this respect, the reproductions made in the computer memory by the software during the analysis phase could be considered transient and part of a technological process, since they are automatically removed from the RAM after the analysis and since they are indispensable to the analysis process. In the information retrieval phase and the creation of a target dataset, the reproductions may only be intended to exist temporarily, but the fact that they probably require human intervention to be removed make it highly unlikely that they could be covered by this exception. The same applies for the inclusion of works in TDM publications, which are permanent by nature[334].

In order to assess the applicability of this exception to TDM activities, it is necessary to ask ourselves whether TDM activities can be considered a lawful use of a work, and whether such lawful use is the sole purpose for which temporary copies of the copyrighted material are made. The directive itself briefly sets out in recital 33 a first meaning of lawful use: "*use should be considered lawful where it is authorised by the right holder or not restricted by law*"[335]. In the *Infopaq II* judgment, another meaning of lawful use can be derived. The CJEU in this case held that extracting 11 words from a newspaper article was a lawful use, even if not properly authorised by the rightholder[336].

The last requirement of the rule is that the transient copy should not have independent economic significance. This means that:

"*the economic advantage derived from their implementation must not be either distinct or separable from the economic advantage derived from the lawful use of the work concerned and it must not generate an additional economic advantage going beyond that derived from that use of the protected work*"[337].

The CJEU also requested in the same judgement that the works should not be modified during the process [338],which excludes from the application of the exception all those cases (most of them, since we have argued that the work of preparing the data is the most relevant part of the work) in which the contents are modified, such as the case in which metadata are added to the original material or unstructured texts are converted into structured data for the purpose of analysis[339].

Notwithstanding the positive element of the mandatory implementation, which makes this exception almost the same and present in all member states' legislations, however, if considered globally, the scope of the exception does not seem sufficiently broad and flexible to accommodate all the different TDM activities. Only the analysis step is likely to benefit from Article 5(1), which on the other hand excludes any TDM technology where information must be either retrieved from an external source or where contents must be modified before analysis[340].

---

330 CJEU 17 January 2012, C-302/10 (Infopaq II), §30.
331 *Ibidem*, §32.
332 *Ibidem*, §65.
333 *Ibidem*, §69.
334 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 28 – 29.
335 Recital 33 InfoSoc Directive.
336 CJEU Infopaq II, §44.
337 *Ibidem*, §50.
338 *Ibidem*, §54.
339 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 29.
340 *Ibidem*.

3.2.1.2. Research exceptions in the InfoSoc and Database Directives

Both the InfoSoc Directive and the Database Directive provide for exceptions to authors' rights for those uses whose purpose is scientific research.

Starting with the exception contained in the InfoSoc Directive, Article 5(3)(a) provides an exception to the reproduction and communication rights discussed above, providing that:

> *Member States may provide for exceptions or limitations to the rights provided for in Articles 2 and 3 in the following cases:*
> *(a) use for the sole purpose of illustration for teaching or scientific research, as long as the source, including the author's name, is indicated, unless this turns out to be impossible and to the extent justified by the non-commercial purpose to be achieved;*[341]

The "*purpose of illustration for teaching"* is not relevant for this work, as in TDM works are used to extract knowledge and patterns and not for illustration purposes in teaching activities[342]. Therefore, we will focus our attention here only on the case where materials are used for scientific research purposes.

Already from the literal wording of the rule, there are several limits to the application of this exception for TDM activities. First, the application of the exception is limited in scope to activities that have "*the sole purpose of illustration for teaching or scientific research"*, thus being inapplicable from the outset to activities that do not qualify as scientific research or that (partially) serve goals other than scientific research. In addition, it is necessary that such activities do not have, not even indirectly, a commercial purpose[343], as may be the case, for example, with any research (albeit scientific) carried out by a commercial company in order to develop a product to be launched on the market[344].

Furthermore, it only applies as long as the source is indicated, with the clarification that this is not necessary if it "*turns out to be impossible*", which perhaps makes its applicability to TDM easier. In fact, the amount of material to be mined makes citing the source effectively impossible most of the time[345]. Even this possible positive aspect is, however, entirely disputable, since what is materially extremely difficult does not necessarily qualify as "*impossible*" from an interpretative point of view.

Moreover, this exception is not mandatory but optional, which means that the choice of whether or not to implement the exception is left to the member states. As a result, not all EU Members have adopted the exception (such as Greece, the Netherlands and Spain) and, among the Member States, many have implemented it in national law with considerable differences in the wording of the provision[346]. The differences among the exceptions provided by the countries concern several aspects, including: the covered exclusive rights, the purpose and beneficiaries of the exception, the subject matter covered, the absolute or relative character of the attribution requirement, the different wording to describe the non-commercial requisite etc.[347].

---

341 Art. 5(3)(a) InfoSoc Directive.
342 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 32.
343 GEIGER, FROSIO, BULAYENKO, *Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU,* cit., 13.
344 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 32.
345 GEIGER, FROSIO, BULAYENKO, *Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU,* cit., 13.
346 M. BOTTIS, M. PAPADOPOULOS, C. ZAMPAKOLAS, P. GANATSIOU, *Text and Data Mining in the EU 'Acquis Communautaire' Tinkering with TDM & Digital Legal Deposit,* in Erasmus Law Review vol. 12 n. 2, 2019, 191.
347 For a full description of the differences between states in the implementation of this exception see: CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 30.

The final limitation is the uncertainty as to whether the applicability of the exception can be overcome contractually[348] and the extent to which the adoption of technological protection measures (TPMs) may cover uses that would normally be permitted under the exception[349].

The exactly same exception is provided for in the Database Directive, Article 6(2)(b). It applies not to the works contained in a database but to the database itself (more exactly, to its structure, or to its selection and arrangement)[350], providing that:

> *Member States shall have the option of providing for limitations on the rights set out in Article 5 in the following cases:*
> *(a) where there is use for the sole purpose of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved;*[351]

This exception, identical to its counterpart in the Copyright Directive (even if it seems to require that the source must be indicated in all cases, there being no safeguard clause in the provision) would apply, in the context of a TDM activity, when there is a reproduction of whole or substantial parts of an original database in its arrangement[352], on the other hand not being covered by Article 5 of the Database Directive the unauthorised use of an insubstantial part of a database.

The need to mention the source of all databases can be a heavy burden, as researchers either necessarily have to indicate the sources (and the list of sources for publication) or, in the impossibility of tracing the author of individual databases, have to abandon the idea of being able to copy the latter for their research project[353].

A further exception is provided for by the same Directive, in Article 9(b), but with reference to the sui generis right of extraction, providing:

> *Member States may stipulate that lawful users of a database which is made available to the public in whatever manner may, without the authorization of its maker, extract or re-utilize a substantial part of its contents:*

---

348 Only four countries have reported that derogations by contract are not allowed: Bulgaria, Hungary, UK and Iceland.

349 See CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 34. Indeed, as seen in the last part of the previous chapter, Article 6(1) of the Copyright Directive obliges Member States to provide legal protection against the circumvention of TPMs applied to works and other material covered by the Directive. This provision potentially restricts the user from certain acts that might otherwise be permitted by one of the exceptions to copyright. Indeed, Article 6(4) provides that, in the absence of voluntary measures by rightholders, Member States must take the necessary measures to ensure that users of works can benefit from the exceptions. In our case, the possibility of circumventing technological protection measures in order to benefit from the exception is allowed only in Slovakia, with the other countries generally providing for a right to request access, whether or not through an assigned committee or body, a process that can be quite cumbersome, especially in situations such as TDM, where the material used is large and often not provided by the same provider. The result of this is that right holders derogate from the research exception by physically or digitally restricting the acts necessary for TDM activities covered by a research exception.

350 TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* cit., 67.

351 Art. 6(2)(b) Database Directive.

352 GEIGER, FROSIO, BULAYENKO, *Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU,* cit., 14.

353 TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* cit., 71.

(b) in the case of extraction for the purposes of illustration for teaching or scientific research, as long as the source is indicated and to the extent justified by the non-commercial purpose to be achieved;[354]

As the exception applies, according to the rule, to those databases "*made available to the public in whatever manner*", the provision seems not to apply to those databases that have not been made public, such as an internal only database or a confidential one. Furthermore, if the database is not freely accessible, the user will be considered a "*lawful user*" when he can claim an authorisation obtained through a subscription[355].

However, this exception seems to be more restrictive than the previous ones, since first of all the exception is provided only with respect to the right of extraction, and not also to the right of reutilization. However, as we have seen at the beginning of the chapter, the right of extraction is, between the two rights that make up the sui generis rights discipline, the most relevant one for the purposes of TDM, since TDM may entail in most cases extractions in whole or in substantial part of the database content, but normally does not entail the reutilization of the same content. Moreover, here too the attribution requirement appears to be absolute, since the rule does not provide any derogation where the attribution "*turns out to be impossible*". On the other hand, the exception seems to be less restrictive as regards the purpose for which the exception is meant. Indeed, whereas for the Copyright Directive and its Database Directive counterpart, the exception allows the use of protected material for the "*sole*" purpose of illustration for teaching or scientific research, this exception omits the word "*sole*". The omission suggests that scientific research should not be the only purpose for which the exception is sought to be applied[356].

3.2.1.3. Private Use exception

Article 5 (2)(b) della InfoSoc Directive provides that:

Member States may provide for exceptions or limitations to the reproduction right provided for in Article 2 in the following cases:
(b) in respect of reproductions on any medium made by a natural person for private use and for ends that are neither directly nor indirectly commercial, on condition that the rightholders receive fair compensation which takes account of the application or non-application of technological measures referred to in Article 6 to the work or subject-matter concerned;

It has been noted that this exception might potentially cover some uses, in particular reproductions, made by individual researchers, especially in the case in which these users operates in Member States that do not have a research exemption, thus allowing reproductions done for research purposes[357].

However, it has two main limitations: the limit of the non-applicability to situations where there is even an indirect commercial use and the limit of the non-applicability to those situations (actually most of them) where the results of the work are not individual but used by a collective group of researchers. Furthermore, the rule requires that a fair remuneration be paid to the rightholder. In this regard, the Court of Justice considers that the fair

354 Art. 9(b) Database Directive.
355 TRIAILLE, DE MEEÛS D'ARGENTEUIL, DE FRANCQUEN, *Study on the legal framework of text and data mining (TDM),* cit., 80.
356 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 37.
357 GEIGER, FROSIO, BULAYENKO, *Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU,* cit., 35.

remuneration is to be calculated by reference to the harm caused to rightholders[358], which in the case of TDM does not theoretically exist at all, even if this issue is rather questionable and uncertain. This is without considering that even this exception is not mandatory but optional, which results in inconsistent national implementation, limited legal certainty and high transaction costs[359].

3.2.1.4. Normal use of a database

This exception is the only mandatory exception provided by the Database Directive, and is provided for in Article 6 (1):

> *"The performance by the lawful user of a database or of a copy thereof of any of the acts listed in Article 5 which is necessary for the purposes of access to the contents of the databases and normal use of the contents by the lawful user shall not require the authorization of the author of the database. Where the lawful user is authorized to use only part of the database, this provision shall apply only to that part".*

The so-called lawful users, can, thanks to this exception, reproduce a database without any authorisation of the rightholder if it is necessary to access the content of the database and make a normal use of it. The difficulties in applying this exception lie in the lack of interpretation as to the concept of "*normal use*" of a database, hence the lack of certainty as to the qualification of TDM activities as normal use and therefore, in the absence of a uniform interpretation of the concept of normal use, the danger of differing interpretations between EU member states increases, such as the legal certainty of users[360][361].

3.2.1.5. Extraction of insubstantial parts of a database

Finally, also Art. 8 (1) of the Database Directive was thought to be abstractly used as an exception (although not labelled as an exception) for TDM activities. It provides that:

> *"The maker of a database which is made available to the public in whatever manner may not prevent a lawful user of the database from extracting and/or re-utilizing insubstantial parts of its contents, evaluated qualitatively and/or quantitatively, for any purposes whatsoever. Where the lawful user is authorized to extract and/or re-utilize only part of the database, this paragraph shall apply only to that part".*

The most positive element of this provision is the possibility to extract and re-utilise insubstantial parts of the content of databases for any purpose. The evaluation of insubstantiality, as required by the provision, must be appreciated in terms of both quantity and quality. In this regard, the CJEU has held that the requirement of insubstantial parts must be assessed through the investment in the creation of the database and the harm that such extraction or reutilisation may cause to that investment[362].

---

[358] See CJEU, C-467/08, Padawan v SGAE (21 October 2008) ECLI:EU:C:2010:620.

[359] GEIGER, FROSIO, BULAYENKO, *Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU,* cit., 35.

[360] For example, we'll see how German national law has extended the notion of normal use of a database to include TDM in its copyright exception for TDM.

[361] GEIGER, FROSIO, BULAYENKO, *Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU,* cit., 36.

[362] CJEU, C-203/02, The British Horseracing Board Ltd and Others v. William Hill Organization Ltd (9 November 2004), § 69. CJUE states: "In that connection, it must be borne in mind that protection by the sui generis right covers databases whose creation required a substantial investment. Against that background,

With reference to TDM, it is therefore possible to argue that even the repeated and systematic extraction of insubstantial parts of the database, provided that the activity carried out on the database is not aimed at reconstructing all or substantial parts of it, thus not leading to a damage of the investment of the author of the database, should be considered lawful[363].

However, there are some limitations to the application of this provision as an exception: first, it only applies to lawful users performing TDM activity on databases, using specific TDM techniques that do not require making a copy of the material but performing crawling activities on the database. Furthermore, it may be contractually overridden[364].

3.2.1.6. Conclusions

The existing limitations in the InfoSoc Directive were not at the time, and do not appear to be now, broad enough to include in their provisions text and data mining, and neither the InfoSoc Directive nor the Database Directive contained an explicit ad hoc exception for text and data mining. In particular, we note that each of the directives has exceptions with more or less serious flaws, and none of the exceptions considered can offer a stable legal framework to engage in TDM research projects safely and to invest resources on it, thus leaving users willing to carry out such projects, whether they are private individuals or research organizations, in a position of extreme uncertainty. Particular problems result from the voluntary nature of the implementation of most of the single exceptions provided by the analysed directives. This makes it even less predictable whether existing exceptions and limitations can be applied to TDM projects, especially the cross-border ones[365].

3.2.2. The first TDM exception in Europe: United Kingdom

We are now going to take a look at some European legislations that introduced copyright exceptions for text and data mining "*before*" the specific exception introduced by the European Union DSM Directive and in the context of the yet existing InfoSoc and Database Directives. We will focus our attention mostly on the United Kingdom, first European country manifestly interested in the possibility of introducing a specific exception for TDM and that is being talked about a lot in the recent post-Brexit period.

If we consider the regulation of TDM in Europe, the UK is certainly to be considered the pioneer state in this field, as it was the first EU member state to introduce a specific copyright exception allowing text and data mining for non-commercial research[366].

Article 7(1) of the directive prohibits extraction and/or re-utilisation not only of the whole of a database protected by the sui generis right but also of a substantial part, evaluated qualitatively or quantitatively, of its contents. According to the 42nd recital of the preamble to the directive, that provision is intended to prevent a situation in which a user 'through his acts, causes significant detriment, evaluated qualitatively or quantitatively, to the investment'. It appears from that recital that the assessment, in qualitative terms, of whether the part at issue is substantial, must, like the assessment in quantitative terms, refer to the investment in the creation of the database and the prejudice caused to that investment by the act of extracting or re-utilising that part".

363 C. GEIGER, G. FROSIO, O. BULAYENKO, *Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU,* cit., 36.

364 *Ibidem*, 37.

365 *Ibidem*, 37.

366 E. ROSATI, *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, 8. Available at: https://op.europa.eu/en/publication-detail/-/publication/fdb4ecaa-20f1-11e8-ac73-01aa75ed71a1/language-en

UK copyright law, despite being a common law country, contains a "*closed-end*" list of copyright limitations, the fair dealing provisions, provided by Sections 28-31 of the *Copyright, Designs and Patents Act* of 1988. These provisions provide the basic structure and are then followed by a series of particular exemptions from copyright liability[367].

It was precisely the curiosity about the opportunity and possibility of adopting a regime similar to the American Fair Use system in place of the closed structure of exceptions in force that led, in November 2010, former Prime Minister David Cameron[368] to commission an independent report, the Hargreaves Review, (from Prof. Ian Harvgreaves), with the specific purpose of receiving answer to the question if the then current "*intellectual property framework might not be sufficiently well designed to promote innovation and growth in the UK economy*"[369] asking for consequent recommendations for reform.

The answer given by the professor's report to the first question was that fair use could certainly bring some benefits in terms of flexibility for copyright, but that fair use was only one aspect that could explain the success of US internet companies, attitudes towards business risk and investor culture being much more important factors. On the possibility of such a transplant, the review therefore concluded that it was not possible under the then existing EU rules[370][371].

Setting aside the first issue, the answer to the second question was that the intellectual property framework of the time, especially the part regarding copyright, was not adequate to support innovation and growth, and in particular research, especially in the medical field[372]. He therefore presented ten recommendations in the review. Among these recommendations, in the fifth point, Hargraves noted that there were at the time activities useful for research but prevented by copyright, and therefore recommended that the government should have explored the possibility of introducing exceptions in national legislation to try to make the best use of the already rather narrow exceptions offered by copyright in the European Union. Among these exceptions, great attention was paid to a possible exception for the activities

---

367 J. BOYLE, *(When) Is Copyright Reform Possible? Lessons from the Hargreaves Review*, in R. L. OKEDIJI, *Copyright Law in an Age of Limitations and Exceptions*, Cambridge University Press, 2017, 9.

368 Famous is what David Cameron said in November 2010, when he announced the Review of IP and Growth. Quoted in HARGRAVES, *Digital Opportunity - A Review of Intellectual Property and Growth,* cit., 44: "The founders of Google have said they could never have started their company in Britain. The service they provide depends on taking a snapshot of all the content on the internet at any one time and they feel our copyright system is not as friendly to this sort of innovation as it is in the United States. Over there, they have what are called "fair use" provisions, which some people believe gives companies more breathing space to create new products and services".

369 I. HARGRAVES, *Digital Opportunity - A Review of Intellectual Property and Growth, An Independent Report by Professor Ian Hargreaves*, May 2011, 1. Available at: https://www.gov.uk/government/publications/digital-opportunity-review-of-intellectual-property-and-growth

370 I. HARGRAVES, *Digital Opportunity - A Review of Intellectual Property and Growth,* cit., 46, paragraph 5.19. "The advice given to the Review by UK Government lawyers is that significant difficulties would arise in any attempt to transpose US style Fair Use into European law. It is against this background that the Review has stuck to its Terms of Reference and sought to isolate the particular benefits for economic growth that Fair Use exceptions provide in the US, with a view to understanding how these benefits can be most expeditiously obtained in the UK".

371 N. JONDET, *The text and data mining exception in the proposal for a directive on copyright: why the European Union needs to go further than the laws of member states*, in *Propriétés Intellectuelles*, no. 67, April 2018, 28.

372 I. HARGRAVES, *Digital Opportunity - A Review of Intellectual Property and Growth, An Independent Report by Professor Ian Hargreaves*, cit., 1. "We have found that the UK's intellectual property framework, especially with regard to copyright, is falling behind what is needed. Copyright, once the exclusive concern of authors and their publishers, is today preventing medical researchers studying data and text in pursuit of new treatments. Copying has become basic to numerous industrial processes, as well as to a burgeoning service economy based upon the internet. The UK cannot afford to let a legal framework designed around artists impede vigorous participation in these emerging business sectors".

of "t*ext and data analytics*". This activity, according to Hargreaves's conclusions, involved (and still involves) "*uses enabled by technology of works in ways which do not directly trade on the underlying creative and expressive purpose of the work*", recommending that this exception could not be overridden by contract, and that the UK should also promote and lobby the European Union for the introduction of this exception in the European copyright legislation[373].

The UK government responded positively to Hargreaves' recommendation, stating that it would pursue reform proposals for exceptions to UK copyright law, including, among others, a broad exception covering text and data mining activities carried out for non-commercial research purposes[374]. This last feature, depending on the narrow exceptions allowed by the European Union we have just analysed, did not allow to go further and provide for an exception also for research activities with commercial purposes. Indeed, the distinction between commercial and non-commercial research was only introduced in the UK by the implementation of the 2001 directive. Commentators have since then highlighted how difficult it can be to implement this distinction, and it was notably to avoid such difficulty that Hargreaves had recommended that the UK government should lobby the European institutions to ensure that any future TDM exception also applied to commercial research[375]. This initial response was followed by a further report from the government, in which it was noted that some publishers were taking "*an active role in developing text and data analytic technologies",* and that someone was already offering "*contracts that support the use of these technologies",* however, under the conditions prevailing at that time, *"research projects may in some cases require specific permissions from a large number of publishers in order to proceed*", considering this to be an "*insurmountable obstacle, preventing a potentially significant quantity of research from taking place at all*"[376]. The government, therefore, explicitly recognizing and quantifying in economic terms[377] the potential of TDM for research and the UK economy[378], and aware that it was unlikely that permitting mining for research would in itself "*negatively affect the market for or value of copyright works*"[379], proposed in the same report "*to amend the Copyright, Designs and Patents Act 1988*" so that should not be anymore "*an infringement of copyright for a person who already has a right to access a work (whether under a license or otherwise) to copy the work as part of a technological process of analysis and synthesis of the content of the work for the sole purpose of non-commercial research*"[380].

This proposed amendment finally resulted, on 1 June 2014, in the introduction of a specific exception for text and data mining in Section 29A of the *UK Copyright, Designs and*

---

373 I. HARGRAVES, *Digital Opportunity - A Review of Intellectual Property and Growth,* cit., 46-47.

374 HM Government. *The Government Response to the Hargreaves Review of Intellectual Property and Growth*. August 2011, 8. Available at: http://www.ipo.gov.uk/ipresponse-full.pdf

375 I. HARGRAVES, *Digital Opportunity - A Review of Intellectual Property and Growth,* cit., 48, paragraph 5.26.

376 HM Government, *Modernising copyright: a modern, robust and flexible framework. Government response to consultation on copyright exceptions and clarifying copyright law*, 2012. Available at: https://www.mpaonline.org.uk/wp-content/uploads/2017/09/Modernising_Copyright_-_a_modern_robust_and_flexible_framework_-_Government_response.pdf

377 See: UK IPO. 'I*mpact Assessment – Exception for Copying of Works for Use by Text and Data Analytics* (BIS0312) (Final Impact Assessement – Part 3)', 13 December 2012. http://www.legislation.gov.uk/ukia/2014/156/pdfs/ukia_20140156_en.pdf.

378 See HM Government, *Modernising copyright: a modern, robust and flexible framework. Government response to consultation on copyright exceptions and clarifying copyright law*, 37: "A more open market for the development of analytical technologies also has potential to offer opportunities to new and existing businesses, with potential knock-on impacts for growth. Evidence provided to the consultation suggests a permitted act would benefit researchers by £124m each year, and given the strength of the UK research base and the growing importance of data sharing/re-use this has a potential to be much higher if exploited effectively".

379 *Ibidem*, 37.

380 *Ibidem.*

Patents Act 1988[381], predicated on article 5(3)(a) of the Information Society Directive, which allows EU Member States to enact exceptions permitting the use of works 'for the sole purpose of illustration for teaching or scientific research [...] to the extent justified by the non-commercial purpose to be achieved'[382], worded as follows:

> *Copies for text and data analysis for non-commercial research*
>
> *(1) The making of a copy of a work by a person who has lawful access to the work does not infringe copyright in the work provided that—*
> *(a) the copy is made in order that a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non-commercial purpose, and*
> *(b) the copy is accompanied by a sufficient acknowledgement (unless this would be impossible for reasons of practicality or otherwise).*
>
> *(2) Where a copy of a work has been made under this section, copyright in the work is infringed if—*
> *(a) the copy is transferred to any other person, except where the transfer is authorised by the copyright owner, or*
> *(b) the copy is used for any purpose other than that mentioned in subsection (1)(a), except where the use is authorised by the copyright owner.*
>
> *(3) If a copy made under this section is subsequently dealt with—*
> *(a) it is to be treated as an infringing copy for the purposes of that dealing, and*
> *(b) if that dealing infringes copyright, it is to be treated as an infringing copy for all subsequent purposes.*
>
> *(4) In subsection (3) "dealt with" means sold or let for hire, or offered or exposed for sale or hire.*
>
> *(5) To the extent that a term of a contract purports to prevent or restrict the making of a copy which, by virtue of this section, would not infringe copyright, that term is unenforceable.*[383]

Of this exception, it can be said that it is certainly a rather broad one, since it applies to all types of content and all types of research, although, as mentioned, exclusively to non-commercial research. The term "*research*", in fact, in the absence of any dissenting jurisprudence, has always been interpreted very broadly, most believing it to apply to any person carrying out research, irrespective of his or her status or that of the institution with which he or she might be associated[384]. The application of this exception is also subject to three conditions: first, the person making the copy must have "*lawful access*" to the work in question; second, the computational analysis must be carried out for the sole purpose of non-commercial research; and third, the copy must be accompanied by a sufficient

[381] UK Copyright, Designs and Patents Act 1988. Available at: https://www.legislation.gov.uk/ukpga/1988/48/contents

[382] Y.H. LEE, *United Kingdom Copyright Decisions and Legislative Developments 2014*, In IIC International Review of Intellectual Property and Competition Law, 46 (2), 2015, 232.

[383] Section 29A of the UK Copyright, Designs and Patents Act 1988. Available at: https://www.legislation.gov.uk/ukpga/1988/48/section/29A

[384] N. JONDET, *The text and data mining exception in the proposal for a directive on copyright: why the European Union needs to go further than the laws of member states*, cit., 29.

acknowledgement, unless this would be impossible for reasons of practicality or otherwise[385]. The positive element of this exception is the fact that it cannot be restricted or overridden by contract, as explicitly provided for in the rule and as strongly advised by Hargreaves himself.

Following the UK's exit from the European Union, the UK government will not implement directive 790/2019. This means that at present only a very narrow TDM exception applies in the UK. As we have seen, however, UK is a pioneer state in TDM and already since the introduction of its exception in 2014 it held that restricting the exception to non-commercial research activities was the only choice available in the framework of the then current EU copyright law, but not the best one.

Even after its recent exit from the EU, the UK has not wasted any time and has once again shown itself to be highly committed to technological innovation and careful about balancing interests in the regulation of intellectual property[386]. In fact, on 7 September 2020, the UK Intellectual Property Office announced a "*Call for views on artificial intelligence and intellectual property*" [387], in order "*to understand the implications that Artificial Intelligence (AI) might have for Intellectual Property (IP) policy*" and, most importantly for us, "t*o understand the impact IP might have for AI, in the near to medium term*"[388]. The consultation lasted about two months and collected 92 responses, coming from a wide range of stakeholders, such as individual IP attorneys, trade bodies, industry associations, tech sectors, creative industries and other sectors.

From technology firms, entrepreneurs and researchers it emerged, on the one hand, the need to reform the current system of exceptions to copyright. It was argued that easy access to copyrighted works is essential for the development of the artificial intelligence system, calling for a reform of the system of exceptions to copyright, giving as examples to follow the American and Japanese systems, which we will address in the next chapter[389]. On the other hand, should be noted some resistance from copyright owners, who have expressed their firm opposition to the possibility of reforming the copyright exceptions to include commercial research activities within the existing exception[390].

---

385 Y.H. LEE, *United Kingdom Copyright Decisions and Legislative Developments 2014*, cit., 232.

386 This is also demonstrated by the words with which the UK IPO describes and justifies the consultation: "The UK is voted one of the best IP environments in the world. To keep it that way we are keen to look ahead to the challenges that new technologies bring. We need to make sure the UK's IP environment is adapted to accommodate them".

387 UK IPO, Artificial intelligence and intellectual property: call for views, 7 September 2020. Available at: https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views.

388 UK IPO, Artificial intelligence and intellectual property: call for views, cit.

389 "Many users of copyright material for AI, including technology firms, entrepreneurs and researchers, argued that easy access to works is crucial for teaching AI systems. They argued that such activities do not conflict with existing markets for work and, in fact, they may create new ones. While some already have licences in place in order to carry out TDM at volume, they also highlighted potential drawbacks of a licensing-only model". "A number referred to international approaches to TDM exceptions as they relate to AI – in particular the Japanese and EU data mining exceptions and US fair use provisions. They argued that these should be better understood in the UK, and that similar approaches may benefit innovation".

390 UK IPO, Consultation outcome - Government response to call for views on artificial intelligence and intellectual property, Updated 23 March 2021. Available at: https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views/government-response-to-call-for-views-on-artificial-intelligence-and-intellectual-property. "Many copyright owners argued that the current copyright exceptions do not apply to the use of copyright works in the AI machine learning process. They said that such use requires the express permission of right holders who should be able to determine whether to license their works for such use and set appropriate terms". Moreover, "A number of respondents gave views on text and data mining (TDM) as a methodology which can support AI in a variety of ways. Many copyright owners understood the rationale for the present copyright exception

The government, for its part, at the end of the consultation and in response to the numerous opinions received, undertook to examine the possibility of introducing into the UK copyright system "*improved licensing or copyright exceptions, to support innovation and research*"[391].

It is difficult to predict now which direction the UK will take after leaving the EU and it is also difficult to predict how the legislative autonomy achieved by the UK will affect the legislative decisions of the EU, which now finds itself facing not only the original competitors (e.g. US and Japan) but also the UK, all of which are inclined to more permissive TDM regimes.

Having in fact a complete autonomy and no European constraints anymore, in shaping its copyright law the United Kingdom could henceforth decide to adopt, like Japan, a very broad TDM exception for commercial research, or to remove the requirement of non-commerciality from its research-related exception, or even, in a more radical hypothesis made likely by the common belonging to the common law tradition, it could decide to adopt a US-style fair use system[392].

Some academics (see Jondet) had in fact hoped in the early stages of the EU directive's legislative process that, in drafting the EU TDM exception, the European institutions would consider the possible scenario where a post-Brexit Britain could lead and adopt a very broad TDM exception, even becoming an outpost of fair use in Europe[393]. However, as predicted by academic and MEP activist Julia Reda in a 2017 interview, this was not the case[394]. Nevertheless, in the situation that has arisen as a result of the so-called "*hard Brexit*", it is highly unlikely that the EU will not already have to reconsider (despite the recently enacted directive that has already addressed this issue) the need to counteract, also at a regulatory level, a potential competitive advantage for research and innovation from a state that is physically anchored within Europe and still strongly integrated into the European market, such as the United Kingdom.

3.2.3. Other successive TDM exceptions in Europe

Other European Union's member states have shown particular sensitivity to TDM issues, anticipating with their own legislative initiatives the entry into force of Directive 790/2019. Some examples are briefly given in the section below.

3.2.3.1. Ireland

---

which allows TDM for the purposes of non-commercial scientific research but expressed concerns about moving towards an exception that would allow commercial TDM. They believed that this could prejudice their legitimate interests and argued that licensing for TDM is already used across their sectors and can provide greater certainty. They argued that any new exception allowing more permissive use would shift the balance unfairly against creators".

[391] The Government, in particular, stated that: "[…] wants to better understand the merits of TDM exceptions in this area, including the approaches taken in other countries. The government remains committed to ensuring that a fair balance is struck between the needs of copyright owners and users. In view of this we will:
Review the ways in which copyright owners license works for use with AI, and consult on measures to make this easier, including improved licensing or copyright exceptions, to support innovation and research".

[392] R. ARNOLD, L. BENTLY, E. DERCLAYE, G. DINWOODIE, *IP Law Post-Brexit*, 101(2) Judicature 65 (2017), 69. Available at: https://judicature.duke.edu/articles/ip-law-post-brexit/

[393] See JONDET, *The text and data mining exception in the proposal for a directive on copyright: why the European Union needs to go further than the laws of member states*, cit., 34.

[394] See the interview of MEP Julia Reda, *Post-Brexit, UK could seize competitive edge in text and data mining,* ScienceBusiness, 23 February 2017. Available at: https://sciencebusiness.net/news/80154/Post-Brexit%2C-UK-could-seize-competitive-edge-in-text-and-data-mining

Among those who first in Europe identified the potential of text and data mining and at least studied the feasibility of a specific exception for text and data mining activities, we find Ireland, where in 2013 the Copyright Review Committee published a report in which was recommended the creation of a TDM exception for research[395]. In fact, starting from the realization that "*Copyright reform is in the air*" and citing other countries that where reforming their copyright law, the Report recommended a number of changes to the Irish copyright framework as embodied in the 2000 Copyright Act, among which even the proposal to introduce an Irish fair use exception[396]. In summer 2016, the government had prepared a bill[397] to reform copyright accordingly, but it was not submitted for discussion, seemingly because of the debate started at EU level by the Commission's directive proposal[398]. Another Bill to reform Irish copyright law[399] dates 2018, strange timing, it was noted, to introduce an exception for TDM, given that the 2019 Directive would come into force shortly thereafter and Ireland, as a member state, would have to implement the same Directive. Perhaps Ireland's move may be justified by the fact that it wished to guide the debate that existed at the time at the Council and European Parliament levels with regard to the purpose and benefits of a TDM exception at European level (The EU proposal would in fact be only available to *"research organisations"*), as well as by the need to ensure how that same exception, once introduced, could be adequately protected against contractual or technological override[400].

This proposal eventually culminated in law on 26 June 2019[401], and the Irish TDM exception provides:

> *53A. (1) Subject to subsection (3), the making of a copy of a work by a person who has lawful access to the work does not infringe copyright in the work where the copy is—*
> *(a) made in order that the person may carry out a computational analysis of anything in the work for the sole purpose of research for a non-commercial purpose, and*
> *(b) accompanied by a sufficient acknowledgement.*
>
> *(2) A copy made under subsection (1) of a work which was, at the time when the copy was made, available without a restriction as to its access does not infringe copyright, and whether or not that work continues to be so available after that time.*
>
> *(3) Where a copy of a work has been made under subsection (1) by a person, the copyright in the work is infringed where the copy—*
> *(a) is transferred to any other person, except where the transfer is authorised by the copyright owner, or*
> *(b) is used for any purpose other than the purpose referred to in subsection (1)(a).*

---

395 Irish Copyright Review Committee, '*Modernising Copyright*', Department of Jobs, Enterprise and Innovation, October 2013, 85-88/157. Available at: http://www.cearta.ie/wp-content/uploads/2013/10/CRC-Report.pdf

396 E. ROSATI, *Irish Government proposes introduction of Irish text and data mining exception*, The IPKat, March 2018. Available at: https://ipkitten.blogspot.com/2018/03/irish-government-proposes-introduction.html

397 Irish Government, '*General Scheme of a Copyright Bill Approved by Government*', Department of Business, Enterprise and Innovation, 4 August 2016.

398 JONDET, *The text and data mining exception in the proposal for a directive on copyright: why the European Union needs to go further than the laws of member states*, cit., 33.

399 Copyright and Other Intellectual Property Law Provisions Bill 2018 (Bill 31 of 2018). Available at: https://www.oireachtas.ie/en/bills/bill/2018/31/

400 E. ROSATI, *Irish Government proposes introduction of Irish text and data mining exception*, cit.

401 Ireland: Copyright and Other Intellectual Property Law Provisions Act 2019 (Number 19 of 2019). Available at: https://data.oireachtas.ie/ie/oireachtas/act/2019/19/eng/enacted/a1919.pdf

(4) Without prejudice to section 374, nothing in Part VII shall be construed as operating to prevent a person from undertaking an act permitted by this section.

(5) Without prejudice to the generality of section 52(1), where the publication of the results of a computational analysis referred to in subsection (1)(a) of a copy of a work includes the reproduction of extracts from the work, such inclusion shall constitute inclusion in an incidental manner referred to in section 52(1) if the extracts are not more than are reasonably necessary to explain, or to assist in explaining, the results of the analysis[402].

The exception is quite similar to the one formulated by the UK, as the Irish TDM exception would be available to persons who have legal access to a work solely for research purposes for a non-commercial purpose, and provided that sufficient acknowledgement is given. The central aspects of this exception are the limitation to non-commercial research, the beneficiaries of the exception, the use of copies made, and the safeguards against contractual and technological invalidation[403]. It is interesting here to note that, unlike the exception introduced by the UK, the Irish exception does not contain an explicit reference to the prohibition of contractual override, but this is because the Irish Copyright Act (section 2(10)) establishes a general prohibition of contractual override[404].

3.2.3.2. France

After a long legislative process[405] on 7 October 2016, a law for a digital republic (*"Loi pour une république numérique"*) was adopted, by which the Parliament amended[406] the Code de la propriété intellectuelle, introducing two narrow exceptions: a copyright exception for the *"extraction"* (in French *"exploration"*) of data and a sui generis right exception for the *"search"* (in French "*fouille*") of data[407]. They provide:

After the second paragraph of 9° of article L.122-5, a 10° is inserted as follows:

10° Electronic copies or reproductions realised from a legal original, for the purpose of text and data mining included or associated in a scientific publication for the needs of the public research, excluding commercial exploitation. A decree lays down the conditions in which text and data mining are employed, as well as the modalities of preservation and communication of the files produced at the end of the research activities for which they have been produced; these files constitute research data;

After the 4° of the article L.342-3 is inserted a 5°, thus written:

---

402 Text of the Irish TDM exception, Copyright and Other Intellectual Property Law Provisions Act 2019 - Number 19 of 2019, cit., point 14, 13.

403 For an in-depth explanation of all of these aspects, which we will not discuss here, see: ROSATI, *Irish Government proposes introduction of Irish text and data mining exception,* cit.

404 *Ibidem.*

405 For an accurate description of the troubled process that led to the introduction of the two exceptions in French legislation, see: N. JONDET, *The text and data mining exception in the proposal for a directive on copyright: why the European Union needs to go further than the laws of member states,* cit., 29 ss.

406 French Intellectual Property Code (Code de la Propriété Intellectuelle, CPI) AMENDEMENT N°180. https://www.assemblee-nationale.fr/14/amendements/3399/AN/180.asp

407 Art. 38 of the Law No. 2016-1231 for a Digital Republic added paragraph 10 to art. L.122-5 and paragraph 5 to art. L.342-3 of the French Intellectual Property Code (Code de la Propriété Intellectuelle, CPI).

5° Electronic copies or reproductions of a database realised by someone who has a legal access to it, for the purpose of text and data mining included or associated to scientific publications for the needs of a research activity, excluding commercial exploitation. The preservation and the communication of the technical copies made during the process, at the end of the research activities for which they have been produced, are provided by institutions appointed by decree. Other copies or reproductions are destroyed[408].

The TDM copyright exception is therefore intended to make possible those digital copies or reproductions made from a lawful source, for the mining of texts or those data included in or associated with scientific texts for public research purposes, while excluding any commercial activity[409]. However, the French legislator then chose to leave the regulation of the conditions under which TDM may be undertaken as well as the procedures for preserving and communicating research files created for the purposes of TDM to a subsequent actualisation decree[410].

What can be said about it is that it is an exception which, when compared with the counterpart TDM exception in the UK, is both broader and narrower in scope. The French approach is broader in the sense that it addresses both copyright and database rights[411]. At the same time, it is more limited than its UK counterpart in that it applies only for the purpose of public research (copyright exception) or in the context of a research (database right exception), limited only to "*texts*" even of a non-scientific nature and "*data included in or associated with scientific texts*" and excluding, therefore, other works such as pictures, musical or audio-visual works[412][413].

It is interesting to note that the French legislative reform finally concluded and materialised with the introduction of a pure exception, in spite of the fact that the previous legislative process, long and rather conflictual, had recorded the repetition of authoritative legal opinions which on the one hand indicated the option of the exception as not available on the basis of the legal framework in force at European level, while, on the other hand, indicated that the so-called contractual option was preferable, i.e. the option that, leaving aside all regulatory/legislative changes both at national and European level, appealed to the possibility of regulating, on an exclusively negotiated basis, any issue concerning the possible interferences between TDM and copyright. An emblematic example of this strong and contrasting line of thought is the report on TDM presented in July 2014 by the Conseil supérieur de la propriété littéraire et artistique (CSPLA), the advisory body in charge of advising the Ministry of Culture on copyright issues, which came up with recommendations completely opposite to the corresponding UK recommendations of the Hargreaves Review. In fact, it manifested a transparent and underlying concern to ensure maximum protection

[408] Unofficial translation provided by: BOTTIS, PAPADOPOULOS, ZAMPAKOLAS, GANATSIOU, *Text and Data Mining in the EU 'Acquis Communautaire' Tinkering with TDM & Digital Legal Deposit,* cit., 197.
[409] JONDET, *The text and data mining exception in the proposal for a directive on copyright: why the European Union needs to go further than the laws of member states*, cit., 31.
[410] BOTTIS, PAPADOPOULOS, ZAMPAKOLAS, GANATSIOU, *Text and Data Mining in the EU 'Acquis Communautaire' Tinkering with TDM & Digital Legal Deposit,* cit., 197.
[411] CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 70.
[412] JONDET, *The text and data mining exception in the proposal for a directive on copyright: why the European Union needs to go further than the laws of member states*, cit., 31.
[413] For a very good overview of the French TDM provisions as well as a comparative analysis with the UK and US see: L. MAUREL, *L'exception TDM dans la loi numérique: mérites, limites et perspectives*, S.I.Lex, 11 November 2016.

for the reasons of copyright with respect to the conflicting reasons of TDM, indicated as nothing less than a "*parasitic*" practice[414][415].

The solution adopted at the end of this tormented debate sought to respond to the need, felt at the parliamentary level following several requests from research and innovation circles, to preserve France from a possible condition of competitive disadvantage in terms of research not only with respect to U.S. researchers but also with respect to the United Kingdom, a nation that had shortly before adopted an effective exception intended, as we have seen, to significantly stimulate the possibilities of exploitation of TDM in terms of scientific research. France's adoption of a text and data mining exception has indeed put an end to the so-called *"contractual"* option as a possible solution to the copyright issues raised by text and data mining. However, the hostile legal-cultural background in which this reform came to fruition subsequently proved to be a serious obstacle to the concrete implementation of the new provisions. In fact, two years after the promulgation of the law, the government decrees for its successful implementation had not yet been issued, so that the arrival of the broader and more compulsory exceptions introduced by the 2019 EU Directive made the French national provisions on TDM substantially outdated, obsolete and subject to inevitable further revisions[416].

3.2.3.3. Estonia

In Estonia, a country that has since long been particularly sensitive to technological innovation[417], attention to the topic of text and data mining was already noticeable back in 2012, when A. Kelli, H. Pisuke and A. Tarvast published a study[418] aimed at understanding whether the development of digital language resources (special databases that consist of and contain many written and oral texts) could be used for the purpose of developing and training machine translation and other language tools and if their use could be compatible with the then Estonian copyright law. The Kelli's study concluded that the Estonian Copyright Act (1992) already had some exceptions, in particular the quotation exception and the research exception (Section 19), which could be applied and could cover scientific research activities in the field of linguistics[419]. Nevertheless, in order to make the law clearer and more certain, and to avoid creating unnecessary obstacles to the use of TDM arising from uncertainty as to the actual legality of TDM and the applicable law, it was considered appropriate to add a

414 J. MARTIN, L. DE CARVALHO, *Rapport de la mission sur l'exploration de données* («*Text and Data mining*»), Conseil Supérieur de la Propriété Littéraire et Artistique, July 2014 [Rapport CSPLA]. Available at: http://www.culturecommunication.gouv.fr/content/download/105451/1236408/version/1/file/Rapport%20Text%20and%20Data%20Mining%20(exploration%20de%20donn%C3%A9es).pdf

415 JONDET, *The text and data mining exception in the proposal for a directive on copyright: why the European Union needs to go further than the laws of member states*, cit., 30.

416 *Ibid.,* 31.

417 Proof of this is the emergence of several very successful technology start-ups, first and foremost Skype. See on this topic: "Estonian president delights in country's high proportion of unicorns", The Guardian. Availbale at: https://www.theguardian.com/world/2018/jun/29/estonia-unicorns-president-kersti-kaljulaid-delight. For a brief explanation of why Estonia is such a technologically advanced country, see: *How Estonia became Europe's tech hotspot,* The New Economy. Available at: https://www.theneweconomy.com/technology/how-estonia-became-europes-tech-hotspot.

418 A. KELLI, A. TAVAST, H. PISUKE, *Copyright and Constitutional Aspects of Digital Language Resources: The Estonian Approach*, in Juridica International 2012 Vol 19. Available at: https://www.researchgate.net/publication/288558711_Copyright_and_constitutional_aspects_of_digital_language_resources_The_estonian_approach

419 *Ibid.*, 48.

specific exception. Indeed, in 2014, the draft Copyright and Related Rights Act was issued[420], which introduced an exception for data mining and text analysis, worded as follows:

> *Reproduction and processing of an object of rights for the purpose of text analysis and data mining, on the condition of attributing the name of the author of the used work, the name of the work and the source of publication, except if such attribution is impossible, and on the condition that such use is not carried out for commercial purposes*[421].

The Estonian legislature, thus, amended the country's Copyright Act of 1992[422] and, in January 2017, introduced TMD in paragraph 3 of Article 19 titled "*Free use of works for scientific, educational, informational and judicial purposes*". Thus, the Estonian Copyright Act include now the following provision:

> *The following is permitted without the authorisation of the author and without payment of remuneration if mention is made of the name of the author of the work, if it appears thereon, the name of the work and the source publication:*
> *3) processing of an object of rights for the purposes of text and data mining and provided that such use does not have a commercial objective*;

It is noticeable that Estonia has taken the rule introduced by the United Kingdom as a model[423]. It is therefore not necessary to elaborate further on this exception.

3.2.3.4. Germany

Finally, a brief mention can be made of what happened in Germany. In January 2017, the Federal Ministry of Justice in Germany issued a draft for a bill to "*Align Copyright Law with the Current Demands of the Knowledge Based Society*"[424]. As a result of this a new law on copyright in the knowledge economy, the Urheberrechts-Wissensgesellschafts-Gesetz[425], was adopted by the Bundestag in September 2017. Eventually, it entered into force the 1st March 2018 with the initial intention that it should only remain valid for a limited period of five years. At the end of this period the legislator would have to decide whether to extend its validity or replace it with different rules, following the likely adoption of a new Directive on the Digital Single Market. This law introduced a specific exception for TDM activities in Section 60d, titled *"Text and data mining"*, with the following text:

> *(1) In order to enable the automatic analysis of large numbers of works (source material) for scientific research, it shall be permissible*

---

420 Autoriõiguse ja autoriõigusega kaasnevate õiguste seaduse eelnõu (The Estonian draft Copyright and Related Rights Act), 2014.

421 A. KELLI, *The conceptual bases for codifying Estonia's IP law and the main legislative changes: From the comparative approach to embedding drafted law into the socio-economic context*, in International Comparative Jurisprudence 1 (2015).

422 Estonian Copyright Act 1992. Available in English at: https://www.riigiteataja.ee/en/eli/510062014003/consolide

423 BOTTIS, PAPADOPOULOS, ZAMPAKOLAS, GANATSIOU, *Text and Data Mining in the EU 'Acquis Communautaire' Tinkering with TDM & Digital Legal Deposit,* cit., 197.

424 The text of the draft proposal is available at (Text in German): https://www.bmjv.de/SharedDocs/Gesetzgebungsverfahren/Dokumente/RefE_UrhWissG.pdf;jsessionid=717C31AA231CB9A6E1BE65058D11699C.1_cid289?__blob=publicationFile&v=1

425 Act on Copyright and Related Rights (Urheberrechtsgesetz, UrhG) as last amended by Article 1 of the Act of 1 September 2017 (Federal Law Gazette I p. 3346), The Federal Ministry of Justice and Consumer Protection [as translated by Ute Reusch]. http://www.gesetze-im-internet.de/englisch_urhg/ , Section 60d.

1. to reproduce the source material, including automatically and systematically, in order to create, particularly by means of normalisation, structuring and categorisation, a corpus which can be analysed and
2. to make the corpus available to the public for a specifically limited circle of persons for their joint scientific research, as well as to individual third persons for the purpose of monitoring the quality of scientific research.
In such cases, the user may only pursue non-commercial purposes.
(2) If database works are used pursuant to subsection (1), this shall constitute customary use in accordance with section 55a sentence 1. If insubstantial parts of databases are used pursuant to subsection (1), this shall be deemed consistent with the normal utilisation of the database and with the legitimate interests of the producer of the database within the meaning of section 87b (1) sentence 2 and section 87e.
(3) Once the research work has been completed, the corpus and the reproductions of the source material shall be deleted; they may no longer be made available to the public. It shall, however, be permissible to transmit the corpus and the reproductions of the source material to the institutions referred to in sections 60e and 60f for the purpose of long-term storage.

Also here, therefore, the exception is restricted exclusively to cases of TDM carried out for non-commercial purposes, and currently covers those acts of reproduction of databases protected under copyright or sui generis database rights necessary to build a corpus for TDM purposes, reproductions that according to the rule are to be considered "*normal use of database*". According to some, this exception has been constructed having in mind mainly TDM activities carried out with text, not data, in the context of academic research. According to the rule, at the end of the TDM project all copies and reproductions must be deleted. Exceptions to this rule are those acts of making available the corpus of materials produced by TDM activity (e.g. source materials that were normalised, structured and categorised) to a specifically limited circle of persons for their joint scientific research, as well as to individual third persons for the purpose of monitoring the quality of scientific research, plus the case of sending the *"corpus"* to institutions designated by law (e.g. libraries, archives, museums or educational establishment) for long-term storage[426]. A positive element is, also here, the non-overridability of the rule by contracts between the parties[427].

Interestingly, on the other hand, the new law requires that flat-rate equitable remuneration need to be paid to a copyright collecting society for the allowed uses[428]. A remuneration right for TDM is currently not found in the TDM exceptions in the UK and France.

3.3. Harmonization in EU through Directive DSM 790/2019

At EU level, the topic of text and data mining has been the subject of interest since the beginning of the last decade. In particular, there have been two types of approaches to regulating TDM activities. The first approach, which turned out to be unsuccessful, focused

426 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 71.
427 Act on Copyright and Related Rights (Urheberrechtsgesetz, UrhG), Section 60g "Lawfully permitted use and contractually authorised use": "(1) The rightholder may not invoke agreements which restrict or prohibit uses permitted in accordance with sections 60a to 60f and such restriction or prohibition is to the detriment of the persons entitled to such use".
428 P. KAMOCKI, E. KETZAN, J. WILDGANS, A. WITT, *New exceptions for Text and Data Mining and their possible impact on the CLARIN infrastructure*, Selected papers from the CLARIN Annual Conference 2018, Linköping Electronic Conference Proceedings 159, 68.

mainly on the use of licences; the second approach, which was eventually followed, was a specific exception approach.

3.3.1. The licence approach: "Licences for Europe"

Starting from the first approach, at the end of 2012 the Commission, on the initiative of then President Barroso, held a Debate on content in the digital economy[429], at the end of which two different parallel tracks of action were announced. A first track of action, called "*Immediate issues for action*", was the launch of a structured stakeholder dialogue that should have been held already at the beginning of 2013, and was designed to seek solutions to six issues, including text and data mining, for which it was considered necessary to make rapid progress. With particular regard to TDM, the aim was to investigate the potentials and limitations of innovative licensing and technological solutions. On the basis of this stakeholder dialogue, by the end of 2013, the outcome should have been summarised and an attempt made to provide effective market-led solutions to the issues, without, however, prejudicing the possibility of intervention by public policy, including legislative reform.

The second track of action instead, "*Medium term issues for decision-making in 2014*", was intended to provide relevant market studies, impact assessments and legal drafting works that could support the eventual decision to propose legislative reform of the copyright system in the European Union. Among the topics covered by these studies was also the appropriateness of the levels of harmonisation, limitations and exceptions to copyright in the digital age.

Following this decision, the Commission announced, less than two weeks later, the opening of the Stakeholder Dialogue[430], the "*Licenses for Europe*"[431]. composed of four different working groups, including a working group dedicated to the promotion of an efficient use of text and data mining (TDM), limited however to scientific research purposes (WG4). The aim of the latter was to "*identify the scale of demand for TDM access at EU level for text mining of scientific publications and underlying data for research purposes, and appropriate means of meeting this demand*", paying particular attention to exploring "*the potential and possible limits of standard licensing models, as well as assess the appropriateness and feasibility of technology platforms to facilitate TDM access*".

The participants of these working groups were hands-on representatives of rights holders, licensing bodies, commercial and non-commercial users of protected content, as well as internet end-users[432]. The work would have started at the beginning of 2013 in an initial plenary meeting, would have continued in a plenary at mid-term (June) and finally in the fourth quarter of 2013.

The first meeting, held on 4 February 2013, seemed to have remained committed to the initial objective, namely, according to the Chair, "*the need to stay within the scope agreed in the*

[429] European Commission. Commission agrees way forward for modernising copyright in the digital economy, MEMO/12/950. Brussels: 05.12.2012. Available at: http://europa.eu/rapid/press-release_MEMO-12-950_en.htm

[430] European Commission. Communication from the Commission on content in the Digital Single Market, COM(2012) 789 final. Brussels: 18.12.2012. Available at: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52012DC0789

[431] European Commission (2013). "Licenses For Europe: Structured stakeholder dialogue 2013". Available at: https://ec.europa.eu/licences-for-europe-dialogue/

[432] Among the 37 organizations invited to participate, relevant participants were: IBM, Microsoft, Reed Elsevier, STM (International Association of Scientific, Technical & Medical Publishers), SAS, Computers and Communications Industry Association, Communia, LIBER, National Centre for Text Mining, University of Manchester. For a complete list of the organizations invited to participate see: https://ec.europa.eu/licences-for-europe-dialogue/sites/licences-for-europe-dialogue/files/131213_wg4-list-of-participants.pdf

Commission's December Communication and to discuss problems on the ground and practical solutions (not the interpretation of, or changes to, the existing legal framework)"[433].

However, this position already changed at the second meeting, in which the Chair pointed out that "*even though the focus of the discussion should be on practical market-based solutions, participants may of course discuss other preferred options, including legislative solutions*". This was probably due to the fact, as also indicated by the Chair, that several researchers and research institutions, including representatives of libraries pointed out, expressing this opinion also in a letter sent by some of the participants on 26 February 2013[434], "*that mining of data and texts to which a lawful access had been obtained should not be subject to a separate licence or separate payment",* moreover arguing that the *"right to read should be the right to mine"*. The representatives of libraries explained that *"they were not opposed to market driven solutions. However, they also expressed the view that an exception for text and data mining was preferred"*[435].

Notwithstanding the reassurances on the still not excluded legislative intervention by the representatives of the Commission, manifested in the third meeting, held on 22 April 2013, in particular by The Internal Market and Services Directorate General (DG MARKT) [436] and the Commission's Directorate-General for Research and Innovation, the academic community, supported by European Technology SMEs and the Open Access Publishers, unhappy with the adopted approach (focused exclusively on licensing rather than on

[433] Licences for Europe, Working Group 4 – Text and data mining, Conclusions of the First meeting of 4 February 2013. Available at: https://ec.europa.eu/licences-for-europe-dialogue/en/content/wg4-conclusions-meetings.html

[434] See: Letter sent by concerned participants in response to the "Licences for Europe – A Stakeholder Dialogue" workshop on text and data mining for scientific research purposes, 26.02.2013. Available at: https://libereurope.eu/document/licences-for-europe-a-stakeholder-dialogue-text-and-data-mining-for-scientific-research-purposes-working-group/. "We write to express our serious and deep-felt concerns in regards to Working Group 4 on text and data mining (TDM). Despite the title, it appears the research and technology communities have been presented not with a stakeholder dialogue, but a process with an already predetermined outcome –namely that additional licensing is the only solution to the problems being faced by those wishing to undertake TDM of content to which they already have lawful access. Such an outcome places European researchers and technology companies at a serious disadvantage compared to those located in the United States and Asia".

[435] Licences for Europe, Working Group 4 – Text and data mining, Conclusions of the Second meeting of 8 March 2013. Available at: https://ec.europa.eu/licences-for-europe-dialogue/en/content/wg4-conclusions-meetings.html

[436] The Internal Market and Services Directorate General is one of the Directorates General and specialised services which make up the European Commission. Its main role is to coordinate the Commission's policy on the European Single Market and to seek the removal of unjustified obstacles to trade, in particular in the field of services and financial markets.

statutory exceptions), largely[437] withdrew[438] from the process[439], thus determining the substantial failure of the project of TDM regulation through licensing[440].

3.3.2. Road to the DSM Directive: preliminary works for the introduction of a TDM exception in European Copyright Law

In the end, therefore, the latter approach for specific exceptions was preferred to the former. This is probably mainly due to the fact that, looking at the international landscape, the most followed approach to regulate TDM was either the US "*fair use*" doctrine, or the introduction of statutory exceptions, as in the case of e.g. Japan and other non-European countries. Moreover, we have seen that the solution for specific exceptions was also the one most desired by most stakeholders other than publishers. Finally, some member states, such as the UK, were already considering at the time the introduction of a TDM exception within the legal framework of exceptions offered by the EU copyright law (e.g., Article 5.3 (a) of the InfoSoc Directive)[441].

The European Commission, after having repeatedly expressed the need to reform and to harmonise copyright law, in particular with a view to establishing a legal framework which would meet the requirements of the digital era appropriately, on December 2013, having managed to digest the failure of the "*Licenses for Europe*" project (at least as far as TDM was concerned), launched a public consultation on the review of EU copyright rules[442]. The main aim for the European legislator was to ensure that the system of rights, limitations and enforcement remained appropriate and was adapted to the new environment, in particular

---

437 Those who withdrew from the project were: The Association of European Research Libraries (LIBER), The Coalition for a Digital Economy, European Bureau of Library Information and Documentation Associations (EBLIDA), The Open Knowledge Foundation, Communia, Ubiquity Press Ltd., Trans-Atlantic Consumer Dialogue, National Centre for Text Mining, University of Manchester, European Network for Copyright in support of Education and Science (ENCES), Jisc.

438 Letter from European Technology SMEs, Open Access Publishers and the Research Sector WG 4, May 22nd, 2013. Available at: https://libereurope.eu/wp-content/uploads/2013/05/Letter_of_withdrawalL4E_TDM_May-24_4.pdf
See also: LIBER (Association of European Research Libraries) (2013). "Stakeholders representing the research sector, SMEs and open access publishers withdraw from Licences for Europe". Available at: https://libereurope.eu/article/stakeholders-representing-the-research-sector-smes-and-open-access-publishers-withdraw-from-licences-for-europe-2/

439 KAMOCKI, KETZAN, WILDGANS, WITT, *New exceptions for Text and Data Mining and their possible impact on the CLARIN infrastructure*, cit., 67.

440 In particular, those who had decided to withdraw their adhesion to the project wrote in the above-mentioned letter: "We believe that any meaningful engagement on the legal framework within which data driven innovation exists must, as a point of centrality, address the issue of limitations and exceptions. Having placed licensing as the central pillar of the discussion, the "Licences for Europe" Working Group has not made this focused evaluation possible. Instead, the dialogue on limitations and exceptions is only taking place through the refracted lens of licensing. This incorrectly presupposes that additional relicensing of already licensed content (i.e. double licensing) – and by implication also licensing of the open internet– is the solution to the rapid adoption of TDM technology. […] We are concerned, therefore, that our participation in a discussion that focuses primarily on proprietary licenses could be used to imply that our sectors accept the notion of double licensing of as a solution. It is not. We firmly believe that "the right to read is the right to mine". […] Furthermore, we would point to the urgent need to be competitive with the United States and the high-tech economies in Japan and South Korea, where legal barriers to TDM are far lower precisely because of the existence of copyright limitations and exceptions there".

441 KAMOCKI, KETZAN, WILDGANS, WITT, *New exceptions for Text and Data Mining and their possible impact on the CLARIN infrastructure*, cit., 67.

442 *Public Consultation on the Review of the EU Copyright Rules,* 2013. Available at: https://digital-strategy.ec.europa.eu/en/library/modernisation-eu-copyright-rules-useful-documents

"*ensuring that the EU copyright regulatory framework*" stayed "*fit for purpose in the digital environment to support creation and innovation, tap the full potential of the Single Market, foster growth and investment in our economy and promote cultural diversity*"[443].

This consultation, which lasted until March 2014, was addressed, in the form of specific questions, to different stakeholders, including end users, institutional users, rightsholders, industries, collective management organisations, public authorities, Member States and others. The stakeholders were asked questions regarding access to content libraries and archives, teaching, research, disabilities, text and data mining[444], user-generated content; questions whose specific purpose was to identify difficulties arising from the fact that limitations and exceptions were mainly optional, to determine whether more harmonisation was needed, and whether it was necessary to make limitations and exceptions mandatory, in order to add to or remove some of them from the existing list[445].

This consultation generated more than 9500 replies, and the report on this consultation was published by the European Commission in July 2014[446].

The first group of questions related in general to the system of limitations and exceptions in European copyright law. Stakeholders' opinions on the subject diverged considerably. For example, end users argued that the optional nature of limitations and exceptions was a source of legal uncertainty and an uneven playing field for market participants, and that the then existing system of exceptions to copyright made copyright law more difficult to understand and apply for users. They also expressed the wish for existing exceptions and limitations to be preserved and wanted to introduce new ones, focusing on the need for at least the exceptions relating to fundamental rights to be mandatory and harmonised. They also argued that at the moment there was not a sufficient balance between the rights of right holders and exceptions and wished to transform exceptions into "*user rights*" and not to allow contracts to override exceptions and to prevent right holders from limiting the use of exceptions by using technological protection measures.

On the same wavelength were the institutional users (as well as some member states and most representatives of academia, civil society or think-tanks), who argued that the exceptions and limitations should be maintained and made mandatory and harmonised. Some of these went as far as to propose the introduction of an open-ended norm, which would complement the list of exceptions in Directive 2001/29. Some also expressed the need to introduce a specific exception for text and data mining or that the research exception should also extend to commercial research.

On the other hand, there were authors and performers, who, due to the 2001 InfoSoc Directive still in force at the time, which we saw in the last chapter was strongly biased in their favour, were rather satisfied and did not see the need to change the list of exceptions and limitations, being also against the idea of further harmonisation or inclusion of new exceptions, considering in particular that the list of exceptions was "*very broad, flexible and fit*

---

443 *Public Consultation on the Review of the EU Copyright Rules,* cit., 2.

444 *Ibidem*, 27.

445 TŪBAITĖ-STALAUSKIENĖ, *EU Copyright Law: Developing Exceptions and Limitations Systematically - An Analysis of Recent Legislative Proposals*, cit., 162.

446 European Commission - Directorate General Internal Market and Services*, Report on the responses to the Public Consultation on the Review of the EU Copyright Rules,* July 2014. Available at: https://digital-strategy.ec.europa.eu/en/library/modernisation-eu-copyright-rules-useful-documents

for purpose" and that it ensured "*a balance between property rights and the public interest*"[447]. Instead of legislation, they advocated the greater speed and efficiency of licensing[448].

A second group of questions related specifically to whether the European Union's system of copyright limitations and exceptions should be made more flexible[449].

Again, predictably, on the one hand the end users and institutional users (supported by the member states and part of the academic community) felt that more flexibility was needed to ensure that copyright exceptions could adapt to technological change and were forward-looking, bringing up TDM as an example, and arguing that the lack of flexibility in the list of exceptions could have the consequence of putting Europe at an unfortunate competitive disadvantage, when compared to other countries where copyright law provides an open ended norm such as the "*fair use*" defence. They therefore suggested adding an open-ended norm to the limited list of exceptions to allow for uses that could not be foreseen at the time the legislation was adopted. Further, some considered that the status of exceptions in EU copyright law should be strengthened and that the *"three-step test"* is often given too strict an interpretation by courts. Some criticised the InfoSoc Directive arguing that it was not drafted in a technologically neutral manner and that this was problematic in times of accelerated technological progress.

On the other hand, however, the rightholders, all agreed that the EU legal framework provided sufficient flexibility with respect to exceptions, arguing that the optional nature of the exceptions allowed member states to adapt the copyright system in accordance with their national traditions and cultural policy, and that the introduction of an open-ended norm would have reduced the level of harmonisation and lowered the level of legal certainty in Europe, and saying that they were particularly against the introduction of an open norm similar to the "*fair use*" principle in the US, according to them not in line with European legal traditions, traditionally based on statutory law and not on (as in the US) judge-made law, whose case law of the previous 200 years justified the existence of the principle in the US.

Concerning the specific questions relating to text and data mining, according to which the stakeholders were supposed to identify possible problems and possibly propose a solution, here too the pattern seems to be very similar to that identified for the previous questions, as a matter of course.

End-users, consumers and institutional users (also supported by service providers, software companies in particular), who were mainly researchers, argued that Europe was missing out on the benefits that text and data mining can bring to competitiveness and innovation, and ultimately to citizens. They identified as obstacles to TDM the legal uncertainty about whether and how copyright could apply to text and data mining and the problems with existing licensing mechanisms, which they considered inadequate. Some also argued that TDM should not be relevant to copyright, since it does not concern the expression of an idea that copyright law intends to protect, but only the analysis of the underlying facts. Others pointed out that the reproduction of copyright works for non-commercial research based on text and data mining might already be covered by existing exceptions and limitations to copyright and database law in the laws of the Member States,

---

[447] European Commission - Directorate General Internal Market and Services*, Report on the responses to the Public Consultation on the Review of the EU Copyright Rules*, cit., 31. They stated that "*current framework is the result of a sustainable compromise which has to be preserved, to ensure in particular legal certainty and a stable and comprehensive framework for all stakeholders. Some note that new and/or broader exceptions could result in businesses using copyright protected works and performances without remunerating rightholders ('free-riding')*".

[448] TŪBAITĖ-STALAUSKIENĖ, *EU Copyright Law: Developing Exceptions and Limitations Systematically - An Analysis of Recent Legislative Proposals*, cit., 163.

[449] European Commission - Directorate General Internal Market and Services*, Report on the responses to the Public Consultation on the Review of the EU Copyright Rules*, cit., 33.

but noted that in many Member States it was still unclear whether these exceptions, in particular the research exception (when implemented), could apply to text and data mining. They argued that licences were not the best solution to solve the problem of uncertainty about TDM, but rather could be an additional barrier and source of transaction costs, since a good TDM project, in order to obtain the material needed to compose a corpus, would require a huge amount of contractual arrangements (with the additional inconvenience of disclosing information about one's own projects), thus limiting the number of articles that can be mined and interfering with the way researchers can make the output of the mining available. They pointed out in their responses that TDM was significantly easier to perform in non-EU countries that contained *"fair use"* provisions in their legal systems, with the result that North American universities could claim a competitive advantage over EU-based universities and companies. They were also concerned about the use of technological protection measures that block access to content, thus preventing text and data mining or making it more difficult, also disagreeing with publishers' concerns about reduced performance and security issues related to their infrastructure when crawled by mining robots. They therefore suggested to the Commission that text and data mining should not be subject to licensing, but rather called for legislative intervention to introduce a specific mandatory exception for text and data mining in EU copyright law. The exception should cover both commercial and non-commercial scientific research, as limiting it to non-commercial uses would have created legal uncertainty and prevent the full development of TDM's potential. In their view, technological protection measures and contracts should not be allowed to override the exception. These respondents also felt that researchers should have the right to share extraction results with other researchers, provided that these results are not substitutable for the original works that were extracted.

On the other hand, some authors, such as journalists and writers and their representatives, as well as CMOs, argued that there were no major problems in the field of text and data mining and that the best approach to the issue was to continue with licensing solutions developed through dialogue between stakeholders and between right holders and governments to improve licensing practices, essentially believing that it was premature to address text and data mining in legislation, considering the fact that it was a rather recent activity and characterised by great uncertainty about even the meaning to be attributed to the term TDM. Although they generally expressed their opposition to a specific exception, they nevertheless expressed their preferences as to the characteristics of a possible exception: first, it should be limited to non-commercial uses, as they considered an exception covering commercial uses to be contrary to the EU's international obligations, and that it could favour commercial operators, in particular news aggregators or commercial news monitoring services. They stressed that it was essential that the output of text and data mining did not become a substitutable product for the original works that were mined. Some respondents also stressed the role that collective management could play in this area and some suggested that if an exception is introduced, it should be linked to the payment of fair compensation to right holders. The introduction of a remuneration fee is also suggested as an alternative by some.

Along the same sentiment were the publishers, who also opposed the introduction of a TDM exception, justifying this opposition by arguing that there was no evidence of market failure. In particular, Science, technology and medical (STM) publishers pointed out that they were already fulfilling the requests of those who asked to do TDM activities (although still rather limited but growing) and that they already had solutions to allow TDM, in particular through licences, granted under standard terms and at no cost to researchers who wanted to mine subscription-based content for the purposes of non-commercial scientific research. An

exception, according to them, would have damaged the licensing offers that publisher were developing and would not have solved the other issues other than copyright raised by text and data mining, such as the protection of data privacy and technical aspects which require the intervention and investments by publishers (e.g. to set up a specific technical environment, such as dedicated platforms from which researchers may download the content before mining it). They also expressed concern about the danger of an increase in abuses and piracy and about the risk of damage to databases and infrastructure hosting their content when they are crawled by mining robots.

Finally, the member states (apart from a few who opposed legislative innovations, which they considered premature, placing more trust in the licences that were being developed at the time), for their part, recognised the benefits that TDM could offer to scientific research and expressed their consent to the opportunity to examine proposals, including legislation, through the introduction of a specific exception, if based on clear evidence. One member state in particular was concerned that European researchers should not be at a competitive disadvantage vis-à-vis other international players.

Jean Claude Juncker himself, on 15 July 2014, newly elected President of the European Commission, included, among his "*Political Guidelines for the next European Commission*", the need to make "*better use of the great opportunities offered by digital technologies*"[450]. This was to be achieved also through the modernisation of the European Copyright in the light of the "*digital revolution and changed consumer behaviour*". The then President of the Commission thus undertook to take "*ambitious legislative steps towards a connected digital single market*" within the first six months of his mandate[451].

On the basis of the above mentioned consultation and the resulting report, the European Commission, building on its renewed drive for innovation, prepared the "*Digital Single Market strategy*", which was presented in a Commission Communication in May 2015[452]. One of the objectives of this strategy was to reduce differences between national copyright legislations[453]. With specific regard to TDM, the Commission argued in the same communication at the same point that

---

[450] In a speech before the European Parliament on 22 October 2014 Jean-Claude Juncker, announcing the first legislative initiatives of the Juncker Commission, said: "*Every day, Europe is losing out by not unlocking the great potential of our huge digital single market. Jobs that should be there are not being created. Ideas – the DNA of Europe's economy! – do not materialise to the extent they should. Let us change this for the better*". *Setting Europe in Motion: President-elect Juncker's Main Messages from his speech before the European Parliament*, Statement in the European Parliament plenary session ahead of the vote on the College, Strasbourg, 22 October 2014, 26. Available at: https://www.eesc.europa.eu/resources/docs/jean-claude-juncker---political-guidelines.pdf.

[451] JEAN-CLAUDE JUNCKER, *A New Start for Europe: My Agenda for Jobs, Growth, Fairness and Democratic Change*, Political Guidelines for the next European Commission, 14 July 2014, 5. Available at https://www.eesc.europa.eu/resources/docs/jean-claude-juncker---political-guidelines.pdf. "*I believe that we must make much better use of the great opportunities offered by digital technologies, which know no borders. To do so, we will need to have the courage to break down national silos in telecoms regulation, in copyright and data protection legislation, in the management of radio waves and in the application of competition law. […] I intend to take, within the first six months of my mandate, ambitious legislative steps towards a connected digital single market, notably by swiftly concluding negotiations on common European data protection rules; by adding more ambition to the ongoing reform of our telecoms rules; by modernising copyright rules in the light of the digital revolution and changed consumer behaviour; and by modernising and simplifying consumer rules for online and digital purchases. This should go hand-in-hand with efforts to boost digital skills and learning across society and to facilitate the creation of innovative start-ups. Enhancing the use of digital technologies and online services should become a horizontal policy, covering all sectors of the economy and of the public sector*".

[452] *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A Digital Single Market Strategy for Europe*, COM(2015) 192 final, 6.5.2015. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52015DC0192

[453] See *2.4. Better access to digital content - A modern, more European copyright framework*, in *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A Digital Single Market Strategy for Europe*, cit., 6.

"Innovation in research for both non-commercial and commercial purposes, based on the use of text and data mining (e.g. copying of text and datasets in search of significant correlations or occurrences) may be hampered because of an unclear legal framework and divergent approaches at national level. The need for greater legal certainty to enable researchers and educational institutions to make wider use of copyright-protected material, including across borders, so that they can benefit from the potential of these technologies and from cross-border collaboration will be assessed, as with all parts of the copyright proposals in the light of its impact on all interested parties"[454].

The Commission therefore made the commitment to present by the end of 2015 a legislative proposal that could lead to a reduction in the differences between national copyright regimes and allow wider online access to works by users across the EU, including through further harmonisation measures. The proposal would have included, among others[455], harmonised exceptions that would have provided greater legal certainty for cross-border use of content for specific purposes (e.g. research, education, text and data mining, etc.).

The first concrete step in the direction of Copyright modernisation and harmonisation was taken in December 2015, when the European Commission presented an "*Action Plan for the modernisation of EU copyright rules*"[456]. The Commission, arguing that "*EU copyright rules need to be adapted so that all market players and citizens can seize the opportunities of this new environment*" and aiming at the realisation of "*a more European framework*" necessary "*to overcome fragmentation and frictions within a functioning single market*", intended with this communication to explain how this objective would be achieved in practice, containing it targeted actions with proposals for the very short term, a set of proposals planned for 2016, and a long-term vision, building them on the previous public consultation in 2013-2014, taking into account the opinions of the European Parliament expressed in its resolution on the implementation of the Directive on Copyright in the Information Society and the conclusions of the European Council meeting of 25-26 June 2015.

The Commission noted here that the area of copyright exceptions was particularly fragmented in the EU, mainly due to the non-mandatory nature of most of the exceptions, pointing out that this had the consequence that some states contained exceptions that others did not and that some states had broader exceptions while others had narrower ones. In addition, noted that some exceptions had to be revised in the light of technological progress over the years, which was causing problems particularly for those exceptions applicable to education and research. This, the Commission argued, was particularly evident in the case of text and data mining, where the lack of a clear provision for scientific research was creating uncertainty among the research community, thus damaging the "*EU's competitiveness and scientific leadership at a time when research and innovation (R&I) activities within the EU must increasingly take place through cross-border and cross-discipline collaboration and on a larger scale […]*".

In light of this, The Commission would accordingly consider options to address the problem and in particular consider a legislative proposal for a specific exception for TDM

---

454 *Ibid.*

455 The Commission mentions for example: portability of legally acquired content; ensuring cross-border access to legally purchased online services while respecting the value of rights in the audiovisual sector; clarifying the rules on the activities of intermediaries in relation to copyright-protected content; modernising enforcement of intellectual property rights, focusing on commercial-scale infringements (the 'follow the money' approach) as well as its cross-border applicability.

456 *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Towards a modern, more European copyright framework*, COM(2015) 626 final, 2.12.2015. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2015%3A626%3AFIN

by spring 2016, to be introduced in order to "*allow public interest research organisations to carry out text and data mining of content they have lawful access to, with full legal certainty, for scientific research purpose*".

In September 2016, in line with the Digital Single Market strategy, the European Commission presented the legislative package[457] of two Directives[458] and two regulations[459] for the modernisation of EU copyright law[460].

The legislative document we are interested in is the first directive, the Digital Single Market Directive, which was designed to "*bring the EU framework on exceptions up to speed with digital uses, in certain key areas like education, research and access to knowledge and are particularly focused on cross-border uses, thus contributing to further deepening the single market*" by introducing "*new mandatory exceptions in the areas of education, research, and preservation of cultural heritage*". Among these is the important introduction of a "*new mandatory exception for text and data mining carried out for the purposes of scientific research*", which would have allowed the "*research organisations to use text and data mining technologies in full legal certainty*" and would "*prevent different approaches being followed in different Member States in an area, like research, where cross-border cooperation on a large scale and cross-discipline collaboration is more and more frequent".* This, according to the Commission, would have supported *"scientific advancement and innovation in the EU*".

3.3.3. The legislative procedure of the DSM Directive: the initial proposal, successive amendments and the adoption

Let us therefore now proceed to examine the complicated and lengthy ordinary legislative procedure that finally led to the adoption of the two new TDM exceptions provided for in Articles 3 and 4 of Directive 790/2019[461]. The process followed the ordinary legislative procedure, which places the European Parliament and the Council of the European Union on an equal footing, so that a separate process of evaluation of the proposal took place within each of the two institutions[462]. Thus, the Commission's proposal for a Directive was followed by proposals from the Parliament and the Council.

The Copyright in the Digital Single Market Directive is one of the longest in the copyright *acquis*, being composed of 86 recitals and 32 articles. It is divided into five titles: general provisions (I), measures to adapt exceptions and limitations to the digital and cross-

457 *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Promoting a fair, efficient and competitive European copyright-based economy in the Digital Single Market*, COM(2016)592. Available at: https://digital-strategy.ec.europa.eu/en/library/promoting-fair-efficient-and-competitive-european-copyright-based-economy-digital-single-market

458 *Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market*, COM(2016) 593, and *Proposal for a Directive of the European Parliament and of the Council on certain permitted uses of works and other subject-matter protected by copyright and related rights for the benefit of persons who are blind, visually impaired or otherwise print disabled and amending Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society*, COM(2016) 596.

459 *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,* COM(2016) 594, and *Proposal for a Regulation of the European Parliament and of the Council on the cross-border exchange between the Union and third countries of accessible format copies of certain works and other subject-matter protected by copyright and related rights for the benefit of persons who are blind, visually impaired or otherwise print disabled,* COM(2016) 595.

460 TŪBAITĖ-STALAUSKIENĖ, *EU Copyright Law: Developing Exceptions and Limitations Systematically - An Analysis of Recent Legislative Proposals*, cit., 163.

461 For a complete chronological account of the legislative process and related documents see: CREATe, EU Copyright Reform – Evidence on the Copyright in the Digital Single Market Directive. Available at: https://www.create.ac.uk/policy-responses/eu-copyright-reform/#ert_pane4-0

462 QUINTAIS, T*he New Copyright in the Digital Single Market Directive: A Critical Look*, cit., 2.

border environment (II), measures to improve licensing practices and ensure wider access to content (III), measures to achieve a well-functioning marketplace for copyright (IV), and final provisions (V)[463].

In the next section we will focus our attention mainly on the events affecting Title II on exceptions and limitations, which is of most interest to us, and which contains the two new exceptions, leaving aside the (even more complicated and discussed) issues relating to the subsequent titles.

3.3.3.1. The (limited) Commission's proposal

In accordance with the ordinary legislative process of the European Union[464], the first Proposal for a Directive on copyright in the Digital Single Market was published on 14 September 2016 by the European Commission, presented by Günther Oettinger shortly before he left his position as Digital Commissioner. The proposal was referred to the European Parliament Committee on Legal Affairs (JURI), which on 12 October 2016 appointed Therese Comodini Cachia (EPP,Malta) as rapporteur[465]. At that time the Proposal had only one article which was devoted to Text and Data Mining, article 3, and read as follows:

Article 3:

1. *Member States shall provide for an exception to the rights provided for in Article 2 of Directive 2001/29/EC, Articles 5(a) and 7(1) of Directive 96/9/EC and Article 11(1) of this Directive for reproductions and extractions made by research organisations in order to carry out text and data mining of works or other subject- matter to which they have lawful access for the purposes of scientific research.*
2. *Any contractual provision contrary to the exception provided for in paragraph 1 shall be unenforceable.*
3. *Rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject-matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.*
4. *Member States shall encourage rightholders and research organisations to define commonly-agreed best practices concerning the application of the measures referred to in paragraph 3.*

The European Commission's draft DSM Directive proposal was somewhat ambiguous on the issue. In fact, although it seemed to have recognised the value of TDM and its potential for "*gain new knowledge and discover new trends*" in recital 8, not only for the scientific community but also for innovation, it provided in the proposal merely a mandatory TDM exception for the benefit of non-commercial research organisations. The exception, thus, only benefitted those entities, such as universities or research institutes, which act in the public interest, as highlighted in recital 11 of the draft directive[466]. However, it had the

463 *Ibid.*, 6.

464 For an overview about the EU Ordinary Legislative Procedure see: EU Parliament, The Ordinary Legislative Procedure. Available at: https://www.europarl.europa.eu/olp/en/ordinary-legislative-procedure/overview

465 For an explanation of the role of the "rapporteur" within the European Union see: European Parliament, A "rapporteur" - the person who presents reports to Parliament. Available at: https://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+IM-PRESS+20060725STO09938+0+DOC+XML+V0//EN

466 Recital 11: "Research organisations across the Union encompass a wide variety of entities the primary goal of which is to conduct scientific research or to do so together with the provision of educational services. Due

positive characteristic of being applicable to any type of research activity, including commercial research, as specified in recital 10 of the proposal, which stated that "*research organisations should also benefit from the exception when they engage into public-private partnerships*"[467]. Moreover, its application could not be excluded by contract. Recital 13 also provided that there was no need "*to provide for compensation for rightholders*" because "*in view of the nature and scope of the exception the harm should be minimal*", the interests of rightholders already being guaranteed by the requirement of lawful access to the content, which mainly allowed academic publishers to have a revenue stream through subscriptions. Finally, the proposal foresaw that rightholders could apply proportionate measures to ensure the security and integrity of their networks and databases[468].

In retrospect, the doctrine, commenting on this first version of the exception, has argued that by restricting the scope of the exception in such a way, i.e. by only authorising the research organizations to search texts and data "*on works or other subject-matter to which they have lawful access for the purposes of scientific research*", which seemed to exclude a large part of online research since the legality of the sources would be uncertain, the Commission risked creating an ineffective and rapidly obsolete provision, in particular with regard to the development of artificial intelligence, but also with regards to other activities of essential research and innovation not conducted by public bodies[469].

3.3.3.2. The Committee stage and the Commodini Draft Report: a bold proposal

After this first proposal by the Commission, the parliament came into play and could thus formulate its first amendments to the Commission's proposal for a directive.

to the diversity of such entities, it is important to have a common understanding of the beneficiaries of the exception. Despite different legal forms and structures, research organisations across Member States generally have in common that they act either on a not for profit basis or in the context of a public-interest mission recognised by the State. Such a public-interest mission may, for example, be reflected through public funding or through provisions in national laws or public contracts. At the same time, organisations upon which commercial undertakings have a decisive influence allowing them to exercise control because of structural situations such as their quality of shareholders or members, which may result in preferential access to the results of the research, should not be considered research organisations for the purposes of this Directive".

[467] Recital 10: "This legal uncertainty should be addressed by providing for a mandatory exception to the right of reproduction and also to the right to prevent extraction from a database. The new exception should be without prejudice to the existing mandatory exception on temporary acts of reproduction laid down in Article 5(1) of Directive 2001/29, which should continue to apply to text and data mining techniques which do not involve the making of copies going beyond the scope of that exception. Research organisations should also benefit from the exception when they engage into public-private partnerships".

[468] JONDET, *The text and data mining exception in the proposal for a directive on copyright: why the European Union needs to go further than the laws of member states*, cit., 32.

[469] C. GEIGER, *The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the failed Text and Data Mining Limitations in the CSDM Directive*, PIJIP/TLS Research Paper Series 2021, No. 66, 7.

Since the proposal of the directive, especially during the parliamentary phase, several criticisms were made of the proposal by the academic community[470][471].

Getting back to the beginning of the parliamentary phase, it began with the so-called committee[472] stage, in which four different committees proposed some Draft reports with amendments. These committees were, in particular: the Culture and Education CULT, Internal Market IMCO, Industry, Research and Energy ITRE, the Legal Affairs JURI (which was the lead committee), with a fifth committee LIBE (Civil Liberties, Justice and Home Affairs). They submitted their amendments to the proposal in February 2017[473], and it has been said that, at this phase alone, the MEPs submitted more than 1000 amendments[474].

This exchange of views between committees thus brought the so-called Committee process to completion, culminating in the so-called "*compromise*" text drafted by the JURI rapporteur. It, the "*MEP Comodini's Draft Report*"[475] was submitted by the JURI Committee on 10 March 2017, in the following version:

Article 3

Text and data mining

1. Member States shall provide for an exception to the rights provided for in Article 2 of Directive 2001/29/EC, Articles 5(a) and 7(1) of Directive 96/9/EC and Article 11(1) of this Directive for reproductions and extractions to be made by a person who has

---

[470] See in this regard C. GEIGER, G. FROSIO, O. BULAYENKO, *The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects*, Study for the Directorate-General for Internal Policies of the Union, Department of Citizens' Rights and Constitutional Affairs, European Parliament, February 2018. For a critical evaluation of the directive proposal, see also C. GEIGER, G. FROSIO, O. BULAYENKO, *Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?*, IIC 2018, Vol. 49, No. 7, 814, and from the same authors: *The EU Commission's Proposal to Reform Copyright Limitations: A Good but Far Too Timid Step in the Right Direction*, EIPR 2018, Vol. 40, p. 4; *European Copyright Society, General Opinion on the EU Copyright Reform Package*, 24 January 2017, p. 5; R. M. HILTY, H. RICHTER, in R. M. HILTY, V. MOSCON, *Modernisation of the EU Copyright Rules, Position Statement*, MPI for Innovation and Competition Research Paper No. 17-12, p. 25 et sq.; N. JONDET, *L'exception pour le data mining dans le projet de directive sur le droit d'auteur - Pourquoi l'Union européenne doit aller plus loin que les legislations des Etats membres*, Propr. intell. 2018, No. 67, p. 25; E. ROSATI, *An EU text and data mining exception: will it deliver what the Digital Single Market Strategy promised?*, The IPKat, 22 May 2017.

[471] These criticisms seem to have been the impetus that subsequently led to the circle of beneficiaries of the exception to been extended to "cultural heritage institutions", which according to recital 13 of the directive includes mainly libraries, museums and archives. In addition, the possibility to store works for search and mining purposes has been added. Finally, a new exception has been introduced in the body of the directive, according to which "Member States shall provide for an exception or limitation to the rights (...) for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining".

[472] For an overview of the role of committees and the difference between Lead Committee and Committee for Opinion, see: *What role for the European Parliament's committees and how do they work?* Available at: https://epthinktank.eu/2020/05/06/what-role-for-the-european-parliaments-committees-and-how-do-they-work/

[473] For an overview of the amendments presented by the Committees see: T. VOLLMER, *The wandering saga of the text and data mining exception in the EU copyright reform*, 5 June 2017. Available at: https://www.communia-association.org/2017/06/05/wandering-saga-text-data-mining-exception-eu-copyright-reform/

[474] E. ROSATI, M. KRETCHMER, *EU copyright reform: quo vadis*?, in The IPKat, 11 May 2017. Available at: https://ipkitten.blogspot.com/2017/05/eu-copyright-reform-quo-vadis.html

[475] Draft Report on the proposal for a directive of the European Parliament and of the Council on copyright in the Digital Single Market (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)). Committee on Legal Affairs Rapporteur: Therese Comodini Cachia. Available at: https://www.europarl.europa.eu/doceo/document/JURI-PR-601094_EN.pdf?redirect

lawful access to works and other subject-matter, provided that reproduction or extraction is used for the sole purpose of text and data mining.

1a. Member States shall provide for rightholders who market works or other subject-matter primarily for research purposes, to have an obligation to allow research organisations not having lawful access to those works or other subject- matter access to datasets that enable them to carry out only text and data mining. Member States may also provide for rightholders to have a right to request compensation for meeting this obligation as long as that compensation is related to the cost of formatting these datasets.

2. Any contractual provision contrary to the exception provided for in paragraph 1 shall be unenforceable.

3. Rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject-matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.

4. Member States shall encourage rightholders and research organisations to define commonly-agreed best practices concerning the application of the measures referred to in paragraph 3.

4a. Member States shall designate a facility to store datasets used in research by text and data mining technologies securely and to make such datasets accessible only for verification purposes.

This first amendment to the draft directive was welcomed by the academic and scientific community[476] as a great step forward compared to the rather narrow exception initially formulated by the Commission. The proposal, to be submitted to the Parliament for approval, envisaged the extension of the benefits of the exception, no longer limited to "*research organisations*", but extended to "*reproductions and extractions to be made by a person who has lawful access to works and other subject-matter, provided that reproduction or extraction is used for the sole purpose of text and data mining*", also by no longer restricting the purposes of TDM to *"scientific research"*, but by extending the scope of the exception to all possible and different uses of text and data mining.

In addition, it proposed an amendment obliging publishers to include a mechanism for users who would otherwise not have legal access to the corpus of works to be able to undertake TDM on the publisher's content, possibly after paying a fee to those publishers. Finally, it also wanted to order Member States to establish facilities to securely store research datasets to ensure the accessibility and verifiability of research made possible through TDM.

A short period of time later, in June 2017, Axel Voss replaced Terese Comodini Cachia as rapporteur for the Copyright Directive. Although Voss initially reassured everyone about the fact that he would have continued to work from the existing draft version of the bill, some noticed some important diversions on opinions between the two MEPs[477]. At this point, however, the parliament's work on the issue slowed down considerably. Indeed, after several successive postponements since July 2017, and a further postponement towards the end of March 2018, the JURI Committee further postponed their vote on agreeing the

[476] See Communia Association, *JURI rapporteur proposes to fix most egregious flaws of the copyright reform proposal*, 20 March 2017. Available at: https://www.communia-association.org/2017/03/20/lead-mep-proposes-fixes-egregious-flaws-copyright-reform-proposal/. See also: T. VOLLMER, *European Parliament legal affairs committee pushes for strong exception for text and data mining,* Communia, 21 March 2017. Available at: https://www.communia-association.org/2017/03/21/european-parliament-legal-affairs-committee-pushes-strong-exception-text-data-mining/

[477] *New lead MEP could shift talks on contentious copyright bill*, Euractiv. Available at: https://www.euractiv.com/section/digital/news/new-lead-mep-could-shift-talks-on-contentious-copyright-bill/

European Parliament's final view and amendments to the existing draft of the proposed Copyright Directive. The meeting for the JURI vote was therefore set for 20/21 June 2018.

3.3.3.3. The Council amendments: the introduction of a new broader optional exception

In the meantime, in parallel, the Council was also working on its proposal and considering amendments to the first Commission Proposal. After an initial proposal for revision under the Maltese Presidency, which added "*cultural heritage institutions"* to the list of beneficiaries of the exception alongside "*research organisations*"[478], Estonia took over the presidency of the Council of the European Union for the second half of 2017. After some initial revisions of the initial proposal[479], all of which left unaffected the presence of the Commission's obligatory TDM exception that would have applied to research organisations for purposes of scientific research, to which was added the recommendation that the beneficiaries originally contemplated by the Commission would need to be expanded to include cultural heritage institutions, an additional and optional exception was also introduced in the last compromise proposal of his term of mandate in Article 3[480]:

> *(5) Member States may provide for an exception or a limitation […] for temporary reproductions and extractions of works and other subject-matter that form an integral part of the process of text and data mining, provided that the works and other subject-matter are accessed lawfully and that the use of the works or other subject-matter for text and data mining is not expressly reserved by the rightholder.*

This additional exception, which was supposed to apply to beneficiaries other than research organisations, and for uses other than scientific research, was, on the one hand, welcomed as it extended the scope of the exception, but on the other hand, criticised for being limited in scope, as it would only cover "*temporary reproductions and extractions*", and only if the rightsholder did not prohibit it. The commentators wondered in particular why an exception for temporary acts of reproduction should be included, when there was already an exception, and moreover the only mandatory one, the exception for temporary acts of reproduction in Article 5(1) of Directive 2001/29/EC, which could cover such acts. Moreover, they challenged the reasonableness of such an exception, which was not only optional for member states, but could also be overcome and avoided either contractually or through technological protection measures, when the rightholders had reserved the right to make reproductions and extractions for text and data mining[481].

Following the change of Presidency at the Council of the European Union, in which the outgoing Estonia was succeeded by Bulgaria for the next six months (January - June

---

478 Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market - Presidency compromise proposal regarding Articles 2 to 9 (Malta Presidency), 8 May 2017. Available at: https://data.consilium.europa.eu/doc/document/ST-8929-2017-INIT/en/pdf

479 Occurred on 30 August, 26 September, 30 October and 29 November respectively.

480 Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market - Presidency compromise proposal (consolidated version) and state of play, 13 December 2017. Available at: https://data.consilium.europa.eu/doc/document/ST-15651-2017-INIT/en/pdf

481 T. VOLLMER, *The Estonian Presidency's new (and optional) TDM exception: small potatoes*, Communia Association, 4 December 2017. Available at: https://www.communia-association.org/2017/12/04/estonian-presidencys-new-optional-tdm-exception-small-potatoes/

2018), the Proposal for a Directive on Copyright in the Digital Single Market was discussed by the Permanent Representatives Committee[482] on 31 January 2018.

The 23 of March 2018, the Bulgarian Presidency presented its first compromise proposal, which was nearly identical to the previous one, maintaining the changes already offered by the earlier Estonian plan and the Commission's obligatory TDM exception that would apply to research organisations (including cultural heritage institutions) for purposes of scientific research. Moreover, also this Proposal included the additional and optional exception in Article 3 for temporary reproductions and extractions. This additional exception would apply to beneficiaries other than research organisations, and for uses other than scientific research. But those acts would be limited in that they only would cover temporary reproductions and extractions, and only if the rightsholder does not prohibit it[483].

Towards the finalisation of the discussions in the Council, the Presidency, in order to obtain a negotiating mandate in view of the inter-institutional negotiations with Parliament, scheduled a COREPER meeting before the JURI vote, and it was brought forward to 27 April 2018, which was preceded by a preparatory meeting of the Attaches, scheduled for 16 April 2018.

On 12 April, the Presidency's questions in relation to Articles 3, 11 and 13 were published with the aim of seeking views from delegates and asking for feedback on possible amendments, stating:

"Building upon on the discussions up to date, we consider that in order to reach an agreement in COREPER, the remaining issues that need to be discussed are Articles 3a (optional TDM exception), 11 (publishers right) and 13 ("value gap")".

The Council Presidency pointed out that there was still no unanimous opinion among the delegations on the proposed legislation on text and data mining. It therefore invited the delegations to express their position on (inter alia, Articles 11 and 13) the article in question, based on the question put by the Presidency itself, without prejudice to the possibility of raising other issues that might exist with regard to that particular provision. On the basis of the proposals made by the delegations, the Presidency would then consider whether to amend the consolidated compromise proposal.

"*Most delegations seem to agree that the mandatory TDM exception for research purposes laid down in Article 3 should be complemented by an optional exception for text and data mining in other situations covering temporary reproductions and with safeguards for rightholders. However, some Member States have expressed concerns about the scope of the optional exception, that they consider too broad, and suggested to limit its application.*
Should the scope of the optional exception for text and data mining provided for in Article 3a be limited and to what extent, for example to temporary copies of works and other subject matter which have been made freely available to the public online?" [484]

[482] Its main task is to prepare meetings at ministerial level of the Council of the European Union. COREPER plays a key role in the development of EU policies, as most of the negotiations between member states on decisions to be taken take place within it.

[483] T. VOLLMER, *Proposed Council compromise on TDM: still not good enough*, 6 April 2018. Available at: https://www.communia-association.org/2018/04/06/proposed-council-compromise-tdm-still-not-good-enough/

[484] Proposal for a Directive of the European Parliament and the Council on copyright in the Digital Single Market - Presidency compromise proposal. https://www.parlament.gv.at/PAKT/EU/XXVI/EU/01/76/EU_17608/imfname_10800647.pdf

The day before the planned COREPER vote, in which the Bulgarian Council Presidency sought a mandate to negotiate copyright reform with the European Parliament, academics from 25 leading intellectual property research centers in Europe published an open letter[485], expressing serious concerns about the legislative direction of the proposed copyright directive. The open letter followed a first open letter[486], published in February 2017, and was sent immediately after a statement from 169 European academics[487], demonstrating the increasing hostility of the academic and scientific community towards the directive. The open letter now added Article 3 on text and data mining to the two most criticised articles, namely Article 11 and Article 13, arguing in summary that the proposed legislation was "*failing its stated goals to improve choice, access and fairness in the digital environment*", asking members of Parliament to oppose the passage of the directive as drafted. In the letter, the academics recalled that from the outset they had demanded great transparency in the legislative process, as copyright was in their view "*a heavily lobbied field*", and there was therefore a great risk that the interests of small innovating firms, of non-organised creators and of consumers would not be taken into account. In this respect, they regretted the lack of transparency and evidence to justify the choices made in the last period in the work of both JURI and COREPER, which had come close to presenting their respective "*compromise*" texts. They noted in particular that the latest revisions of the Proposed Directive had been negotiated behind closed doors, as opposed to the period during which Cachia Comodini was rapporteur, a period in which The Committees of the European Parliament did their job in scrutinising the proposals and presenting explained reports[488].

With regard to text and data mining, they pointed out that there was a broad scientific consensus that the proposed exception for TDM in Article 3 would never be able to "*achieve its goal to stimulate innovation and research if restricted to certain organisations*". They also noted that rapporteur Therese Comodini Cachia (EPP) of the Legal Affairs committee JURI had initially submitted a balanced draft report, which opened the TDM exception under art. 3 to all, arguing that instead the legislative drafts that emerged from the offices of newly appointed rapporteur Axel Voss as well as from the Bulgarian Council presidency (since January 2018) "*pay lip service to authors' interest but respond in effect to the agenda of powerful corporate interests*". They therefore hoped that Comodini's draft report for the JURI Committee, which could still offer a solid basis for progress, would be taken up again, being it "*based on wide and transparent consultation, taking into account scientific evidence*". They therefore asked the Parliament and the Council, if the legislative process had progressed in the form proposed by the drafts of the Bulgarian Presidency and the JURI rapporteur Voss, to reject the Proposed Directive altogether, since it would have not served the public interest, "*failing its stated goals to improve choice, access and fairness in the digital environment*".

At the COREPER meeting on 27 April 2018, the mandate to negotiate was not granted to the Bulgarian Presidency of the Council on the draft Copyright Directive, as issues remained, including the final wording of Article 13. One month later, on 25 May 2018, the

---

485 *The Copyright Directive is failing* (Open Letter to Members of the European Parliament and the Council of the European Union), 26 April 2018. Available at: https://www.create.ac.uk/wp-content/uploads/2018/04/OpenLetter_EU_Copyright_Research_Centres_26_04_2018.pdf

486 EU Copyright Reform Proposals Unfit for the Digital Age, (Open Letter to Members of the European Parliament and the Council of the European Union), 24 February 2017. Available at: https://www.create.ac.uk/wp-content/uploads/2017/02/OpenLetter_EU_Copyright_Reform_24_02_2017.pdf

487 Academics Against Press Publishers' Right. Academics launch final appeal to European Parliament. Available at: https://www.ivir.nl/academics-against-press-publishers-right/

488 According to the academics, for example, MEP Zdzisław Krasnodębski (ECR) for the Industry committee ITRE drew attention to the narrow scope of the proposed Art. 3 exception for text-and-data-mining.

EU Council agreed its position on the draft Copyright Directive. By adopting this position, the Bulgarian Presidency thus received a mandate to start negotiations with the European Parliament (once its own position was agreed, brought forward to 20/21 June 2018).

This is the draft report approved by the Council:

Article 3
Text and data mining for the purposes of scientific research

1. *Member States shall provide for an exception to the rights provided for in Article 2 of Directive 2001/29/EC, Articles 5(a) and 7(1) of Directive 96/9/EC and Article 11(1) of this Directive for reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out text and data mining of works or other subject-matter to which they have lawful access, for the purposes of scientific research.*

1a. *Copies of works or other subject-matter made in compliance with paragraph 1 shall be stored with an appropriate level of security and not be retained for longer than necessary for achieving the purposes of scientific research.*

2. *Moved to Art. 6(1).*

3. *Rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject-matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.*

4. *Member States shall encourage rightholders, research organisations and cultural heritage institutions to define commonly-agreed best practices concerning the application of the obligation and measures referred to respectively in paragraphs 1a and 3.*

Article 3a
Optional exception or limitation for text and data mining

1. *Without prejudice to Article 3 of this Directive Member States may provide for an exception or a limitation to the rights provided for in Article 2 of Directive 2001/29/EC, Articles 5(a) and 7(1) of Directive 96/9/EC and Article 11(1) of this Directive for temporary reproductions and extractions of lawfully accessible works and other subject-matter that form a part of the process of text and data mining.*

2. *The exception or limitation provided for in paragraph 1 shall apply provided that the use of works and other subject matter referred to therein has not been expressly reserved by their rightholders including by technical means.*[489]

3.3.3.4. Parliament Position: The Voss Draft report and the disappointment of the European TDM community

At this point, after countless postponements, on 24 May 2018, the first compromise amendments with regard to the text and data mining (TDM) provisions of Article 3[490] were finally published by the Copyright Directive rapporteur, Axel Voss, with the following text:

[489] Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market - Agreed negotiating mandate, 25 May 2018. Available at: https://www.consilium.europa.eu/media/35373/st09134-en18.pdf

[490] Draft compromise amendments on Text and data mining. Version 1. 24 May 2018. Available at: https://www.communia-association.org/wp-content/uploads/2018/05/24052018-Draft-CA-on-Art.-3-v1.pdf

1. Member States shall provide for an exception to the rights provided for in Article 2 of Directive 2001/29/EC, Articles 5(a) and 7(1) of Directive 96/9/EC and Article 11(1) of this Directive for:

a) reproductions and extractions made by research organisations in order to carry out text and data mining of as regards works or and other subject-matter to which they have lawful access for the purposes of scientific research that are lawfully available online, provided that the rightholder has not reserved such uses in a machine-readable format.

This exception shall not apply to text and data mining of press publications within the meaning of Article 2 paragraph 4 of this Directive or to text and data mining of any works and other subject-matter incorporated in a press publication provided that publishers of these press publications express such reservation by listing their websites in a central point of information online.

Reproductions and extractions made for the process of text and data mining of such works and other subject-matter shall be deleted as soon as they are no longer required for this purpose.

b) reproductions and extractions of works or other subject-matter to which they have acquired lawful access made in order to carry out on a non-for-profit basis text and data mining for the purposes of scientific research by research organisations and cultural heritage institutions.
Any contractual provision contrary to the exception provided for in paragraph 1 b) shall be unenforceable.

Reproductions and extractions made for text and data mining purposes shall be stored in a secure manner. As soon as the research activity has ended the copies shall be deleted or, if Member States choose to make use of this possibility, stored by trusted bodies appointed for this purpose.

2. In cases other than the ones mentioned in paragraph 1, a license under which the licensee is allowed to carry out extractions and reproductions from a work or other protected-subject matter shall be deemed to allow for text-and-data-mining, including, where applicable, by their subcontractors, without requiring any specific permission of the rightholders, unless the parties agree otherwise and the rightholder reserves such uses in a machine-readable format.

Reproductions and extractions made for the process of text and data mining of such works and other subject-matter shall be deleted as soon as they are no longer required for this purpose. Any reproductions of works or other subject matter retained for longer than required are excluded from the scope of this paragraph.

3. Rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject-matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.

4. Member States shall encourage rightholders, research organisations and cultural heritage institutions to define commonly-agreed best practices concerning the application of the measures referred to in paragraph 1 point b and paragraph 3[491].

This initial amendment was described as *"reasonable"* by commentators[492], as for the first time since Voss took office, an attempt had been made to create a mandatory exception for all private sector entities (instead of being an optional restriction only for research and heritage institutions) and with purposes outside of strictly scientific research. Indeed, we have seen that one of the main limitations and problems with the exception in the Commission's proposal was the fact that the exception would only apply to research organisations operating on a non-profit basis or on the basis of a public interest mission recognised by a Member State. Therefore, the practical effect of this limitation was that the private sector would be completely excluded from the benefits of the exception, including important stakeholder groups such as journalists, citizen scientists, social enterprises, civil society organisations and cultural heritage organisations, which could certainly benefit from automated data analysis. This version of the exception was seen as a mashup of Article 3 of the Commission's proposal and Article 3a of the Council's text, "*a reasonable attempt at arriving at a compromise between those who agree with the Commission's original narrow approach and those, on the other hand, who argue for a much broader exception that allows anyone to engage in text and data mining for any purpose*". This seemed to be even a better approach than the Council's "3a" wording, because the exception was here mandatory, not merely optional. Moreover, this clause meant that any user would be allowed to conduct TDM on a work, except for those materials for which the rightsholders have reserved to do so because they have made them available in a machine-readable format. It was observed how the requirement that such reservations shall be made "*in a machine-readable format*" was particularly important, and went beyond the Council's suggestion that such requirements could be made "*by technical means*", since it opened up the possibility of engaging in text and data mining of online resources on a massive scale[493].

However, a few days later, on 5 June 2018, a second version of Article 3 on TDM included in the JURI Compromise agreement circulated on Twitter[494], which was significantly less positive for the text and data mining community. In this second version, the JURI Committee essentially aligned its position with the Council Common position, approved on 25 May (the day after the first version of JURI) by COREPER.

The second version of compromise included, like the Council version, two different articles, Article 3 and Article 3a, which read as follows:

Article 3

---

491 Draft compromise amendments on Text and data mining, 24.05.2018. Available at: https://www.communia-association.org/wp-content/uploads/2018/05/24052018-Draft-CA-on-Art.-3-v1.pdf

492 For a comment on the First Version of the Compromise amendment provided by JURI see: Communia Association, *Can Voss salvage the text and data mining exception*?, 4 June 2018. Available at: https://www.communia-association.org/2018/06/04/can-voss-salvage-text-data-mining-exception/

"Voss' compromise amendment is a mashup of Article 3 of the Commission's proposal and Article 3a of the Council text. In opposition to his approach in many other areas, the changes here seem to be a reasonable attempt at arriving at a compromise between those who agree with the Commission's original narrow approach and those (like us) — who argue for a much broader exception that allows anyone to engage in text and data mining for any purpose".

493 Communia Association, *Can Voss salvage the text and data mining exception?,* 4 June 2018. Available at: https://www.communia-association.org/2018/06/04/can-voss-salvage-text-data-mining-exception/

494 Second Version of the EU parliament draft proposal. https://twitter.com/communia_eu/status/1004047868280360963/photo/1

Text and data mining

1. *Member states shall provide for an exception to the rights provided to in Article 2 of Directive 2001/29/EC, Articles 5(a) and 7(1) of Directive 96/9/EC and Article 11(1) of this Directive for reproductions and extractions of works or other subject-matter to which they have acquired lawful access made in order to carry out text and data mining for the purposes of scientific research by research organisations and cultural heritage institutions.*
2. *Reproductions and extractions made for text and data mining purposes shall be stored in a secure manner. As soon as the research activity has ended the copies shall be deleted or, if Member States choose to make use of this possibility, stored by trusted bodies appointed for this purpose.*

Article 3a
Optional exception or limitation for text and data mining

Without prejudice to Article 3 of this Directive Member States may provide for an exception or a limitation for temporary reproductions and extractions of lawfully accessible works and other subject-matter that form a part of the process of text and data mining undertaken by others than research organizations, provided that the use of works and other subject matter referred to therein has not been expressly reserved by their rightholders including by machine readable means.

The public reaction to this sudden and unexpected change was not long in coming. Already on 8 June, 24 organisations representing universities, large and small technology companies, telecommunications and Internet service providers, start-ups and scale-ups, libraries, open access publishers, investigative and data journalists and non-profit organisations sent a letter[495] to Voss and the members of the JURI Committee expressing their disappointment and deep concern about the wording of Article 3 in the second version of the draft compromise amendments. They complained in particular about the fact that the JURI Committee had complied with the COREPER decision of 25 May, which contained the optional exception to Article 3a. This optional, not mandatory, exception would, according to them, lead to regulatory fragmentation in Europe, making it more difficult for most of the European research and innovation ecosystem (i.e. companies, in particular SMEs and start-ups, but also journalists or individual researchers) to carry out TDM activities. Being optional for Member States, it would have particularly affected those engaging in cross-border research and data analysis, defeating the purpose of building a digital single market. It would also have penalised research organisations engaging in public-private partnerships. It would also only allow *"temporary"* reproductions and extractions, which would significantly limit the techniques used by most TDM users today, simply negating the benefits of this optional exception and not reflecting how TDM works. They therefore asked the JURI Committee to return to their initial compromise proposal, which was in line with the objective of the European institutions' "*of reaching the right balance between the protection of effective business models and the need to provide a framework fit for innovation*". They also hoped that the European Parliament would propose a "*futureproof, broad mandatory exception for TDM for all*

[495] European Alliance for Research Excellence, *24 Organisations urge Rapporteur Axel Voss MEP to strike a more ambitious deal on TDM*, 8 June 2018. Available at: https://eare.eu/24-organisations-urge-rapporteur-voss-reach-ambitious-deal-tdm/

parties that have lawful access to the underlying work, and ensure Europe lives up to its ambition of becoming a global leader in AI, machine learning and data analytics".

Despite this, on 20 June 2018, the European Parliament's Committee on Legal Affairs (JURI), narrowly voted to approve the draft report, containing Article 3 and 3a: the mandatory exception for text and data mining (TDM) for the purposes of scientific research by research organizations and the optional exception allowing temporary reproductions made as part of the process of TDM. This vote also gave a mandate to Rapporteur MEP Axel Voss to start trilogue negotiations with the Council and Commission.

A full plenary vote to grant the negotiating mandate was then scheduled by the European Parliament in early July.

The following is therefore the version finally approved by the JURI Committee[496]:

Article 3
Text and data mining

Member States shall provide for an exception to the rights provided for in Article 2 of Directive 2001/29/EC, Articles 5(a) and 7(1) of Directive 96/9/EC and Article 11(1) of this Directive for reproductions and extractions of works or other subject-matter to which research organisations have lawful access and made in order to carry out text and data mining for the purposes of scientific research by such organisations.

Member States shall provide for educational establishments and cultural heritage institutions conducting scientific research within the meaning of point (1)(a) or (1)(b) of Article 2, in such a way that the access to the results generated by the scientific research cannot be enjoyed on a preferential basis by an undertaking exercising a decisive influence upon such organisations, to also be able to benefit from the exception provided for in this Article.

1a. Reproductions and extractions made for text and data mining purposes shall be stored in a secure manner, for example by trusted bodies appointed for this purpose.

2. Any contractual provision contrary to the exception provided for in paragraph 1 shall be unenforceable.

3. Rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject-matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.

4. Member States may continue to provide text and data mining exceptions in accordance with point (a) of Article 5(3) of Directive 2001/29/EC.

Article 3a
Optional exception or limitation for text and data mining

1. Without prejudice to Article 3 of this Directive, Member States may provide for an exception or a limitation to the rights provided for in Article 2 of Directive 2001/29/EC, Articles 5(a) and 7(1) of Directive 96/9/EC and Article 11(1) of this Directive for reproductions and extractions of lawfully accessible works and other subject-matter that form a part of the process of text and data mining, provided that the use of works and other subject matter

[496] https://www.europarl.europa.eu/doceo/document/A-8-2018-0245_EN.pdf?redirect

referred to therein has not been expressly reserved by their rightholders, including by machine readable means.

2. Reproductions and extractions made pursuant to paragraph 1 shall not be used for purposes other than text and data mining.

3. Member States may continue to provide text and data mining exceptions in accordance with Art. 5 (3) (a) of Directive 2001/29/EC.

Following the vote of the JURI Committee, on 26 of June 2018, The European Alliance for Research Excellence (EARE) expressed in a statement[497] once again its disappointment with the TDM provisions of Article 3 of the Copyright Directive, calling on the members of the European Parliament to reject the mandate to initiate inter-institutional negotiations in plenary session, in order to revise the current TDM exception and give the opportunity to policymakers to adopt a broad, future-proof mandatory TDM exception for all parties that have legal access to the underlying work. A few days later, in an open letter [498], the members of EuroDoc (The European Council of Doctoral Candidates and Junior Researchers), also agreed to the call to reject the Copyright Directive (with a plenary vote now scheduled for 5 July 2018), asking in particular that TDM should not qualify as copyright infringement and that Article 3 should be deleted or minimally expanded to allow text-and-data mining of copyrighted works for non-scientific purposes and organisations across Europe. On the same day, three days before the plenary vote in the European Parliament, more than 145 organisations, including start-ups, human rights organisations and researchers sent an open letter[499] to the European Parliament. The open letter called on MEPs to vote against the proposed reform, arguing that it created "*a huge gap between the expected value of the directive for the European economy and citizens and the harm the text will cause*" and that the proposal as formulated by the JURI Committee presented several risks, as widely highlighted by an impressive array of academics, internet pioneers, civil society organisations, start-ups and other experts.

On 5 July 2018, the Proposal for a Directive, in a full plenary vote by the European Parliament challenged the JURI vote and rejected the negotiation mandate and JURI report with 318 to 278 votes (and 31 abstentions), thus reopening the debate, which was postponed to September 2018. On 6 September 2018, a list of 45 amendments to the Directive was published, but there were no major amendments to the TDM provisions.

Two months after the first plenary vote, on 12 September 2018, the European Parliament finally adopted a position on the Copyright Directive, with 438 votes in favour, 226 against, and 39 abstentions.

Following the parliamentary adoption of the copyright directive on 12 September, several commentators expressed concerns about its potential harmful impact. For instance,

497 EARE's position on the JURI Committee Report on the Copyright Directive, 26 June 2018. Available at: https://eare.eu/assets/uploads/2018/06/EARE_Statement_JURI-Vote_Copyright-Directive.pdf

498 Eurodoc Open Letter to European Parliament on Copyright Directive, 2 July 2018. Available at: http://eurodoc.net/news/2018/eurodoc-writes-open-letter-to-european-parliament-on-copyright-directive

499 Call to Members of the European Parliament – Europe's citizens, startups, human rights organisations, publishers, creators, educators, cultural heritage professionals, librarians, and researchers ask for your support, 2 July 2018. Available at: https://copybuzz.com/wp-content/uploads/2018/07/Copyright-Open-Letter-on-EP-Plenary-Vote-on-Negotiation-Mandate.pdf

the European University Association[500] said it regretted the "*missed opportunity to bolster R&I*" due to the fact that the European Parliament ultimately did not improve on the Commission's already limited proposal for text and data mining (TDM) in Article 3 of the directive, and that it would have been better to adopt a broader and more mandatory copyright exception that would allow any user with legitimate access to data to engage in text and data mining for research, innovation and educational purposes, following the principle that "*the right to read is the right to mine*". The same disappointment was expressed by the usual European Alliance for Research Excellence[501], which said it was concerned that the exception formulated so far would probably "*hinder Europe's Artificial Intelligence (AI) ambitions*" and hoped that, as the European Parliament and the Council entered inter-institutional negotiations, there would "*still a chance to adopt a simple broad and mandatory future-looking TDM exception, which would protect the interests of rightsholders while encouraging European innovation*".

3.3.3.5. The Trilogue procedure and stakeholders' pressures to make the exceptions broad and mandatory. Directive adoption

Thus, having given the Parliament a mandate to negotiate, negotiations began behind closed doors on 2 October 2018 as part of the trilogue process between three key EU bodies: the Commission, the Parliament and the Council, in an attempt to agree a common position on the text of the Directive with a view to a final vote expected in 2019. The second round of trialogue negotiations on the Copyright Directive took place on 25 October 2018.

In the meantime, several research institutions (including Liber, Allied for Startups, and the European University Association) continued to push, by signing a letter[502] on 13 September 2018, for the EU policy makers to change their minds and decide to adopt a broad and mandatory text and data mining exception in the Copyright Directive (at that time article 3a). The letter stressed the fact that text and data mining tecniques are essential to the development of AI industries, and that an exception was required to remain competitive with countries such as the US, China, and Japan. The signatories also expressed concern that the text was apparently detached from the reality of text and data mining practices (see, for example, the allowance for "*temporary reproductions*"), and consequently would have been "*bad for science, research and innovation, AI, and ultimately bad for Europe*".

The third round of negotiations took place on 26 November, followed by a fourth round held on 3 December. At the final meeting, set for 13 December 2018, no agreement was reached with respect to the final text of the Copyright Directive.

A new date was set for 21 January, which was chaired by the new Romanian Presidency (rotating from the Austrian on 31 December). Before that, there was to be a vote by national governments on 18 January to give a new negotiating mandate to the new Romanian Presidency.

Just before the final trilogue negotiations, again, on 15 January 2019, several European research and innovation organisations (including Communia, Liber, and Allied for Startups)

---

[500] European University Association, *EU Copyright Reform: European Parliament vote is a missed opportunity to bolster R&I*, 13 September 2018. Available at: https://eua.eu/news/150:eu-copyright-reform-european-parliament-vote-is-a-missed-opportunity-to-bolster-r-i.html

[501] European Alliance for Research Excellence, *The European Parliament's Position on the Copyright Directive Will Hold Back European Research And Innovation*, 13 September 2018. Available at: https://eare.eu/european-parliaments-position-copyright-hold-back-research-innovation/

[502] LIBER, *Europe Needs A Broad & Mandatory TDM Exception*, 13 November 2018. Available at: https://libereurope.eu/article/europe-needs-a-broad-mandatory-tdm-exception/

sent another letter[503] to the EU legislators explicitly requesting to make the text and data mining (TDM) exception mandatory under Article 3a of the proposed Copyright Directive, highlighting how the then wording of the new TDM exception, limited to research organisations only, would have the capacity to suppress the use of TDM, potentially being used to support Europe's AI innovation in the current global race.

On 17 January, the Romanian presidency presented an updated proposal to be voted by the COREPER the following day. With regard to text and data mining, the proposal of the Council presidency specified that during the meeting with the other EU bodies, the Commission requested to turn Article 3a into a mandatory provision in order to avoid fragmentation in the DSM, pointing out that this was an issue on which no agreement had yet been reached, and that the EP required further internal reflections. He therefore asked COREPER, with a view to granting a mandate to negotiate the final trilogue on 21 January 2019, "*to confirm whether in an overall compromise package the Presidency can show flexibility to making Article 3a a mandatory provision in case the EP requests such change*"[504].

However, at the Council meeting on 18 January, the Romanian Presidency was not able to obtain the agreement of all national governments on the final text of the copyright directive, as the French and German positions, in particular on Article 13, were too far apart to obtain a mandate. Eleven countries would have voted against the proposed compromise text: Germany, Belgium, the Netherlands, Finland, Slovenia, Italy, Poland, Sweden, Croatia, Luxembourg and Portugal. As a result, what was supposed to be the final trialogue on 21 January was cancelled.

On 4 February 2019, an agreement was finally reached between the French and German governments in an attempt to obtain a negotiating mandate at the next COREPER meeting on 8 February 2019[505]. On 8 February 2019, the Romanian Presidency successfully obtained an updated negotiating mandate regarding the Copyright Directive. Eight countries reportedly dissented: Italy, Poland, the Netherlands, Sweden, Finland, Luxembourg, Malta and Slovakia.

On 11 February 2019, Final trilogue negotiations on the Copyright Directive began and were scheduled to finish on 13 February. On this day, a deal has been reached, meaning that the Copyright Directive would have continued through the legislative process. During this last meeting, the exception contained in Article 3a was finally made mandatory for member states[506]. The research community, in particular the European Alliance for Research Excellence, welcomed this latest unexpected development, stating that with this decision "*negotiators have sent a strong positive signal to European innovators*" and that "*a mandatory TDM exception will significantly encourage and bolster Europe's Artificial Intelligence ambitions*"[507].

A compromise text of the Copyright Directive was finally approved by EU governments at a COREPER meeting on 20 February 2019 by qualified majority[508]. The

---

[503] *Do not block Europe from becoming a leader in Artificial Intelligence (AI): Adopt a broad and mandatory Text and Data Mining (TDM) exception.* Available at: https://cdt.org/wp-content/uploads/2019/01/2019-01-15-Open-Letter-TDM-Copyright.pdf

[504] https://data.consilium.europa.eu/doc/document/ST-5138-2019-INIT/en/pdf

[505] Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market - Update of negotiation mandate. 7 February 2019. Available at: https://data.consilium.europa.eu/doc/document/ST-5893-2019-ADD-1/en/pdf

[506] The final text of the compromise agreement is now available here: https://juliareda.eu/wp-content/uploads/2019/02/Copyright_Final_compromise.pdf

[507] *Eare's position on the results of the trilogue discussion on the copyright*, Feb 22, 2019. Available at: https://eare.eu/eares-position-on-the-results-of-the-trilogue-discussions-on-the-copyright-directive/

[508] In particular, the Netherlands, Italy, Poland, Finland and Luxembourg would not support the agreement, labelling it "a step backwards for the digital single market rather than a step forward". Belgium and Slovenia were among those who abstained.

agreement paved the way for a vote at the JURI committee meeting on 26 February, achieved by 16 votes to 9, with no abstentions, and a final vote at the plenary of the European Parliament in mid-April. About a month later, on 26 March, a vote in the plenary of the European Parliament saw the Copyright Directive adopted without amendment[509], with 348 votes in favour and 274 votes against. To complete the co-decision procedure, the Council of the European Union approved the Copyright Directive by qualified majority on 15 April[510].

3.3.4. The official text, reactions to the directive and critical analysis of the new TDM exceptions at Arts. 3 and 4 of Directive 790/2019

The final text of the directive, as approved by the Parliament and the Council, therefore includes two different mandatory exceptions dedicated to text and data mining: the first in article 3, dedicated to text and data mining carried out in the context of scientific research, and the second in article 4, addressed to other uses not limited to scientific research, plus article 7, which prevents the contractual override of the provision contained in article 3.

The following is the official text of the provisions:

Article 3
Text and data mining for the purposes of scientific research

1. *Member States shall provide for an exception to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, and Article 15(1) of this Directive for reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access.*
2. *Copies of works or other subject matter made in compliance with paragraph 1 shall be stored with an appropriate level of security and may be retained for the purposes of scientific research, including for the verification of research results.*
3. *Rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.*
4. *Member States shall encourage rightholders, research organisations and cultural heritage institutions to define commonly agreed best practices concerning the application of the obligation and of the measures referred to in paragraphs 2 and 3 respectively.*

Article 4
Exception or limitation for text and data mining

1. *Member States shall provide for an exception or limitation to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of Directive 2009/24/EC and Article 15(1) of this Directive for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining.*
2. *Reproductions and extractions made pursuant to paragraph 1 may be retained for as long as is necessary for the purposes of text and data mining.*

[509] https://www.europarl.europa.eu/doceo/document/A-8-2018-0245-AM-271-271_EN.pdf
[510] Italy, Luxembourg, the Netherlands, Poland, Finland and Sweden voted against the directive while Belgium, Estonia and Slovenia abstained.

3. The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.
4. This Article shall not affect the application of Article 3 of this Directive.

Article 7
Common provisions

1. Any contractual provision contrary to the exceptions provided for in Articles 3, 5 and 6 shall be unenforceable.
2. Article 5(5) of Directive 2001/29/EC shall apply to the exceptions and limitations provided for under this Title. The first, third and fifth subparagraphs of Article 6(4) of Directive 2001/29/EC shall apply to Articles 3 to 6 of this Directive.

To summarise, therefore, and to give a final overview of the features of the two articles finally approved, the body of EU Directive 790/2019, in the part concerning the TDM, thus outlines a substantially bipartite discipline, based on two provisions with a parallel content: it requires all EU member states to implement Article 3, as a specific exception for the research sector, and Article 4, as a general closing exception.

This last provision seems only apparently to provide for a wide-ranging exception. Starting from the common prerequisite of lawful access to the target sources, it does not provide for any subjective limitation and is therefore addressed to any person intending to engage in text and data mining activities and does not even set any limitation as to the individual purposes pursued, thus including commercial exploitation. These seemingly broad and generalised starting conditions are however followed by the provision of the third paragraph, according to which the exception applies provided that the use for text and data mining purposes has not been expressly reserved by the rightholder in an appropriate manner. In essence, the exercise of the mining activity by third parties depends on a substantially potestative right of the rightholder to be expressed in advance and in a technically appropriate way, thus confining the actual operation of the exception to a hypothesis (presumably residual) of "*silent consent*", with a consequent contraction of the practical impact of the provision.

The European legislator, with the general exception and its dual basis on the condition of legitimate access and the reservation mechanism, rather than identifying a truly exceptional case, has created or rather institutionalised a derivative market for the extraction of text and data in favour of rightholders, leaving the offer to the respective company policy choices.

On this basic assumption, with the parallel introduction in Article 3 of the directive of the mandatory exception in favour of research organisations, the European legislator at the same time intended to support the European research sector, enhancing its contribution to the collective interest and providing for a favourable regime in terms of exemption from the payment of royalties.

On a subjective level, the categories of beneficiaries are research organisations and cultural heritage institutions. The definition of "*research organisations*" is however not new in EU law, since it is already familiar from the state aid rules. According to Article 2(1) of EU Directive 790/2019, the definition of "*research organization*" covers all those entities

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 basis or by reinvesting all*

the 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*[511].

The reactions to the directive as such were generally rather negative. We have seen how already during the legislative process various stakeholders such as academic, scientific and startup communities expressed almost at each new step their disappointment with the chosen approach to the TDM issue, the lack of transparency of the work and the lack of ambition[512] in the drafting of the provisions therein contained. As seen before, these criticisms about the proposal gradually led to some improvements between the first version of the exception proposed by the Commission and the final version text of the directive. As a result, for example, the circle of beneficiaries of the exception has been extended to "*cultural heritage institutions*". In addition, the possibility to store works for search and mining purposes has been added, which is an important feature, since research takes time, and it must therefore be possible to get back to the works carrying the data without having to reproduce them from time to time. Finally, a new exception has been introduced in the body of the directive, according to which "*Member States shall provide for an exception or limitation to the rights (...) for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining*"[513].

However, this was clearly not enough to definitively put an end to the controversy that had been surrounding the directive for some time. After the directive came into force, the criticism became, if possible, even stronger, and some commentators have described it as one of the most controversial pieces of legislation in recent decades[514].

As regards the directive as a whole: first of all, it has been pointed out that, having been initially conceived with the aim of being a legislative instrument meant to promote the Digital Single Market, it soon turned into an "*industry policy tool, shaped more by effective lobbying than evidence and expertise*", the results of which were found to be at least unsatisfactory, containing multiple problematic provisions, such as the new press publishers' right (then Article 11, now 15) and the so-called value gap or "*upload filter*" provision (then Article 13, now 17). The legislative process was actually affected by multiple pressures from interest groups as well as civil society, through protests reminiscent of the ACTA debate, oppositions by digital rights NGOs and Internet luminaries, and multiple expert statements by research institutes and academics. However, the lobbying by rights holders' representatives certainly surpassed in intensity and effectiveness the opposite positions. A lot of criticism has been made for having given a regulatory preference to private over public choice in EU copyright law, without adequate safeguards for users[515].

With specific regard to TDM, it has been pointed out that the discipline resulting from the legislative process clearly shows a legislative preference for the creation of a secondary TDM market delivered into the hands of right holders, rather than making data analysis

---

511 Article 2 Digital Single Market Directive: Definitions

512 For an early critique of the lack of ambition of the EU legislator in the field of copyright, see the foundational reflections in: P.B. HUGENHOLTZ (ed.), *The Future of Copyright in a Digital Environment*, Den Haag, Kluwer, 1996, in particular the chapter by the editor himself, *Adapting Copyright to the Information Superhighway*, 81 ss.

513 C. GEIGER, *The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the failed Text and Data Mining Limitations in the CSDM Directive*, PIJIP/TLS Research Paper Series 2021, No. 66, 8.

514 R. CASO, *Il conflitto tra diritto d'autore e ricerca scientifica nella disciplina del text and data mining della direttiva sul mercato unico digitale*, Trento LawTech Research Paper nr. 38, 9.

515 QUINTAIS, *The New Copyright in the Digital Single Market Directive: A Critical Look*, cit., 2-3.

activity a more free and certain activity. As has been astutely observed that, while there are many rights holders who may not have an interest in such markets, there are others who have already showed this interest and will therefore try to defend their exclusive control of the data, such is the case with the big scientific publishers, that have now substantially turned into data analysis companies[516].

Several authors have also highlighted the complexity and length of the directive, characterised by its massive size, confusing language, contradictory regulatory requirements and inability to bring about effective harmonisation, which eventually delivers the Member States a text that will be quite difficult to implement and interpret and which will almost certainly lead to legal disputes[517]. These textual complexity and prolixity end up making the provisions contained in the exceptions, which should benefit all citizens, almost useless. In fact, it has been observed that nowadays, when, as we saw in the previous chapter, the digital revolution has levelled out most of the technological and economic barriers that separated authors and publishers from readers, the category of "*scrilettori* (*write-readers*)" (those who, thanks to the internet, share, quote and rework other people's works), becomes the category that will be most affected by the complexity of the exceptions. These provisions are more likely meant to respond to particular interests and could result in a real pervasive economic censorship rather than a rethink of a system that in reality has more write-readers than traditional mediators, bearers of the old monopoly[518].

However, coming to the content of the directive, and in particular the two new exceptions for text and data mining, several commentators managed to underline some positive elements in the wording of the two exceptions, among the negative aspects. Some, for instance, welcomed first of all the fact that the problem was acknowledged and the issue addressed; furthermore, some noted positively that the rather narrow original wording proposed by the committee was eventually partly improved[519]; others have welcomed the fact that recital 9[520] has finally clarified how TDM does not necessarily interfere with copyright and neighbouring rights (Recital 9 CDSMD) and that it could also involve the processing of mere facts or data, which are not copyrightable as such[521]; another positive aspect highlighted by many commentators is the fact that at least the TDM exception for research is binding for the parties and cannot be overridden by contract[522]; another important step, that materialised during the legislative procedure, has been the recognition of the importance of TDM not just for research purposes, but also due to its various applications in other fields

---

516 CASO, *Il conflitto tra diritto d'autore e ricerca scientifica nella disciplina del text and data mining della direttiva sul mercato unico digitale*, cit., 22.
On the aim of the Directive to create a secondary market, see also: B. CALABRESE, *Sulla dimensione imprenditoriale e societaria degli organismi di ricerca ai fini di text and data mining*, in Contratto e impresa, 2020, 4, 1598.

517 CASO, *Il conflitto tra diritto d'autore e ricerca scientifica nella disciplina del text and data mining della direttiva sul mercato unico digitale*, cit., 9.

518 M. C. PIEVATOLO, *L'età del privilegio*, Rivista Il Mulino. Available at: https://www.rivistailmulino.it/a/l-et-del-privilegio

519 QUINTAIS, *The New Copyright in the Digital Single Market Directive: A Critical Look*, cit., 3.

520 Recital 9: "Text and data mining can also be carried out in relation to mere facts or data that are not protected by copyright, and in such instances no authorisation is required under copyright law. There can also be instances of text and data mining that do not involve acts of reproduction or where the reproductions made fall under the mandatory exception for temporary acts of reproduction provided for in Article 5(1) of Directive 2001/29/EC, which should continue to apply to text and data mining techniques that do not involve the making of copies beyond the scope of that exception".

521 R. DUCATO, A. M. STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, in European Intellectual Property Review, 2021, 8.

522 *Ibid.*, 10.

(see Recital 18)[523]. Therefore, to the initial and only TDM research exception proposed by the Commission in 2016, the final text of the Directive added a second TDM exception available to anyone for any purpose (Article 4)[524]. Moreover, both exceptions introduced were finally made mandatory, which means that all Member States will have to adopt them in their national law[525].

Despite the presence of these positive features, a mood of disfavour and disappointment remained about the whole outcome of the directive.

Actually, it is difficult not to notice already from the literal tenor of the rules contained in the proposal for the Directive, and from the recitals which partly explained the reasons for its introduction, how the legislator initially (and perhaps also later), with the introduction of the sole exception for scientific research purposes, had not fully understood the real potential of text and data mining for the future of the Union in terms of innovation and wealth, and above all its essentiality for the development of the artificial intelligence industry of the European Union[526]. Even in its current formulation, Article 3 seems not to recognise the value and reality of scientific research nowadays, which has long since ceased to be confined to the walls of traditional research institutions, and whose public benefit goes far beyond what the law seems to admit. Examples of this were the so-called behaviourally informed regulations, where TDM could be used by legislators to test new legislation; or the use that journalists could make of it to check the authenticity of news. Moreover, private commercial researchers also play a key role in research nowadays, above all in the development of the artificial intelligence industry. All examples of research that are now potentially excluded from the scope of the exception[527].

It is only later, in response to criticism during the proceedings, that the European legislator seems to have partly understood the importance of TDM and introduced the exception for cases not covered by article 3 of the directive in order to authorise text and data mining for those and for the purposes excluded from the scope of article 3[528]. This shift is clearly seen in recital 18, which states that:

> *in addition to their significance in the context of scientific research, text and data mining techniques are widely used both by private and public entities to analyze large amounts of data in different areas of life and for various purposes, including for government services, complex business decisions and the development of new applications or technologies.*

Overall, therefore, the opinion that has emerged most among the commentators is the feeling that this could be considered a "*missed opportunity*". There are those who, for example, commenting on the Directive, have effectively sustained, using a football metaphor, that with this Directive the strategy of the European Union with regard to artificial intelligence can be likened to a "*football team missing a goal-scorer to win any of the competitions with other jurisdictions*

[523] Recital 18: "In addition to their significance in the context of scientific research, text and data mining techniques are widely used both by private and public entities to analyse large amounts of data in different areas of life and for various purposes, including for government services, complex business decisions and the development of new applications or technologies".

[524] DUCATO, STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out,* cit., 10.

[525] *Ibidem.*

[526] GEIGER, *The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the failed Text and Data Mining Limitations in the CSDM Directive*, cit., 5.

[527] DUCATO, STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, cit., 12.

[528] GEIGER, *The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the failed Text and Data Mining Limitations in the CSDM Directive*, cit., 8.

having more flexible limitations to copyright"[529]. Skepticism is even more justified if we consider the recent new plans to strengthen the European artificial intelligence industry. In February 2020, in fact, the European Commission announced an ambitious digital strategy for the European Union, setting out the objectives to be achieved in two communications - one on Europe's digital future[530] and one on data[531] - both complemented by a "*White Paper on artificial intelligence*"[532]. From these documents, from which emerges a great ambition of the European Union in the field of technology and artificial intelligence, one would hope that there would be a determination to create, as has happened, we shall see, in other states outside of the European Union, a legal framework capable of supporting and incentivizing such plans. Indeed, the desire to modernise and adapt the IP legal framework to the challenges posed by the digital environment was once again stated in the "*European Commission's IP action plan*"[533], published at the end of November 2020. However, it has been pointed out that none of these documents seems to reflect the will of the European legislator to study what are the real free spaces left by IP law to allow the development of a balanced digital ecosystem in the EU. This issue was also totally ignored by the recent "*Proposal for a Regulation on a European approach for Artificial Intelligence*"[534], published on April 2021, notwithstanding, as stated in the proposal itself, one of the main goals put forward by the European Commission is to "*ensure legal certainty to facilitate investment and innovation in AI*". It seems, therefore, that the Commission considers that the matter has already been resolved through the "*Directive of 17 April 2019 on Copyright and related rights in the Digital Single Market"* (CDSM-Directive), and that there is no need to conduct further reflections on the role and current appropriateness of European copyright law, and in particular of its exceptions and limitations, for the development of artificial intelligence, having evidently not understood that the CDSM-Directive itself is, according to the majority doctrine, far too restrictive with regard to the exceptions for text and data mining ed in direct contradiction with the objectives it has set and that, therefore, the "*many ambitions newly put forward are likely to remain a dead letter*"[535].

Let us now look specifically at why these exceptions have failed to meet expectations. Even from a superficial reading of the two rules it is impossible not to notice that a little more could have been done and dared. In regulating the new exceptions, it does indeed seem, as has been pointed out, that they have been formulated in such a way as to "*take back with one hand what they grant with the other*"[536]. In fact, one wonders why one cannot submit to automatic analysis the texts one is entitled to read unless one works for research organisaitons and cultural heritage institutions.

With regard to article 3, the exception can only be invoked by two categories of beneficiaries (research organisations and cultural heritage institutions) and exclusively for a specific purposes, i.e. scientific research, with the inevitable and unfortunate inconvenience

[529] *Ibid.*, 5.
[530] Communication from the Commission, "Shaping Europe's digital future", Brussels, 19.02.2020, COM(2020) 67 final.
[531] Communication from the Commission, "A European strategy for data", Brussels, 19.02.2020, COM(2020) 66 final.
[532] European Commission, "White Paper on Artificial Intelligence - A European approach to excellence and trust", Brussels, 19.02.2020, COM(2020) 65 final.
[533] Communication from the Commission, "Making the most of the EU's innovative potential, An intellectual property action plan to support the EU's recovery and resilience", 25 Nov. 2020, COM(2020) 760 final.
[534] Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, Brussels, 21.4.2021 COM (2021) 206 final.
[535] GEIGER, *The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the failed Text and Data Mining Limitations in the CSDM Directive*, cit., 2-3.
[536] PIEVATOLO, *L'età del privilegio*, cit.

that, for example, independent researchers carrying out individual research projects in the public interest or the whole nebula of researchers linked to the institution who are between two short-term employment contracts, may be automatically excluded from the possibility of benefiting from the exception. The fact that the exception is exclusively linked to the concept of scientific research may mean that the same research organisations cannot benefit from it for other purposes indirectly linked to research but equally important and functional, such as administrative or educational activities[537]. The same applies to start-ups operating in the digital environment, which remain subject to exclusive rights in their data mining activities, even though their potential for innovation in the area of artificial intelligence is universally recognised[538].

This is without considering that the very notion of "*scientific research*" itself has been unclear from the very beginning. Recital 12 of the directive specifies that scientific research refers both to works belonging to the natural sciences and to those traditionally belonging to the humanities[539]. This, it has been noted, is somewhat vague, and could potentially narrow the scope of the exception, as it is not easy to subsume certain activities under one or the other notion. Applied research, for example for public health or relating to technology, might fall outside the scope of the Article 3 exception, such as computer science that is not definitely classified as natural science *stricto sensu*. This might of course have implications in the field of AI development. Recital 12 could, therefore, have been formulated in a better way by referring to scientific research tout court, thus including any activity performed according to the pertinent methodological standards, aimed at generating new knowledge and advancing the state of the art in a given field[540].

Going back to the issue of start-ups or individual researchers not permanently embedded in a research organisation, fortunately, after considerable pressure, Article 4 was subsequently inserted, which finally extended the range of beneficiaries of the exception. Sure, but Article 4 is only apparently broader than the counterpart TDM research exception, since the rightholder has the power to opt-out, by expressly reserving the right to make reproduction or extraction for TDM in an appropriate manner, for example via the terms and conditions of the website or via machine-readable means. Actually, some platforms have already moved to prevent TDM in their online terms of use, thus contractually prohibiting users from any act of reproduction made on the content of their website. Another means by

---

537 DUCATO, STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, cit., 12.

538 GEIGER, *The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the failed Text and Data Mining Limitations in the CSDM Directive,* cit., 6.

539 Recital 12: "Research organisations across the Union encompass a wide variety of entities the primary goal of which is to conduct scientific research or to do so together with the provision of educational services. The term 'scientific research' within the meaning of this Directive should be understood to cover both the natural sciences and the human sciences. Due to the diversity of such entities, it is important to have a common understanding of research organisations. They should for example cover, in addition to universities or other higher education institutions and their libraries, also entities such as research institutes and hospitals that carry out research. Despite different legal forms and structures, research organisations in the Member States generally have in common that they act either on a not-for-profit basis or in the context of a public-interest mission recognised by the State. Such a public-interest mission could, for example, be reflected through public funding or through provisions in national laws or public contracts. Conversely, organisations upon which commercial undertakings have a decisive influence allowing such undertakings to exercise control 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, should not be considered research organisations for the purposes of this Directive".

540 DUCATO, STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, cit., 11.

which rightholders prohibit TDM is by using an exclusion protocol that prevents crawling or indexing (the robot.txt[541] protocol)[542].

While one can hope that the majority of rightholders have no interest in preventing or monetising TDM activities on the material they hold, it would be utopian to think that this will be the rule, especially among scientific publishers, multinational corporations that are veritable money machines[543], there is a growing awareness of the additional value attached to the aggregate set of scientific articles published from an informational perspective. It is foreseeable, therefore, that in the not-too-distant future publishers will gear up to offer pay-for-text and data mining services and other strategies aimed at monetising access to a new and promising market such as the TDM market[544].

Some have pointed out that the legislator's dual objective of considering the legitimate interests of rightholders on the one hand and promoting research on the other, with a legal framework that could support innovation, both expectations could have been met by subjecting the exploitation of the TDM activity carried out for commercial purposes to a correlative right of remuneration to the rightholder, namely the one who has borne the costs of creating and maintaining the databases[545].

What is mainly criticised by the doctrine is therefore the basic political choice of the EU legislator, i.e. the need to give more freedom of action to research organisations (including universities) and cultural heritage institutions while giving less freedom to other public and private stakeholders in TDM[546]. The TDM rules have been interpreted by some as a kind of European aid in the sense of a regulatory measure to support and encourage research, at the same time restricting the group of beneficiaries just to those institutions involved in research[547]. Criticism of the doctrine has pointed to the fact that restricting the mandatory exception only to certain categories of subjects is not justified, especially taking into account that research is becoming increasingly precarious nowadays and that the EU itself expressly declares its intention to promote a science that involves citizens in research activities[548].

A further point of disagreement by the doctrine criticises the legislator for having formulated the two exceptions so that they are not a means to provide access to content for TDM purposes, considering that they already require lawful access[549] as a prerequisite to the

---

[541] The robot.txt is an exclusion protocol that content providers can insert into the root directory to prevent crawling or indexing activities on certain pages of their website. See, http://www.robotstxt.org/robotstxt.html.

[542] DUCATO, STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, cit., 13.

[543] Think for example of Reed-Elsevier, a multinational giant giant with annual revenues exceeding £6bn. See: S. BURANYI, I*s the staggeringly profitable business of scientific publishing bad for science?*, The Guardian, 27 June 2017. Available at: https://www.theguardian.com/science/2017/jun/27/profitable-business-scientific-publishing-bad-for-science

[544] B. HUGENHOLTZ, *The New Copyright Directive: Text and Data Mining (Articles 3 and 4)*, Kluwer Copyright Blog, July 24 2019. Available at: http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/

[545] GEIGER, *The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the failed Text and Data Mining Limitations in the CSDM Directive*, cit., 9.

[546] CASO, *Il conflitto tra diritto d'autore e ricerca scientifica nella disciplina del text and data mining della direttiva sul mercato unico digitale*, cit., 18.

[547] CALABRESE, *Sulla dimensione imprenditoriale e societaria degli organismi di ricerca ai fini di text and data mining*, cit., 2.

[548] CASO, Il conflitto tra diritto d'autore e ricerca scientifica nella disciplina del text and data mining della direttiva sul mercato unico digitale, cit., 19.

[549] Where by "Lawful access" the directive means at Recital 14 as "access to content based on an open access policy or through contractual arrangements between rightholders and research organisations or cultural heritage institutions, such as subscriptions, or through other lawful means [...] Lawful access should also cover access to content that is freely available online"

possibility of using the exceptions themselves. According to some in fact[550], this could have the consequence of excluding from the use of the exceptions those research organisations or cultural heritage institutions that do not have the means to afford to pay for subscriptions or entities that do not have lawful access to the content. Commentators note that if Article 3 is read in conjunction with recitals 10 and 17 (which, respectively, warns about the disadvantages that research organisations may have when the terms of licences on content to which they have lawful access may exclude TDM, and that Member states should not provide for compensation for rightholders), might lean towards the interpretation that the rightholder should not be able to make separate and additional licences for the datasets for the purpose of TDM and that compensation should not be provided, meaning that where there is already lawful access, TDM for research purposes should be allowed without any additional burden. However, they also note that the recitals of the directive are non-binding, and this has the potential consequence that unless the resource is otherwise publicly available online, application of the exception could, according to the rule, be circumvented by the publishers', for example, through a change in business model, even just by increasing the subscription fees indiscriminately for all the research organisations[551].

The enormous potential limitation to the successful application of the second exception was also pointed out. In fact, the requirement of lawful access, which is reiterated by recital 18[552] also for the exception provided for in Article 4, provides however that the user must have lawful access also with regard to the content made available to the public online, and that this is the case only when the rightholders have not reserved in an appropriate manner the rights to make reproductions and extractions for TDM. It has been observed that if such reservation of rights is not easily and immediately intelligible to the user (e.g. through machine-readable means) this could translate into a difficulty for the user in verifying the existence of any reservation of the rights on resources accessible online, who would thus have to check one by one the terms and conditions of each website before performing TDM, which is a rather impracticable alternative[553].

Further concerns raised by commentators relate, in the case of Article 3, to the status of copies of protected material made during the TDM process. The provision of the law, according to which such copies may be retained, is excessively general because it does not specify whether they may be reused, again for purposes of scientific research, by persons

---

[550] See: DUCATO, STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, cit.; R. HILTY, H. RICHTER, *Position Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules Part B Exceptions and Limitations (Art. 3–Text and Data Mining)* (2017); See, GEIGER, FROSIO, BULAYENKO, *The exception for text and data mining (TDM) in the proposed Directive on Copyright in the Digital Single Market: legal aspects: in-depth analysis* (European Parliament 2018).

[551] DUCATO, STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, cit., 19.

[552] Recital 18: "This exception or limitation should only apply where the work or other subject matter is accessed lawfully by the beneficiary, including when it has been made available to the public online, and insofar as the rightholders have not reserved in an appropriate manner the rights to make reproductions and extractions for text and data mining. In the case of content that has been made publicly available online, it should only be considered appropriate to reserve those rights by the use of machine-readable means, including metadata and terms and conditions of a website or a service. Other uses should not be affected by the reservation of rights for the purposes of text and data mining. In other cases, it can be appropriate to reserve the rights by other means, such as contractual agreements or a unilateral declaration. Rightholders should be able to apply measures to ensure that their reservations in this regard are respected. This exception or limitation should leave intact the mandatory exception for text and data mining for scientific research purposes provided for in this Directive, as well as the existing exception for temporary acts of reproduction provided for in Article 5(1) of Directive 2001/29/EC".

[553] DUCATO, STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, cit., 19.

outside the research organisation that carried out the mining, also possibly for the purpose of verifying the quality of the research itself. The question is crucial, because it involves the very possibility of verification by the scientific community, also composed of individuals, of the quality of the research. It is not by chance that Germany, in its exception to the TDM we have seen before, Section 60d of the Urheberrechtsgesetz, has expressly established that the corpus created through TDM must be accessible to individual third parties to check the quality of scientific research[554].

As regards Article 4, on the contrary, unlike the corresponding exception for scientific research activities in Article 3, the storage of the copies made for TDM is limited to the activity of mining per se, meaning that once the TDM process has been completed, the reproduced material should be deleted. This reading of the transitional nature of the copies made is also confirmed by the fact that the legislature does not require in this case, contrary to Article 3(2), the adoption of particular security measures for the storage of the mined corpus[555].

Other criticisms have focused on the lack of symmetry of the two provisions with regard to the rights they restrict. Article 3 in fact restricts the reproduction right of the author of the database (Art. 5(a) Dir. 96/9), the right of extraction and/or re-use of the maker of the database (Art. 7(1) Dir. 96/9), the reproduction right of the InfoSoc. directive (Art. 2 Dir. 2001/29) and the new right on journalistic publications under Art. 15 of the same CDSM directive. Article 4, on the other hand, restricts all the rights now listed plus the reproduction right of the software directive (Art. 4.1(a) and (b) Dir. 2009/24). This choice is not easily explainable and there is no justification in the recitals. On the other hand, the software directive does not contain a provision allowing universities, other research bodies and cultural heritage institutions to carry out TDM on software code. The exceptions and limitations provided for in Directive 2009/24 (in particular those provided for in Article 5) do not seem to overlap with that provided for in Article 3 of the CDSM Directive[556].

As if these exceptions were not already sufficiently limited and rather restrained in their scope, the regulatory text finally provides, in Article 7(2), that the three-step test (TST) and the provisions on technological protection measures apply to the new exceptions and limitations, including those concerning the TDM. Thus, Article 5(5) and the first, third and fifth subparagraphs of Article 6(4) of Directive 2001/29 apply[557].

To start with the three-step-test, the fact that the provision states that it applies to the two new exceptions means that they must be interpreted restrictively. As already mentioned at the end of the previous chapter, the European implementation of the three-step-test in Article 5(5) InfoSoc Directive is usually interpreted by the European Court of Justice restrictively, making it a tool to support restrictive interpretations of exceptions and limitations. It cannot yet be predicted how the application of this test will affect the application of the two exceptions in the future. We have said, however, that although there are already some publishers who already offer paid-for TDM, this is not yet very developed, as it is also unlikely that they will be able to demonstrate how the two TDM exceptions conflict with the "*normal exploitation*" of their journals, books or newspapers. The fact remains, however, that invoking the application of the TST could potentially lead as an additional limitation to the TDM exceptions[558].

554 *Ibidem*, 12.
555 *Ibidem*, 13.
556 CASO, *Il conflitto tra diritto d'autore e ricerca scientifica nella disciplina del text and data mining della direttiva sul mercato unico digitale*, cit., 19.
557 *Ibid.*, 21.
558 DUCATO, STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, cit., 16.

As regards the relationship between the two exceptions and the aforementioned provisions on technological protection measures, the reference made by Article 7 to the latter suggests that they are an important tool to protect the legitimate interests of the rightholders. In fact, it has been noted that the provision of the above-mentioned Article 6(4) of the InfoSoc Directive[559], which provides that rightholders, in implementing TPMs must in any case take care to ensure through "*voluntary measures*" the usability by users of the exceptions, and that in case of inertia of the latter the Member States will intervene with appropriate measures, has never demonstrated an adequate functioning in the last 20 years, and it is likely that this will not change now or in the future. The Directive therefore seems to have missed the opportunity to introduce a mandatory prohibition for TPMs to override the exceptions[560] and to use TPMs as a means of circumventing the application of the exceptions.

It is also difficult to foresee how the new technological measures, provided for in art. 3 (3), which rightholders are authorized to apply in order "*to ensure the security and integrity of the networks and databases where the works or other subject matter are hosted*", will affect the practical usability of the exceptions, being them clearly different from the traditional TPMs of the InfoSoc Directive, and which may include, as mentioned in recital 16, access control measures, but also integrity protection mechanisms, such as cryptography, watermarking, etc. [561]. Although recital 16 states that these measures should not preclude the effective application of the exception and, being also provided by the second sentence of paragraph 3 that the same measures cannot go beyond what is necessary to achieve the specific objective of security and integrity of networks and databases, it is not clear at this time what remedies the user can take in order to protect against an instrumental use of such technical measures by the rightholder[562].

Finally, it is precisely the reference of the new directive to the provisions contained in the InfoSoc directive and, in general, to the directives of the previous acquis, which risks leading to a new framework (moreover still characterised by the prevalence of private order over public interest), resulting from the intersection of several directives, which is even more complex and uncertain than the previous one, thus seriously jeopardising the harmonisation

---

[559] Art. 6 (4) InfoSoc Directive: "Notwithstanding the legal protection provided for in paragraph 1, in the absence of voluntary measures taken by rightholders, including agreements between rightholders and other parties concerned, Member States shall take appropriate measures to ensure that rightholders make available to the beneficiary of an exception or limitation provided for in national law in accordance with Article 5(2)(a), (2)(c), (2)(d), (2)(e), (3)(a), (3)(b) or (3)(e) the means of benefiting from that exception or limitation, to the extent necessary to benefit from that exception or limitation and where that beneficiary has legal access to the protected work or subject-matter concerned.
A Member State may also take such measures in respect of a beneficiary of an exception or limitation provided for in accordance with Article 5(2)(b), unless reproduction for private use has already been made possible by rightholders to the extent necessary to benefit from the exception or limitation concerned and in accordance with the provisions of Article 5(2)(b) and (5), without preventing rightholders from adopting adequate measures regarding the number of reproductions in accordance with these provisions.
The technological measures applied voluntarily by rightholders, including those applied in implementation of voluntary agreements, and technological measures applied in implementation of the measures taken by Member States, shall enjoy the legal protection provided for in paragraph 1.
The provisions of the first and second subparagraphs shall not apply to works or other subject-matter made available to the public on agreed contractual terms in such a way that members of the public may access them from a place and at a time individually chosen by them.
When this Article is applied in the context of Directives 92/100/EEC and 96/9/EC, this paragraph shall apply mutatis mutandis".

[560] DUCATO, STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, cit., 17.

[561] *Ibidem*.

[562] CASO, *Il conflitto tra diritto d'autore e ricerca scientifica nella disciplina del text and data mining della direttiva sul mercato unico digitale*, cit., 21.

and simplification intention which the directive was intended to achieve, allowing all the actors, including researchers, to have clear guidance on what they are allowed to do or not, also in a cross-border context. And this picture is even more at risk if one considers that the directive is without prejudice to the possibility for Member States to maintain the already enacted TDM exceptions or to adopt broader provisions compatible with those established in by the Database and InfoSoc Directives (Article 25 CDSMD) as long as they do not limit the scope of the mandatory exceptions or limitations provided for in the CDSMD (Recital 5)[563].

3.3.5. The national implementation of the Directive

We have seen how the directive could have done better and how, in order to be able to compete in the field of scientific research and artificial intelligence, it will be necessary, as Geiger claims[564], to reopen the question of the appropriateness of exceptions for TDM as soon as possible, or else the ambitious plans of the European Union mentioned above will not come to fruition.

The directive was voted in 2019 and member states had until 7 June 2021 to implement its provisions into their national legislation. At the time of writing this paragraph, only a few countries have yet implemented the directive[565][566].

On 26 June 2021, the European Union Commission initiated infringement proceedings against 23 EU member states for failure to comply with their obligations within the deadline[567].

Given these premises, something can therefore still be done at national level when implementing the directive to prevent the narrow requirements of the two new exceptions from excessively and unjustifiably restricting text and data mining. Articles 3 and 4, in fact, serve as a minimum benchmark with which Member States have to comply with. The member states could therefore more specifically improve the regulatory aspect in several

563 DUCATO, STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, cit., 20.

564 GEIGER, *The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the failed Text and Data Mining Limitations in the CSDM Directive*, cit., 11.

565 In particular, the Netherlands, Hungary, Germany, Malta and, only partially, France have implemented the directive within a reasonable period of time. Other Member States, like Spain, Ireland, Estonia, Croatia and Italy have proceeded to implement the directive in the second half of 2021.
For the status of individual Member State implementations see: COMMUNIA DSM Directive Implementation Portal. Available at: https://www.notion.so/DSM-Directive-Implementation-Portal-97518afab71247cfa27f0ddeee770673. A further useful and nice resource to follow the national implementation of the Directive is: COMMUNIA Eurovision DSM Contest. Available at: https://eurovision.communia-association.org.

566 Only a brief mention can be made here, due to the general nature of this work, about the Italian implementation of the directive. Italy finally transposed the provisions of the CDSM Directive on 12th December 2021, through the Legislative Decree 8 November 2021, n.º 177. With respect to the implementation of the exception for the extraction of texts and data for scientific research purposes (Article 3), the Italian legislator has surprisingly gone beyond what was provided in the CDSM Directive. In fact, by making use of the space available under the exception for scientific research in the InfoSoc Directive (Article 5(3)(a)), the new Italian exception covers not only the acts of reproduction and extraction from copyrighted materials, but also the act of communicating research results to the public, which is deemed essential to disseminate such results and to foster collaboration among researchers. This has been welcomed as a positive aspect of the Italian implementation by the doctrine, despite the fact that the transposition of other aspects of the directive have been subject to numerous and harsh criticisms.

567 F. Y. CHEE, *Commission starts legal action against 23 EU countries over copyright rules*, Reuters, 26 June 2021. Available at: https://www.reuters.com/world/europe/commission-starts-legal-action-against-23-eu-countries-over-copyright-rules-2021-07-26/

crucial provision, which have already been criticised by several parties. This would be possible thanks to the wording of Article 25 CDSMS, which provides that member states may adopt "*broader provisions, compatible with the exceptions and limitations provided for in Directives 96/9/EC and 2001/29/EC, for uses or fields covered by the exceptions or limitations provided for in this Directive as long as they do not limit the scope of the mandatory exceptions or limitations*" introduced with the CDSM Directive[568].

For example, when implementing Article 3, member states could give a very broad definition of "*research organisations*", which Article 3 refers to subjectively without defining them[569]. It should be recalled that the rule in fact lays down the legal status of the free activity of text in data mining, placing a mandatory exception, not subject to any conditional reservation, recognised as not overridable by contracts and also free of charge. Indeed, from an interpretative point of view, doctrine has stressed that the wording of the directive can be identified in several places as evidence of the European legislator's intention not to confine the applicability of the exception to the field of public research institutions alone. The notion outlined in Article 2(1) of EU Directive 790/2019, as we have seen, undoubtedly gives central prominence to universities, including their libraries and other public research institutions. The fact, however, that the reference has been expressly extended to "*any other entity*" that satisfies the finalistic and substantive requirements, underlines a purely functional conception that makes it possible to broaden the scope of the exception also to other private and even entrepreneurial and corporate entities. The notion is based on three constituent elements:

1. the scientific research mission;
2. the non-profit or public interest purpose;
3. the prohibition of privileged access to results by a controlling company.

It has been stressed that none of these constituent elements excludes per se from the group of research organisations private collective entities, as seems to be confirmed by the corresponding notion in the field of State aid. On the contrary, the definition of research organisation would in theory show significant margins of compatibility with its possible entrepreneurial and corporate dimension. The objective of scientific research, which, as the *"primary"* (though not necessarily only) qualifying feature of the organisation, must be explicitly reflected in its constitution and/or statutes, can be fully included in the corporate purpose of both an association and a company. The very adjective "*scientific*" of the research has not been deemed irreconcilable with a possible commercial dimension of innovation[570].

It was then noted that a systematic view of the policy objectives set out in Article 179 TFEU[571] provides a framework in which research, whose "*scientific*" bases are mentioned, is

---

568 DUCATO, STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, cit.,

569 V. FALCE, *Direttiva Copyright 2019: fair use ed eccezioni al copyright tra esigenze di «apertura» e necessità di indirizzo*. Available at: https://www.filodiritto.com/direttiva-copyright-2019-fair-use-ed-eccezioni-al-copyright-tra-esigenze-di-apertura-e-necessita-di-indirizzo

570 CALABRESE, *Sulla dimensione imprenditoriale e societaria degli organismi di ricerca ai fini di text and data mining*, cit., 3.

571 Art. 179 TFUE:

1. The Union shall have the objective of strengthening its scientific and technological bases by achieving a European research area in which researchers, scientific knowledge and technology circulate freely, and encouraging it to become more competitive, including in its industry, while promoting all the research activities deemed necessary by virtue of other Chapters of the Treaties.

2. For this purpose the Union shall, throughout the Union, encourage undertakings, including small and medium-sized undertakings, research centres and universities in their research and technological development activities of high quality; it shall support their efforts to cooperate with one another, aiming, notably, at permitting researchers to cooperate freely across borders and at enabling undertakings to exploit the internal market potential to the full, in particular through the opening-up of national public contracts, the definition of common standards and the removal of legal and fiscal obstacles to that cooperation.

framed in terms of the competitiveness of the entire European system "*including in its industry*", and also expressly encourages the participation of "*small and medium-sized undertakings*" in strategic innovation paths, thus demonstrating the coexistence of a commercial soul of research that is broadly scientific in European support programmes[572]. The scientific research mission significantly identified as a "*primary*" but not necessarily exclusive target is reflected in the day-to-day operations of research organisations, which, alongside their scientific activities, carry out entrepreneurial activities related to research, but seen as a contractual service. The fact that the research organisation must act on a non-profit basis "*or*" with lucrative results provided that all profits generated are reinvested in the scientific research itself, by equating the option of non-distribution constraints with non-profit, confirms the possible entrepreneurial nature of the activity carried out alongside the research organisation as well as the possibility and even the purpose of making profits and, with this, the possibility of its declination in corporate form. Finally, further "*entities*" are considered to be identified on the basis of the recognition of a public task without any preclusion referring to the profit profile and therefore also purely commercial entities, for which the qualification of research organisation may derive from the awarding of a public contract, further confirming the fact that the figure of the research organisation may also be extended to subjective types outside the public dimension of the traditionally understood research organisations. Finally, the very fact that the research organisation must behave in such a way that it does not allow a preferential transfer of the results of the TDM activity to a possible undertaking exercising a decisive influence over it implies *"literally"* that the possibility of corporate control by an undertaking does not as such preclude its inclusion in the category of research organisation[573].

Moving on two the second exception, also when implementing the so-called "*opt out*" of Article 4, member states could give the term "*expressively reserved in an appropriate manner*" a restrictive meaning, imposing for instance on rightholders the possibility to use the reservation only under certain conditions, with the introduction of formalities or registration requirements[574], or, for instance, the express indication of the reasons that make it reasonable, also in the light of the common interest and the right to information, also protected by Article 10 ECHR[575].

Furthermore, Member States could grant a broader exception to users acting for commercial purposes who play a decisive role in research in the field of artificial intelligence and also remove any further doubts about the possibility to share the TDM-generated corpora with other researchers, at least for verification purposes and specifically allow TDM on software for research purposes[576].

3.4. Alternative proposals for reforming the copyright system in order to exclude TDM from the copyright scope of protection

3. All Union activities under the Treaties in the area of research and technological development, including demonstration projects, shall be decided on and implemented in accordance with the provisions of this Title.

[572] For a more in-depth look at the role of public action and related models of innovation market competitiveness, see M. LIBERTINI, *La nuova disciplina degli aiuti a favore di ricerca, sviluppo e innovazione*, in E. A. RAFFAELLI, Antitrust between EU and national law, Milano-Bruxelles, 2015, 460 ss.

[573] CALABRESE, *Sulla dimensione imprenditoriale e societaria degli organismi di ricerca ai fini di text and data mining*, cit., 4.

[574] GEIGER, *The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the failed Text and Data Mining Limitations in the CSDM Directive*, cit., 11.

[575] *ibidem*.

[576] DUCATO, STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out,* cit., 28.

As we have seen, one of the main criticisms about the directive with regard to the regulation of text and data mining is that the European legislator has decided to go on with the traditional construction of copyright as composed of rights and exceptions and for not having taken the opportunity to evaluate whether instead, it was worth exploring other alternatives, such as, for example, a reverse approach, in which exclusivity would be the exception and the public domain the rule[577].

Some of the components of the European academic community have, over time, tried to propose alternative and different solutions that would lead to a reform of the European copyright and, as a consequence, indirectly, allow the free functioning of TDM, excluding it from the scope of application of copyright[578].

3.4.1. The "use as a work" theory

Let us look at some of these alternative theories, starting with the one proposed first by Strowel[579] and then by Ducato[580]: the "*use as a work theory*".

These two academics first of all claim that the European legislator, regulating text and data mining with a specific exception, has in fact confirmed, despite the clarification in recital 9, that TDM is an activity relevant to copyright (and database right), emphasizing how a great opportunity to discuss the real boundaries of the right of reproduction and the foundations of copyright has been wasted. Having done this, it would have been possible to restore an effective balance of the same, changing the paradigms, now old, belonging to the era of reproduction and analogue copying. They claim that an alternative could have been a clarification about the conditions for an infringement of the right of reproduction, and that it would have represented a viable and less costly alternative to follow.

In fact, they argue that copyright should not extend to the TDM process, since no author in the past has ever intended to limit the use of his work as a useful source of information or, in producing a work, has ever thought of earning revenue from derivative uses arising from searching or indexing a corpus that includes that work. Moreover, going back to the origins of copyright, they further point out, as we extensively saw in chapter two, that copyright was originally not just intended to remunerate authors (and publishers) for creating (and disseminating) works, but was also meant to expand public learning. This is not only true for the US copyright system, but also for the continental droit d'auteur system. Copyright, in fact, on both sides of the Atlantic, does not cover facts (including correlations), information or other elements considered as non-protectable ideas.

The main problem, according to the authors, is due to the fact that the birth and development of digital technology has somewhat raised the expectations of economic returns for rights holders, who now tend to believe that any reproduction, even without economic reason, must be covered by copyright, in a way that in the author's opinion is not commensurate with the economic fundamentals that apply to digital uses, especially on the internet, since the explosion of digital copies cannot be translated into the same increase in value that would be generated by further non-digital copies. In fact, they claim, many digital copies have no value in and of themselves and are merely the basis or material for value-

577 M. C. PIEVATOLO, *L'età del privilegio,* cit.

578All the theories mentioned below are contained in the book: P.B. HUGENHOLTZ, *Copyright Reconstructed: Rethinking Copyright's Economic Rights in a Time of Highly Dynamic Technological and Economic Change*, Information Law Series, Vol. 41, Alphen aan den Rijn: Wolters Kluwer.

579 A. STROWEL, *Reconstructing the Reproduction and Communication to the Public Rights: How to Align Copyright with Its Fundamentals.*

580 DUCATO, STROWEL, *Ensuring Text and Data Mining: Remaining Issues With the EU Copyright Exceptions and Possible Ways Out*, cit., 27.

added services (e.g. copies made to detect significant word occurrences). The notion of technology-neutral reproduction fits poorly into the digital realm because of the massive technology-driven copies that have no commercial value.

They therefore propose a concrete solution that, in their view, could be extremely useful in solving the current problem arising from the limitations of the two exceptions introduced and their likely fragmented implementation by the member states. In particular, they present a purpose-oriented reading of the conditions for infringing the reproduction and extraction rights that exonerates the acts of copying or extracting made during the TDM process.

The solution to this problem is a qualitative test that could help distinguish an infringing copy from a non-infringing copy. While on the one hand, as we have seen, a communication must be addressed "*to a public*" to be considered copyright relevant, this is not the case for acts of reproduction, which do not mention the need for a public or the possibility of the public having access to the work.

The solution would therefore be the introduction of a real test for an infringement of the reproduction right that would require the work to be used as a work and perceived as such by an audience. This would be a solution for TDM precisely because in TDM the use of the work is never "*as a work*"[581] but "*as a tool*", thus leaving it outside the application of copyright. The expressive characteristics of the work are not used, and there is no audience to use the work, as the work is only an input in a process of searching a corpus and identifying occurrences and possible trends or patterns. The same can be said about the copyright on the database provided by the database directive. When datasets that are original in the selection or arrangement of elements or that contain original works are extracted in order to find correlations, patterns and links between points of information, the dataset or its original elements are not used as works for an audience, and this should not constitute copyright infringement, even if partial copies are made.

Also talking about the sui generis database rights contained in the Database Directive, the authors put forward a similar solution, arguing that the TDM acts do not infringe the sui generis (database) right, stressing that also in the opinion of the CJEU the underlying ratio of the introduction of these rights is to protect the maker of the database "*against the unauthorised appropriation of the results of that investment by acts which involve, in particular, the reconstitution by a user or a competitor of that database or a substantial part of it at a fraction of the cost needed to design it*" [582]. TDM in fact does not involve "*the reconstitution*" of (part of) the database in a competing product. TDM is, on the contrary, to be considered as a form of "*consultation*" and, as underlined by the CJEU, the protection granted by Article 7 "*does not, however, […] cover consultation*" [583]. According to the Court, "*where the maker of a database makes the contents of that database accessible to third parties, even if he does so on a paid basis, his sui generis right does not allow him to prevent such third parties from consulting that database for information purposes*"[584].

---

581 As put by the 2nd Circuit in the *Authors Guild v. Google, Inc.* U.S. case 'the purpose of [defendant]'s copying of the original copyrighted books is to make available significant information *about those books*, permitting a searcher to identify those that contain a word or term of interest' (emphasis added).

582 Case 304/07 Directmedia Publishing GmbH v. Albert-Ludwigs-Universität Freiburg [2008] ECLI:EU:C:2008:552, § 33, relying on the earlier decisions in Case 46/02 Fixtures Marketing v Oy Veikkaus Ab [2004] ECLI:EU:C:2004:694, § 35; Case 203/02 The British Horseracing Board and Others v William Hill Organisation Ltd [2004] ECLI:EU:C:2004:695, §§ 32, 45, 46 and 51; Case 338/02 Fixtures Marketing Ltd v Svenska Spel AB [2004] ECLI:EU:C:2004:696, § 25; and Case 444/02 Fixtures Marketing v Organismos prognostikon agonon podosfairou [2004] ECLI:EU:C:2004:697, § 41.

583 Directmedia Publishing GmbH v. Albert-Ludwigs-Universität Freiburg (n 100), § 51 relying on The British Horseracing Board and Others v William Hill Organisation Ltd (n 100), § 54.

584 Directmedia Publishing GmbH v. Albert-Ludwigs-Universität Freiburg (n 100), § 53.

According to the authors, what the CJEU refers to as "*appropriation*" does not therefore presuppose a "*proprietary*" protection, understood merely as a right on a subject matter, but it must instead be seen and judged in relation to rights defining the relations between the rights owner and third parties. Those relations are always determined in the infringement test. The sui generis right is based on a misappropriation or unfair competition rationale (Recital 39 of the Database Directive indicates that it "*seeks to safeguard the position of makers of databases against misappropriation"*). This is also confirmed by the opinion of Advocate-General Szpunar, who has recently claimed that "*the sui generis right provided for in Article 7 of Directive 96/9 has as its objective to protect database makers against the creation of parasitical competing products*"[585]. Therefore, according to the AG opinion,

"*the national courts should therefore verify not only whether the extraction or reutilisation of the whole or a substantial part of the contents of a database has taken place and whether it is shown that there has been a substantial investment in either the obtaining, verification or presentation of those contents, but also whether the extraction or reutilisation in question constitutes a risk to the possibilities of recouping that investment*" [586].

In other words, "*the condition that there be an adverse effect on the investment of the maker of a database*" should be factor in the infringement analysis.

Improperly allowing the "*propertization*" of the sui generis right resulting from the absence of a correct analysis of the infringement conditions, could actually lead to the creation of a property right on data, a form of appropriation of mere facts and data, which is exactly what the directive itself intended to avoid[587].

3.4.2. The "right to a reasonable exploitation"

Rognstad and Poort[588] instead propose an alternative copyright model that[589], by simplifying the economic rights provided by copyright, can easily adapt them to meet the challenges brought about by technological change in recent years. This alternative model takes the form of granting the right holder an exclusive right to a "*reasonable exploitation*" of the work, the further scope of which should be specified in statutory provisions. The latter should thus definitively replace the traditional construction of the copyright system, which takes the form of a positive definition of acts subject to the exclusive rights, such as reproduction, communication to the public and distribution, subject to exceptions that need justification.

Everything revolves around the concept of *"reasonable exploitation"*, and the whole Copyright System would thus be based on a *"one stage"* test according to which the delineation of the right lies in the definition of what is reasonable rather than in what constitutes *"exploitation"*. This would permit to replace the classic *"six stage assessment"* consisting of

---

585 Opinion of Advocate General Szpunar, 14 Jan. 2021 in Case C-762/19 SIA 'CV-Online Latvia' v. SIA 'Melons', § 40.

586 Idem, § 47 (see also § 59).

587 Recital 45: "The right to prevent unauthorized extraction and/or re-utilisation does not in any way constitute an extension of copyright protection to mere facts or data"
Recital 46 "[…] should not give rise to the creation of a new right in the works, data or materials themselves".

588 O.-A. ROGNSTAD and J. POORT, *The Right to a Reasonable Exploitation Concretized. An Incentive Based Approach* in .B. HUGENHOLTZ, *Copyright Reconstructed: Rethinking Copyright's Economic Rights in a Time of Highly Dynamic Technological and Economic Change*, Information Law Series.

589 This model was initially proposed by O.-A. ROGNSTAD, *Restructuring the Economic Rights in Copyright – Some Reflections on an "Alternative Model"*, Journal of the Copyright Society of the U.S.A 62(4), 503–544.

definitions of exclusive rights, limitations to them, secondary liability, safe harbour and protection of technical protection measures (TPMs) with exceptions.

They too criticise the current formulation of copyright as being based on the broader concept of *"exploitation"*, thus covering all kinds of behaviour and business models that are not only in the current but also in the future financial interest of the right holder, covering not only existing exploitation models but also those yet to be developed.

According to the alternative model proposed by the authors, however, the delineation of exclusive rights is not derived from a classification of what constitutes an act of exploitation but rather from what is considered to be reasonable exploitation. Everything therefore depends on how the right of reasonable exploitation is understood and realised. It must be constructed having regard to the principle of efficiency. They recall, as we have already seen in the previous chapter, how the economic literature on copyright is basically in agreement in arguing that the logic of copyright protection is the incentive to create works of literature, science and art, at the cost of creating a temporary monopoly on these works, the static welfare losses due to this monopoly being justified by the dynamic welfare gains from new creation. Put simply, this exclusive right to certain acts of exploitation allows the right holder to generate revenue and recoup the investment made to create the work. As with many other economic activities, costs precede benefits and will only be incurred if future benefits can be expected.

The concrete implementation of the right of reasonable exploitation must also be built around the principle of proportionality, which consists in making sure that the protection is not excessive or disproportionate to what is considered fair. This implies that even if considerations of efficiency may point in the direction of giving the right holder the right to control a certain type of conduct, it should be possible to adjust the result by reference to what, under the circumstances, would result in disproportionate protection for the right holder, and in principle the reverse.

In addition, the public domain also comes into play, according to which not all information can or should be protected by IP rights. In copyright, this is reflected in various types of rules, for example, the distinction between unprotected ideas and protected expression, the originality criterion, and duration limitation. Not to mention other principles used as criteria to assess the reasonableness of use, such as the principle of dignity, freedom of expression and the principle of market integration.

With regard specifically to the application of the right to reasonable exploitation to TDM, they argue for the weakness of the economic arguments used to place TDM within the ambit of economic rights, based mostly on the argument that any unlicensed use that has economic significance and consequently generates additional profits unreasonably prejudices the legitimate interests of the right holder and therefore should be licensed. This reasoning resembles the argument criticised in earlier sections of the same paper, namely, in the case of digital resale, that authors should somehow be entitled to a share of any surplus created by or with the works in downstream markets. However, whereas in the case of digital resale the market failure can be considered resolved for every physical or digital copy that is sold, in the case of TDM, it can be observed that there has never been a market failure in relation to TDM that needs to be resolved, as long as access to the items to engage in TDM is obtained legally.

In particular, the authors go on to analyse TDM in the light of the principles that serve to construct the right to reasonable exploitation. Regarding the efficiency principle, they argue that the additional profits generated by TDM have never been part of the authors' incentive system. In their opinion, it makes no sense to grant rights that provide ineffective incentives or none at all. In this sense, the emergence of TDM can be considered

unpredictable in the past and cannot have retroactive incentives. Now that TDM is an emerging field of research, it may have become predictable and part of the incentive system for existing publishers who aggregate large databases of articles. But to the extent that TDM was not foreseen in existing contracts for access to such databases, it can be assumed not to have been the case. In that case, requiring an additional license for TDM would give windfall profits to rightsholders without any incentive effect, while imposing deadweight losses and transaction costs on TDM.

Moreover, as regards the other principles, the guiding principles other than efficiency, TDM control may hinder freedom of expression and does not fit well with the principle of proportionality. The relevance in copyright terms of this kind of exploitation can therefore be highly contested. Therefore, the control should not be considered as a reasonable exploitation of the work.

Finally, after outlining the features of this theory, they propose a concrete normative solution to the problem, a *"third category"* style regulation, encompassing both the benefits of rules and standards, i.e. predictability and flexibility respectively. A typical catalogue comprises "*two or more enumerated instances (or items) followed by a general provision authorising courts to recognise other non-enumerated instances (or items) that have the same key characteristics as the enumerated ones*".

CHAPTER 4

COMPARATIVE ASSESSMENT OF THE RELEVANT NON-EUROPEAN LEGISLATIONS ON TEXT AND DATA MINING

After having analysed in detail the European TDM discipline, both in its pre-reform features and in those resulting from the recent European Copyright reform, in this final chapter we are going to turn our attention to other non-European jurisdictions that have accommodated or are considering accommodating TDM in their copyright legislation, and, where present, in their database protection legislation.

Some of these legal systems, in particular the US (worldwide main model of the "*fair use*" approach) and the Japanese ones, as seen before, have been taken several times as paradigm by those stakeholders who, pending the legislative process of Directive 790/2019, were calling for a broad exception, benefiting a large number and variously composed users, covering all activities, carried out for any purpose, with both non-commercial and commercial purposes.

Indeed, several countries have long since recognised the importance of TDM as a useful tool for the development of artificial intelligence and how it can lead and support the digitisation process of their industries and, consequently, benefit the country's overall economy.

We will face some of these countries and see how they have addressed this issue, starting with the United States, the main exponent among those countries that have a fair use clause; we will then move on to Japan, an example of a particularly wide and well thought-out TDM-specific exception; we will then look at some countries that have determined to introduce a fair use clause, modifying their copyright system, notably Israel; finally, we will analyse some other countries that have thought or are thinking of regulating TDM in various other ways.

4.1. Text and Data Mining in the United States: the "fair use" clause

The US approach to text and data mining is markedly different from the European Union one. We have seen in the second chapter how, generally, there are two approaches to regulate exceptions to copyright: on the one hand, the approach of specific exceptions, more frequent among countries belonging to the civil law tradition; on the other hand, the approach of a general clause, more frequently adopted (although, as we shall see when discussing some Asian countries, not always) by the so-called common law countries. We have also explained its foundation, motivation, advantages and disadvantages.

The US approach belongs to the second category and is certainly its most cited example.

Indeed, in a number of cases involving digital technologies, since the early years of the turn of the century, federal courts have held that the fair use clause makes lawful the activities of computational analysis along with the creation of digital archives that enable research services. These statements have greatly supported the arguments of those who wanted to recognise the same archiving activities carried out in TDM equally as fair use[590].

Also with regard to TDM, therefore, it can be claimed that the US courts have in a certain way authorized the acts of systematic copying and reproduction carried out in order

[590] M. W. CARROLL, *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful*, 53 *UC Davis L. Rev.* 893 (2019), 916.

to create a database necessary to perform text and data mining (for example to create a search service for these works) and how they may consider such activities as a "*transformative use*"[591], falling within the scope of application of the general clause called "*fair use*", even when the content copied in their database is protected by copyright. In addition, in the United States the doctrine identifies the concept of "*non-consumptive*" or "*non-expressive*" uses in those copying activities that are performed not to provide others with access to the expressive content of the works in the database[592].

The United States copyright law, therefore, does not have a specific limitation or exception explicitly allowing TDM, but has allowed the creation and development of TDM as a new research tool through case law, believing that fair use could apply to it[593].

Now, before analysing the relationship between Fair Use and TDM, let's see what this theory actually means and how it developed over time.

Formally, fair use is one of the limitations or exceptions to copyright rights provided to rightholders by Section 106(a), nonetheless holding that they are "*subject to sections 107 through 122*". These sections therefore provide a set of uses of protected material that Congress has decided are "*not an infringement of copyright*", and otherwise fully within the scope of the rights provided by 106(a)[594].

Fair use is one of these non-infringing uses and is included in the list in Section 107. It has long been considered a critical component of the American copyright system, a very useful tool designed to make sure that copyright does not hinder the development of changing and evolving technologies and to allow new uses of protected material that may not (or could not) have been conceived of when the Copyright Law was last amended: a limitation that allows copyright to adapt to new and emerging technologies and changing circumstances, thus ensuring a prompt and rapid balancing of copyright. The doctrine of fair use originated as an equitable flexible doctrine, thus being initially merely a common law principle (i.e. a judge-made law through resolving cases in which judges apply existing legal principles to facts of the case). In 1976, however, this principle was also codified in Section 107 of the Copyright Law[595]. The House Report, published after the enactment of the 1976 Copyright Law, explained its introduction into the Copyright Act:

"*The statement of the fair use doctrine in section 107 offers some guidance to users in determining when the principles of the doctrine apply. However, the endless variety of situations and combinations of circumstances that can rise in particular cases precludes the formulation of exact rules in the statute. The bill endorses the purpose and general scope of the judicial doctrine of fair use, but there is no disposition to freeze the doctrine in the statute, especially during a period of rapid technological change. Beyond a very broad statutory*

---

591 *Ibid.,* 941.

592 See, e.g., M. SAG, *Copyright and Copy-Reliant Technology*, (analyzing cases); M. SAG, *Predicting Fair Use*, 73 *Ohio St. L.J.* 47 (2012), 56-57 (showing that a finding of transformativeness predicts a holding of fair use); M. SAG, *The Google Book Settlement and the Fair Use Counterfactual*, 55 *N.Y.L. Sch. L. Rev.* 19 (2010), 54 (discussing, in the context of the Google Book settlement, "the creation of a 'Research Corpus' for non-consumptive and non-commercial research by certain qualified users"); M. SAG, *Orphan Works as Grist for the Data Mill*, 27 *Berkeley Tech. L.J.* 1503 (2012), 1503; J. GRIMMELMANN, *Copyright for Literate Robots,* 101 *Iowa L. Rev.* 657 (2015), 665.

593 K. L. COX, *Text and Data Mining and Fair Use in the United States*, Association of Research Libraries, 2015, 2. Available at: https://www.arl.org/resources/text-and-data-mining-and-fair-use/

594 CARROLL, *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful*, cit., 911.

595 The reason why Fair Use was codified in 1976 is explained in the H.R. Rep. No. 94-1476, pag 66. Available at: https://www.copyright.gov/history/law/clrev_94-1476.pdf: "Although the courts have considered and ruled upon the fair use doctrine over and over again, no real definition of the concept has ever emerged. Indeed, since the doctrine is an equitable rule of reason, no generally applicable definition is possible, and each case raising the question must be decided on its own facts. On the other hand, the courts have evolved a set of criteria which, though in no case definitive or determinative, provide some guage for balancing the equities".

explanation of what fair use is and some of the criteria applicable to it, the courts must be free to adapt the doctrine to particular situations on a case-by-case basis"[596].

Congress, therefore, in spite of its codification through its introduction within the Copyright Law, firmly stated the intention not to limit the flexibility it had experienced over time in its various applications in case law[597], hoping that it would maintain this essential feature and continue to be an equitable rule, capable of accommodating the rapid technological change already foreseeable in those years[598].

The concrete operation of the Fair Use clause, as a limitation on copyright, is similar to that of a specific exception clause: when one of the criteria of the clause is alleged to be met, the user need not seek permission from the rightholder to use the otherwise protected material.

Let us now see what the clause specifically provides and the conditions under which it can be considered applicable to a particular situation. It is useful here to specify that the four factors listed below have to be assessed as a whole and how a defendant need not win on every factor for a court to rule in favour of fair use[599].

The fair use statute, codified at 17 U.S.C. 107, reads:

> *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 include:*
>
> *(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."*

As we have seen in chapter one, listing the many applications that text and data mining can serve, TDM can be carried out for a countless number of different purposes, some of which are also set out in Section 107, namely scholarship and research. Text and data mining has repeatedly been declared by the courts to be a fair use of protected materials. In particular, there have been at least eight cases[600] since 2003 indirectly involving TDM and the creation of databases with protected material, where the purposes for which TDM was

---

596 H.R. Rep. No. 94-1476, 66.

597 The Congress also explicitly said so, stating: "Section 107 is intended to restate the present judicial doctrine of fair use, not to change, narrow, or enlarge it in any way", in H.R. Rep. No. 94-1476, 66.

598 The Courts have also repeatedly stressed the importance of applying fair use in a flexible manner. See, for example, Perfect 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146, 1166 (9th Cir. 2007), in which it says: "[W]e note the importance of analyzing fair use flexibly in light of new circumstances".

599 COX, *Text and Data Mining and Fair Use in the United States*, cit.

600 In descending chronological order, the following cases can be mentioned: *Authors Guild v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014); *White v. West* (S.D.N.Y. 2014); *Fox v. TVEyes* (S.D.N.Y. 2014); *Authors Guild v. Google*, 770 F.Supp.2d 666 (S.D.N.Y. 2011); *A.V. v. iParadigms, LLC* (4th Cir. 2009); *Perfect 10 v. Amazon*, 508 F.3d 1146 (9th Cir. 2007); *Field v. Google*, 412 F.Supp.2d 1106 (D. Nv. 2006); *Kelly v. Arriba Soft*, 336 F.3d 811 (9th Cir. 2003).

exploited were research, use by policymakers or plagiarism check[601]. Most of these cases, however, concerned the creation of databases with protected material and their use in the context of search engines, which have been appearing around the turn of the century. For example, the first case, dating 2003, *Kelly v. Arriba Soft*, 336 F.3d 811 (9th Cir. 2003)[602], concerned a search engine company, Arriba Soft, which operated a search engine that produced thumbnail-sized images in response to search queries. These thumbnails were collected by a "*crawler*", an automated computer programme that goes through the web, visiting and indexing the pages it encounters. When it came across an image, Arriba's crawler would download full-size copies of that image to Arriba's servers. Arriba's software would then reduce the images to thumbnail size, delete the full-size copies, and present the thumbnails in its search results[603].

The court held here that Arriba Soft's search engine could be seen as a useful tool to help index and improve access to images on the Internet[604].

Other subsequent cases[605] involved more or less the same topic, where search engines basically made available copyrighted works (images or book extracts). For instance, a later case, *Perfect 10 v. Amazon*[606], from 2007, concerned Google Image Search, which created, stored, and displayed thumbnails in more or less the same way[607]. Other cases, such as *A.V. v. iParadigms, LLC* (4th Cir. 2009)[608], concerned the creation of databases for the operation of anti-plagiarism software (in this case the famous TurnItIn), which allowed teachers to compare work submitted by their students on their own site with content on the net as well as with papers previously posted on the service to determine whether a work had been copied or not.

Many of the courts that have ruled on the issue, we will now see as we address the four factors individually, have particularly focused on the major benefits that TDM can bring to the public, because they "*enhance[e] information-gathering techniques*"[609] and noted that the creation of search engines and databases were to be considered highly transformative uses, recognising that these activities do not serve as a substitute for the original work and that these databases satisfy a new purpose[610].

---

601 COX, *Text and Data Mining and Fair Use in the United States*, cit.
See also: *Campbell v. Acuff-Rose*, in which the Court states that the factors must be "explored, and the results weighed together, in light of the purposes of copyright."

602 *Kelly v. Arriba Soft*, 336 F.3d 811 (9th Cir. 2003). Available at: https://openjurist.org/336/f3d/811

603 B.L.W. SOBEL, *Artificial Intelligence's Fair Use Crisis*, 41 *Colum. J.L. & Arts* 45 (2017), 53.

604 The dispositive element of Kelly's fair use finding was Arriba's lack of artistic purpose in reproducing Kelly's images. While Kelly's photographs are "artistic works intended to inform and to engage the viewer in an aesthetic experience[,]" Arriba's thumbnails are merely instrumental: they are part of a "tool to help index and improve access to images[.]" Arriba's use is not artistic expression, and, the court reasons, the thumbnails' low resolution makes it unlikely that any user would attempt to consume them for aesthetic, rather than referential, purposes: "The thumbnails do not stifle artistic creativity because they are not used for illustrative or artistic purposes and therefore do not supplant the need for the originals." This formulation of Arriba's purpose undergirded the court's conclusion that Arriba's use was transformative, which in turn allowed Arriba's fair use defense to prevail.

605 *Field v. Google*, 412 F.Supp.2d 1106 (D. Nv. 2006); *Perfect 10 v. Amazon*, 508 F.3d 1146 (9th Cir. 2007); *Authors Guild v. Google*, 770 F.Supp.2d 666 (S.D.N.Y. 2011); *Fox v. TVEyes* (S.D.N.Y. 2014); *White v. West* (S.D.N.Y. 2014); *Authors Guild v. HathiTrust,* 755 F.3d 87 (2d Cir. 2014).

606 *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146 (9th Cir. 2007). Available at: https://cdn.ca9.uscourts.gov/datastore/opinions/2007/12/03/0655405.pdf

607 SOBEL, *Artificial Intelligence's Fair Use Crisis*, cit., 53.

608 *A.V. v. iParadigms, LLC* (4th Cir. 2009). Available at:

609 *Kelly v. Arriba-Soft*, 336 F.3d 811 (9th Cir. 2003)

610 COX, *Text and Data Mining and Fair Use in the United States*, cit.

4.1.1. The four factors and TDM

The following part is aimed at analysing all four factors in the light of TDM's activities, in order to determine in which cases and for what reasons TDM's activities have been considered by the courts to fall within the scope of the Fair Use clause.

4.1.1.1. The purpose and character of the use

The first factor, which has been mostly used in the above mentioned cases to argue that TDM falls within the scope of Fair Use, is the one that looks at "*the purpose and character of the use, including whether such use is of a commercial nature or is for non profit educational purposes*"[611]. Since the Supreme Court's decision in *Campbell*[612] in 1994, the main concept used in assessing the purpose and character of the use in light of the first factor was focused on whether or not the defendant's use was "*transformative*".

The transformative nature of a use of protected material, it has been said, must be assessed considering whether the use "*adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message*"[613]. The criterion of transformativeness of the use is not only the first and the key factor and the one with the greatest weight[614] among the four factors in most cases[615], it is also an argument which recurs and has important effects in the evaluation process carried out by looking at the other three factors too[616].

We have seen how, in fact, what can generally be considered the purpose of TDM is the search among the mined material, whether it is composed of data or text, for recurring patterns or the arrangement of large sets of data or text, resulting the appearance of the copied material almost undetectable or at most only minimally found within the final output of the mining process. Indeed, American courts have recognised this conclusion and held that the creation of a searchable database or search engines has a highly transformative character so that the presence of this first factor has repeatedly persuaded courts to affirm that TDM is an activity qualifying as Fair Use[617].

---

611 17 U.S.C. § 107(1) (2012).

612 *Campbell*, 510 U.S. at 579;

613 *Campbell*, 510 U.S. at 579 (citing P. N. LEVAL, *Toward a Fair Use Standard*, 103 *Harv. L. Rev.* 1105, 1111 (1990)). "The use must be productive and must employ the quoted matter in a different manner or for a different purpose from the original. A quotation of copyrighted material that merely repackages or republishes the original is unlikely to pass the test; in Justice Story's words, it would merely "supersede the objects" of the original. If, on the other hand, the secondary use adds value to the original — if the quoted matter is used as raw material, transformed in the creation of new information, new aesthetics, new insights and understandings — this is the very type of activity that the fair use doctrine intends to protect for the enrichment of society".

614 SOBEL, *Artificial Intelligence's Fair Use Crisis*, cit., 50.

615 See in this respect N. W. NETANEL, *Making Sense of Fair Use*, 15 *Lewis & Clark L. Rev.* 715 (2011), in which the rise to prominence of the transformative use paradigm is traced; see also C. D. ASAY, *Is Transformative Use Eating the World*, 60 *B.C. L. Rev.* (2020), who argues that about 80% of fair use opinions since Campbell have applied the doctrine of transformative use and that this figure rises to 90% in the period starting from 2011.

616 M. SAG, *The New Legal Landscape for Text Mining and Machine Learning*, 66 *J. Copyright Soc'y of the U.S.A.* 291 (2019), 315.

617 COX, *Text and Data Mining and Fair Use in the United States*, cit.

A famous example is *Authors Guild v. HathiTrust*[618], in which Authors Guild[619] brought before the court the case of the creation of a digital library by HathiTrust[620]. The case involved a collaboration between eighty libraries and other institutions aimed to combine the digital copies they had each received from Google to create the HathiTrust Digital Library ("HDL"), with a collection of more than ten million works. The HDL made three uses of these copies. First, it provided a search service that allowed a user to identify relevant works that responded to his query; the search results displayed only the page number(s) of the relevant works on which the search term appeared. Second, for users with certified print disabilities, HDL members could make the full text of the works available through adaptive technologies. Third, a member could make a replacement copy for his or her collection if he or she originally owned the title, and that copy had been lost, destroyed or stolen, and a replacement copy was not available at a fair price[621].

The plaintiff alleged that the defendant used TDM to recognise and locate information sources, thereby making reproductions, claiming copyright infringement. A federal court initially ruled against the plaintiffs in October 2012, holding that the use of HathiTrust was permissible under fair use[622]. The plaintiffs then appealed the decision to the Second Circuit, an appeal that was rebuffed in 2014. The Second Circuit largely upheld the lower court's fair use findings for accessibility and search.

The lower court noted that:

"*The MDP[623] was undertaken with several goals in mind. The MDP allows scholars to identify relevant works far more efficiently*". "*In addition, the program helps Defendants preserve their collections in the face of normal deterioration during circulation, natural disasters, or other catastrophes that decimate library collections, as well as loss due to theft or misplacement*". "*The program provides print-disabled individuals with "access to the wealth of information within library collections*"".

The court continued, saying that "*transformative uses are likely to satisfy the first factor*", defining a transformative use as "*one that actually changes the original work*", or "*one that serves an entirely different purpose*".

Applying this definition to the case, the court held that "*the use to which the works in the HDL are put is transformative because the copies serve an entirely different purpose than the original works: the purpose is superior search capabilities rather than actual access to copyrighted material".* It then noted how *"the search capabilities of the HDL have already given rise to new methods of academic inquiry such as text mining*".

Several cases were cited in order to support this view, stating the court that "*several courts have upheld wholesale copying of works where the use and purpose for the copies was clearly distinguishable from those of the original*". The first case cited was the already mentioned case concerning anti-plagiarism software, *A.V. v. iParadigms, LLC*, 562 F.3d 630, 640 (4th Cir.

---

618 *Authors Guild v. HathiTrust*, 755 F.3d 87 (2d Cir. 2014). Available at:

619 Authors Guild is an organization for writers, supporting them in the defense of their copyrights by representing them in copyright infringement cases. See: Website of Author's Guild, *Who We Are*. Available at: https://www.authorsguild.org/who-we-are/.

620 The HathiTrust Digital Library (HDL) is a spin-off of the Google Books Library Project, founded in 2008 by the Committee on Institutional Cooperation and the University of California system.

621 CARROLL, *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful*, cit., 917.

622 See *Authors Guild v. HathiTrust*, 902 F.Supp.2d 445 (SDNY October 10, 2012). Available at: https://cases.justia.com/federal/district-courts/new-york/nysdce/1:2011cv06351/384619/156/0.pdf

623 The Mass Digitisation Project is referred to in the Authors Guild, Inc. et al v. Hathitrust et al case is a project through which Google creates digital copies of works in the Universities' libraries in exchange for which Google provides digital copies to Hathitrust.

2009), in which the court concluded "*that copying and archiving of student papers "was completely unrelated to expressive content and instead aimed at detecting and discouraging plagiarism*"". The documents were copied here in their entirety and without modification but were only used to generate metadata in the form of plagiarism reports. The Fourth Circuit pointed out that "*iParadigms' use of plaintiffs' works had an entirely different function and purpose than the original works*", thus recognising that, although the papers had been copied in their entirety, greater importance should be accorded to the fact that the purpose of that copying was completely independent of the expressive value of the works themselves[624].

It also brought up the two cases, already mentioned, of thumbnails of protected material being used by search engines to then redirect the user to the page with the original content. The first case is *Perfect 10 v. Amazon.com, Inc.*, 508 F.3d 1146, 1165 (9th Cir. 2007), in which the court found "*that Google's copying of Internet content to make it searchable was transformative because "a search engine transforms the image into a pointer directing a user to a source of information*[625]"". Finally, it cited *Kelly v. Arriba Soft Corp.,* 336 F. 3d 811, 819 (9th Cir. 2003), in which the court found "*that copying to produce exact replicas of artistic works displayed in thumbnail form on the internet to facilitate searches was transformative because it was "unrelated to any aesthetic purpose*"". These thumbnails were in a certain way complete copy, but their size and quality suggested that they would not be a substitute for the original images. These same characteristics also made it clear that the thumbnails were used as a pointing device towards the original images, a completely different purpose from that of the original photos[626].

The cases cited by the court are also important for another reason: they show that although the use of the protected material was of a commercial nature, contrary to the case of HathiTrust which was instead a non-profit company, the compliance with the first factor and the fact that the use was a transformative one, on their own could lead to the position supported by the defendant. In particular, the court held that: "*Defendants satisfy the first factor not merely because they are non-profit institutions, but because the use to which the copies have been put is transformative*"[627]. Indeed, despite an isolated judgment, *Sony Corp. of America v. Universal City Studios, Inc.*, in which the Supreme Court stated that "*every commercial use of copyrighted material is presumptively an unfair exploitation*", where the case itself did not concern a commercial use and the *dictum* was not the final word of the court, in *Campbell*, the Court ruled differently, observing that: "*Congress could not have intended*" a broad presumption against commercial fair uses, as "*nearly all of the illustrative uses listed in the preamble paragraph of § 107 [...] are generally conducted for profit in this country*". In *Campbell*, the Court rejected the notion that commerciality by itself had any "*hard presumptive significance*". Thus, it has been argued that the commercial/non-commercial distinction may be relevant in the assessment of the fourth factor and may certainly be a factor in favour of the qualification of fair use under the first factor, but there is no presumption against commercial use. This is also confirmed by the Second Circuit in *Google Books*, which holds that there is no reason "*why [a defendant's] overall*

---

624 SAG, *The New Legal Landscape for Text Mining and Machine Learning*, cit., 316.

625 The court in *Perfect 10*, 508 F.3d at 1165, goes on to say: "Just as a parody has an obvious claim to transformative value because it can provide social benefit, by shedding light on an earlier work, and, in the process, creating a new one, a search engine provides social benefit by incorporating an original work into a new work, namely, an electronic reference tool. Indeed, a search engine may be more transformative than a parody because a search engine provides an entirely new use for the original work, while a parody typically has the same entertainment purpose as the original work".

626 SAG, *The New Legal Landscape for Text Mining and Machine Learning*, cit., 315.

627 Another case, again involving Authors Guild, but this time against Google, obviously engaging in commercial activities, confirms this view. See *Authors Guild v Google* No. 13-4829-cv (2d Cir. Oct. 16, 2015). See also: *Fox News v. TVEyes* (S.D.N.Y. 2014).

profit motivation should prevail as a reason for denying fair use over its highly convincing transformative purpose, together with the absence of significant substitutive competition, as reasons for granting fair use"[628].

This clearly shows the different political choice of the US compared to the EU, where, as we have seen, TDM activities carried out for purposes other than scientific research and by entities other than research institutes are viewed less favourably, which is why the application of the exception provided for in Article 4 of Directive 790/2019 can be easily overridden by rightholders, who can reserve in various ways to exclude text and data mining activities on their protected material. In the US, therefore, the copyright infringement of commercial companies, such as Google, Amazon or the countless startups, does not necessarily constitute copyright infringement merely because the use of protected material may be commercial in nature, since case law has repeatedly confirmed that also such uses can be considered transformative and therefore not infringing. This, as repeatedly underlined by the letters of criticism submitted by several stakeholders in the process of discussing the European directive, certainly gives the United States a great competitive advantage and, as a consequence, start-ups will probably choose this system to start their business, considering the greater legal certainty and flexibility it offers.

The Second Circuit, the court to which the losing plaintiff appealed, then upheld the lower court's view, ultimately holding that the "*creation of a full-text searchable database is a quintessentially transformative use [and] 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*".

Then, in October 2015, a different panel of the Second Circuit decided the case *Authors Guild v. Google* in favour of Google, stating that Google's copying of all the text of the books contained in a library to create a search index was to be deemed "*highly transformative*", confirming that the unauthorised reproduction of protected material by Google Books is a transformative fair use of the texts, mainly because Google Books provides information "*about*" books, not the books' expression[629].

The Court held here that:

"*As with HathiTrust (and iParadigms), the purpose of Google's copying of the original copyrighted books is to make available significant information about those books, permitting a searcher to identify those that contain a word or term of interest, as well as those that do not include reference to it. In addition, through the ngrams tool, Google allows readers to learn the frequency of usage of selected words in the aggregate corpus of published books in different historical periods. We have no doubt that the purpose of this copying is the sort of transformative purpose described in Campbell*".

From the analysis just made of the American case law, therefore, it is clear that TDM is likely to meet the requirements deemed necessary by the first factor to be considered a fair use, as it should be agreed that, using the words of the Supreme Court in *Campbell*, the reproductions that occur in the process of TDM do not "*merely supersede the objects of the original work*" but "*instead add something new, with a further purpose or different character*". Indeed, the purpose of text and data mining is to generate valuable information about a work or a set of works, which is totally different from what is expressed by the works. As Sag argues, the fact that the insights that are achieved from TDM activities cannot, in most cases, be derived from a mere reading of the material used for mining confirms and reinforces this conclusion[630].

---

[628] SAG, *The New Legal Landscape for Text Mining and Machine Learning*, cit., 322.
[629] SOBEL, *Artificial Intelligence's Fair Use Crisis*, cit., 54.
[630] SAG, *The New Legal Landscape for Text Mining and Machine Learning*, cit., 316.

However, in the judgments of the courts, the four factors must be considered all together and are weighed against each other, which means that satisfying the first requirement does not automatically lead to a fair use conclusion, having to proceed to analyse the other three determining factors too.

4.1.1.2. The nature of the work

The second factor taken into consideration by the courts is not considered to be as important as the first[631]. Most of the cases decided by the courts, in which activities comparable to text and data mining were discussed, involved creative work, which is often given greater protection. Some courts, as also mentioned in *Authors Guild v. HathiTrust,* have held that this factor strongly favors the plaintiff's position, such as in *Campbell*, 510 U.S. at 586, where the court held that *"[S]ome works are closer to the core of intended copyright protection than others*", or in *Blanch v. Koons*, 467 F.3d 244, 256 (2d Cir. 2006)[632], where the court held that copying factual works is more likely fair use than copying creative works.

In other decisions, on the other hand, it has been held that this second factor is to be considered, for the purposes of qualifying a use as fair, either neutral or in favour of the defendant[633]. In fact, in the *AuthorsGuild vs. HathiTrust* judgment cited above, it is argued that, again quoting *Campbell*, 510 U.S. at 586 "*where a use is transformative, the nature of the copyrighted works is not likely to "separate the fair use sheep from the infringing goats*", finally arguing how "*because the use is transformative, intended to facilitate key-word searches or access for print-disabled individuals, the second factor is not dispositive*", citing, to reinforce its argument, the case of the *Bill Graham*, 448 F.3d at 612[634] in which it was stated that "*[t]he second factor may be of limited usefulness where the creative work of art is being used for a transformative purpose*".

It does not therefore seem to be necessary to dwell on this second factor as it is not as decisive as the others.

4.1.1.3. The amount and substantiality of the portion used in relation to the copyrighted work as a whole

The third fair-use factor evaluates whether the amount of copying is reasonable in relation to the purpose of the use. This factor, like the previous one, has been seen by most of the Courts as rather neutral, with some rulings even stating that the copying of the whole work in databases for TDM can be considered as reasonable. Also this factor must be analysed together with the first and fourth ones. It has to be assessed, in particular, whether the amount of material used is appropriate for the purpose already analysed in the light of the first factor, and if there could be economic effects of the use in the traditional market of the copyright owner. When the use can reasonably be considered transformative under the first factor and the amount of material is appropriate, then the use is likely to be considered fair, since, as we shall see shortly, courts give little weight to the impact a transformative use may have on rightholders' lost licensing opportunities[635].

---

631 See, for example: B. BEEBE, *An Empirical Study of U.S. Copyright Fair Use Opinions, 1978-2005,* 156 U. Pa. L. Rev. 549, 610-11 (2008), arguing that the second fair use factor is not decisive.

632 *Blanch v. Koons*, 467 F.3d 244, 256 (2d Cir. 2006). Available at: https://cyber.harvard.edu/people/tfisher/IP/2006%20Blanch%20Abridged.pdf

633 COX, *Text and Data Mining and Fair Use in the United States*, cit.,

634 *Bill Graham Archives v. Dorling Kindersley Ltd.,* 448 F.3d 605, 612 (2d Cir. 2006). Available at: https://cite.case.law/f3d/448/605/

635 CARROLL, *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful*, cit., 945.

We have seen in the first chapter that the greater the amount of material analysed, the better the output of the TDM process. TDM would therefore theoretically require the copy of the whole text or work into the database to be properly processed. Being impossible to know in advance, in the phase of constitution of the corpus to be mined, how the material to be analyzed will be variously combined, without making a copy of the work in its entirety, a researcher could not efficiently use TDM because potentially necessary portions of the work would not be analyzed[636].

In *Sony*, 464 U.S.[637] at 449-50 it was held that *"[T]he extent of permissible copying varies with the purpose and character of the use*". The court thus effectively emphasised that the factors should not be considered separately but rather as a whole, also implicitly suggesting that a copy of an entire text not necessarily should lead the court to consider the use as unfair. In this regard it is useful to cite a further case, *Sundeman v. Seajay Soc'y, Inc.*, 142 F.3d 194, 206 (4th Cir. 1998)[638], in which the court held that there was fair use, where the case concerned a copy of an entire fragile manuscript so that the author of a critical review could study it without inflicting damage.

Moreover, in *Campbell*, 510 U.S. at 586–87 in order to consider the use as fair the principle was stated that "*no more was taken than necessary*". As we have just said, however, in some cases, such as text and data mining, sometimes it is necessary to copy entire works. In the usual case of *Authors Guild vs HathiTrust*, the court states, in this regard, how "*here, entire copies were necessary to fulfill Defendants' purposes of facilitation of searches and access for print-disabled individuals*". It also argues this by referring to *Arriba Soft*, 336 F.3d at 821, citing the part in which the court in that case stated: "*If Arriba only copied part of the image, it would be difficult to identify it, thereby reducing the usefulness of the visual search engine*", and thus leaning towards the qualification of TDM as fair use as regards the second factor, or at least towards its neutrality.

This is further confirmed by the more recent *Google Books*, in which the court of appeals concluded that:

"As with HathiTrust, not only is the copying of the totality of the original reasonably appropriate to Google's transformative purpose, it is literally necessary to achieve that purpose. If Google copied less than the totality of the originals, its search function could not advise searchers reliably whether their searched term appears in a book (or how many times)".

4.1.1.4. The effect of the use upon the potential market for or value of the copyrighted work

The fourth factor examines whether the unauthorised use leads to substantial harm to the market of the copyrighted work, or, as *NXCIM Corp.* 364 F.3d put it, at 482 "*whether the secondary use usurps the market of the original work*". The main focus of this inquiry is therefore to see whether the use of the material has the capacity to replace the work of the copyright owner or its derivative in the marketplace. The courts, in order to assess the impact on the market, also look at the impacts that the use may have on the market of the licences of the copyright owner, with some limitations[639], which will be better seen in a moment[640].

---

636 COX, *Text and Data Mining and Fair Use in the United States*, cit.,

637 *Sony Corp. of America v. Universal City Studios, Inc.,* 464 U.S. 417 (1984). Available at: https://supreme.justia.com/cases/federal/us/464/417/

638 *Sundeman v. Seajay Soc'y, Inc.,* 142 F.3d 194, 206 (4th Cir. 1998). Available at: https://caselaw.findlaw.com/us-4th-circuit/1286064.html

639 E.g., *Am. Geophysical Union v. Texaco, Inc.*, 60 F.3d 913, 930 (2d Cir. 1994): "[o]nly an impact on potential licensing revenues for traditional, reasonable, or likely to be developed markets should be legally cognizable."

640 CARROLL, *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful*, cit., 946.

Again, in analysing this factor, the degree of transformativeness of the use of the protected material must certainly be taken into account, since the greater the transformative aspect of the use, the less likely it is that such use will affect and damage the work-market in which the rightholder derives his profits[641]. As we have said, in the case of TDM, the courts have repeatedly emphasised the transformative and non-expressive nature that TDM makes of the protected material. This suggests that the use of the material for TDM purposes is necessarily less likely to have an adverse impact on the market of the original because it is unlikely that the use could supersede the copyrighted work[642].

The *Authors Guild v. Google Inc.* case is particularly relevant to this factor as well, as it uses the concept of non-expressive use to circumscribe the "*potential market*" for a copyrighted work in a way that the image search engine cases did not. Indeed, the aforementioned *Kelly* and *Perfect 10* did not spend much time discussing the types of markets that may exist for a work and which of these markets the copyright owner has the right to control. Indeed, in *Kelly*, it was emphasized that there was little likelihood of significant harm to the market, Arriba's thumbnails being a redirect of users to plaintiff's site, rather than a damage to plaintiff's business. Similarly, *Perfect 10* concluded that the factor did not favour any party because the evidence of market harm remained "*hypothetical*". On the contrary, *Authors Guild* explicitly notes that Google Books may also harm authors' markets, but such harm "*will generally occur in relation to interests that are not protected by the copyright*". *Authors Guild* suggests that, for purposes of the fourth fair use factor, the relevant potential market includes only consumers' interest "*in the protected aspect of the author's work*"[643].

In *Authors Guild vs. HathiTrust*, in which the court discusses a case of use of protected material for non-commercial purposes, it is argued, referring to pronouncement *Sony*, 464 U.S. at 451, that the plaintiff must show "*by a preponderance of the evidence that some meaningful likelihood of future harm exists*". Also, in *AuthorsGuild v. HathiTrust*, the Court held, in response to plaintiff's argument that *"[e]ach digital copy of a book that Defendants created [...] rather than [purchased] through lawful channels, represents a lost sale*", that the purchase of an additional copy would allow neither full-text searches nor access for individuals with print disabilities, both of which are considered transformative uses.

A second argument was then put forward by the Plaintiff, who claimed that Defendants had "*expose[d] Plaintiffs' property to immense security risks that have the potential to cannibalize the book market through [...] widespread internet piracy*". Defendants responded by illustrating the security measures in place, and by showing that their records were not accessible to the public. For this reason, the court rejected the plaintiff's argument according to which this could have a future effect upon his market.

Finally, going back to the concept mentioned at the beginning of this paragraph about the impact that use may have on the licensing market, in the present case Plaintiffs argued that "*Defendant's activities will harm Plaintiffs by undermining existing and emerging licensing opportunities*" such as a "*collective management system [which would] permit certain of the activities of the Defendants in this case while providing compensation to copyright owners*". The court, however, rejected the plaintiffs' argument, holding that, although the Plaintiffs had correctly cited the Second Circuit's decision in *Texaco*, 60 F.3d at 930, which provided that one must consider the "*impact on potential licensing opportunities*", they failed to cite the second part of the citation, according to which courts should consider "*only traditional, reasonable or likely to be developed markets*". That said, the court held -again highlighting the transformative nature of HathiTrust's use of the

641 See *Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994)*, at 592: "[W]hen [...] the second use is transformative, market substitution is at least less certain, and market harm may not be so readily inferred."
642 COX, *Text and Data Mining and Fair Use in the United States*, cit.
643 SOBEL, *Artificial Intelligence's Fair Use Crisis*, cit., 55.

protected material, i.e., the provision of search capabilities and access for print-disabled individuals- that no harm emerges "*to a "transformative market*"", citing *Bill Graham*, 448 F.3d at 614, where it appears that "*a use that "falls within a transformative market*" does not cause the copyright holder to "*suffer market harm due to the loss of license fees*"".

As Carroll correctly notes, further demonstrating the interconnectedness and necessary joint reading of the four factors, the limitation to the argument of harm to the licensing market applies most frequently where the use is not transformative. In fact, he notes how "*even when a use involves copying enough expressive content that one might expect to need a license, if the copyright owner has no plausible economic reason to develop and support a licensing scheme to monetize such a use, then the fourth factor generally favors the user*"[644].

4.1.2. Conclusions on the qualification of TDM as Fair Use

Putting all four factors together and considering them in conjunction with each other, the American courts have therefore held that TDM's use of the protected material is a highly transformative use, not capable of being considered substitute for the original works, even if integral copies of the protected material are made[645].

In conclusion, it is useful to underline the close connection between fair use and American copyright justification. Again, taking *AuthorsGuild v. HathiTrust* as an example, the court held, citing *Bill Graham*, 448 F.3 d at 608, that "*The totality of the fair-use factors suggests that copyright law's "goal of promoting the Progress of Science [...] would be better served by allowing the use than by preventing it*"", thus finding the defendants' copying to be fair although the plaintiffs complained a *prima facie* case of infringement.

4.2. Text and Data Mining in Japan: the broad specific exceptions in Japanese copyright law

A country that is often cited as a virtuous example in terms of regulation of text and data mining is Japan, which is even described by some as the "*paradise for machine learning and TDM*"[646]. Japan, unlike the United States, has chosen to adopt a specific exception for text and data mining activities. In fact, Japanese copyright belongs to the droit d'auteur system, thus providing for a catalogue of specific exceptions and without a general clause such as the fair use clause or the fair dealing provisions.

Nevertheless, here too, as in the UK and the EU, a discussion has been going on since 2007 about the possibility and desirability of including an open clause along the lines of the US fair use in the Japanese system[647]. The beginning of this discussion about the provisions on limitations of rights conferred by Japanese copyright law can be traced back to a presentation held by Tatsuhiro Ueno at the Copyright Research and Information Center in Tokyo on 19 September 2007. When on May 24, 2008, the Copyright Law Association of Japan held a symposium on "*limitation of rights*", chaired by Ueno himself, the discussion about

644 CARROLL, *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful*, cit., 946.
645 COX, *Text and Data Mining and Fair Use in the United States*, cit.
646 T. UENO, *The Flexible Copyright Exception for 'Non-Enjoyment' Purposes – Recent Amendment in Japan and Its Implication*, GRUR Int. 2021, 149.
647 See, for example: T. UENO, *Rethinking the Provisions on Limitations of Rights in the Japanese Copyright Act, Toward a Japanese-style "Fair Use" Clause*, in AIPPI Journal, July 2009. Available at: http://www.f.waseda.jp/uenot/Japanese-style%20fair%20use%20AIPPI%202009.pdf; Y. CHANG, *Dabates on Introduction of "Fair use" to the Copyright Act of Japan and Korea. Do Japan and Korea need Fair use?*, Comparative IP Academic Workshop Working Paper No. 2, 2009. Available at: http://www.win-cls.sakura.ne.jp/pdf/22/16.pdf

a Japanese-style "*fair use*" clause developed rapidly. In fact, shortly afterwards came the publication of "*The Intellectual Property Strategic Program*"[648], published on 18 June 2008 by the Intellectual Property Strategy Headquarters of the Japanese Government, which proposed a discussion about the possibility of introducing a general clause on copyright exceptions, which stated that:

"[i]n order to promote the content industry in the increasingly digitized and networked world, it is urgent to discuss such issues as the future intellectual property system that can handle newly developed technologies and ways of exploitation and the reinforcement of measures against illegal exploitation of works on the Internet".

According to this document, the discussion should have included "*such topics as the framework to promote distribution in consideration of new ways of content exploitation and the introduction of comprehensive provisions on limitations of rights".*

This discussion was led by the Specialized Committee on the Intellectual Property System for the Digital-Network Generation, of which Ueno himself was a member, set up by the aforementioned Headquarters. At the end of November 2008, the Committee finally issued a report entitled "*Intellectual Property System for the Digital-Network Age*"[649], which concluded that:

"[w]hile the current Act has specific provisions on limitations of rights that stipulate specific cases only in which copyrights shall be limited, it would be appropriate to add to those provisions a general clause on limitation of rights (Japanese-style "fair use" clause) that comprehensively permits fair use of a work to the extent that does not unreasonably prejudice the legitimate interests of right holders"[650].

The details of such a provision were being discussed by the Subcommittee on Legal Affairs of the Subdivision on Copyright of the Council for Cultural Affairs, but in the end, nothing came out of it.

Instead of a fair use provision, a large number of specific provisions on copyright exceptions were introduced into the Japanese Copyright Act.

The first of these were added by the 2009 Amendment[651], which introduced the copyright exceptions regarding internet auctions (Art. 47-2); web search engines (Art. 47-6 before the enactment of the 2018 Amendment); text-and-data mining (Art. 47-7 before the enactment of the 2018 Amendment); and browser caching (Art. 47-8 before the enactment of the 2018 Amendment) etc. Then came the 2012 Amendment[652], which introduced copyright exceptions regarding incidental inclusion (Art. 30-2), and experiments for technological development (Art. 30-4 before the enactment of the 2018 Amendment).[653]

Japan was thus the first country in the world to include a specific exception for text and data mining in its copyright legislation, as it decided already in 2009 to amend its copyright laws to enable text and data mining (TDM). This first amendment introduced a

---

[648] Intellectual Property Strategic Program 2008. Strengthening the Intellectual Property Strategy, Targeting the World. Intellectual Property Strategy Headquarters, June 18, 2008. Available at: http://www.kantei.go.jp/jp/singi/titeki2/keikaku2008_e.pdf

[649] See Digital Net Jidai niokeru Chitekizaisan Senmonchôsakai, Digital Net Jidai niokeru Chizaiseido no Arikata nitsuite (Report) [Intellectual Property System for the Digital-Network Age] 11 (2008) (in Japanese).

[650] UENO, *Rethinking the Provisions on Limitations of Rights in the Japanese Copyright Act, Toward a Japanese-style "Fair Use" Clause*, cit., 195-196.

[651] Act No 53 of 19 August 2009.

[652] Act No 43 of 27 June 2012.

[653] UENO, *The Flexible Copyright Exception for 'Non-Enjoyment' Purposes – Recent Amendment in Japan and Its Implication*, cit., 147.

new Article 47(7), which authorised TDM by all users and for all purposes (commercial and non-commercial).

The article provided:

Article 47(7)**:** (Reproduction, etc. for information analysis)

> *For the purpose of information analysis ("information analysis" means to extract information, concerned with languages, sounds, images or other elements constituting such information, from many works or other much information, and to make a comparison, a classification or other statistical analysis of such information; the same shall apply hereinafter in this Article) by using a computer, it shall be permissible to make recording on a memory, or to make adaptation (including a recording of a derivative work created by such adaptation), of a work, to the extent deemed necessary. However, an exception is made of database works which are made for the use by a person who makes an information analysis*[654].

The exception, if compared with the first TDM provisions entered into force in Europe (and, perhaps, also with the current one resulting from Directive 790/2019) seemed to be already rather broad. First of all, as can already be seen from the wording of the rule, the term "*information analysis*" was defined rather extensively, encompassing various and undefined activities on different types of material. More importantly, it was not limited to activities with non-commercial purposes, as was the case with the first specific exceptions for TDM in European countries or the previous exceptions for scientific research. Moreover, the information analysis did not necessarily have to be carried for (scientific) research purposes; this was in line with the rationale for which the exception was introduced, namely to boost the digital economy in Japan[655].

However, many commentators have also immediately pointed out potential shortcomings to this provision which, perhaps due to its premature introduction, were considered to potentially undermine the successful future application of the Japanese provision. First of all, the acts that were allowed, namely the making of a "*recording on a memory*", suggested that only the reproductions made in the computer's RAM were permitted, which could lead to the consequence that the exception would thus cover and allow only the actual analysis stage in the TDM process and not instead the preceding stages where information is retrieved and stored, which we have seen are most often the most problematic ones. Others have highlighted critical issues related to the fact that there was a risk that the text of the provision, as it was formulated at the time, could be quickly outdated. In fact, the phrase "*by using a computer*" limited the permitted information analysis activities only to those carried with a computer. A further source of uncertainty was the last sentence, which excluded from the application of the exception TDM activities carried out on "*database works which are made for the use by a person who makes an information analysis*". The meaning of this phrase was rather uncertain, but the interpretations of the doctrine held that it exempts the mining of works in databases that are created specifically to use for TDM purposes[656].

Because of the legal uncertainties about which acts of reproduction were covered, about the use of databases and storage of works used for AI purposes, and also because of the rapid growth of IoT, AI, Big Data and Robotics over the past decade, the development

---

654 Copyright Law of Japan, Copyright Research and Information Center, October 2014, 95. Available at: https://www.cric.or.jp/english/clj/doc/20150227_October,2014_Copyright_Law_of_Japan.pdf

655 CASPERS, GUIBAULT, *Baseline report of policies and barriers of TDM in Europe*, cit., 75.

656 *Ibid.*, 75.

of which risked jeopardising the future application of the exception as formulated, it was decided that the rule should be amended to prevent it from becoming obsolete[657].

It is not a secret that Japan has been developing a particularly aggressive strategy in the field of artificial intelligence and big data for some time now, also in view of the fierce competition from China. The country has in fact targeted robotics and artificial intelligence, seeing them as potential solutions to the country's social and economic issues, such as the labour shortage and the ageing population. In line with these aims, former Prime Minister Shinzo Abe presented a plan in March 2017 at the CeBIT expo, held in Germany, to develop a Japanese "*Society 5.0*", which envisages the deep integration of cutting-edge technologies within Japanese society, which would consequently become a *"super-smart society*" capable of providing customized solutions through the adoption of new technologies like Artificial Intelligence (AI), robotics, Big Data, and drones - as well as through policy and regulatory reform. Two months later, in June 2017, Abe himself presented the "*Japan's Growth Strategy 2017*", which lays out a strategic blueprint for Japan's "*Society 5.0*", with specific plans for each of these areas[658]. Within this plan, artificial intelligence plays a major role, that is why back in April 2016 the "*Japan's Artificial Intelligence Technology Strategy Council*" was established, which could support AI research and development in the country, as well as develop a roadmap for AI development, a plan in which industry, academia, and government come together and pursue cooperation in this field[659].

Japan therefore understood and recognised that machine learning and text and data mining are indispensable to the development of artificial intelligence, and wanted to make sure that the legislative environment in which this was to be developed was as suitable as possible to fully support its ambitions in this field. That also meant allowing researchers and private companies to carry out machine learning activities through an update to its copyright rules.

Therefore, the need to reform copyright again, given the new innovative spirit and the already mentioned limitations of the old provision, reignited the discussion about the introduction of more flexible provisions or a general provision on copyright exceptions. In fact, the Intellectual Property Strategy Headquarters, on 9 May 2016, issued the "*Intellectual Property Strategic Program 2016*"[660], again proposing to consider the possibility of introducing flexible provisions on copyright exceptions in order to promote new innovation in the digital network era[661]. However, the Japanese government again considered that a fair use clause along the lines of the American fair use clause could not be easily adapted within a civil law legal model such as the Japanese one. Moreover, much criticism also came from Japanese copyright holders[662].

---

657 EARE, Japan Amends Its Copyright Legislation To Meet Future Demands In Ai And Big Data, *3 September, 2018* https://eare.eu/japan-amends-tdm-exception-copyright/

658 N. JAO, *Japan's 'Society 5.0' will integrate cutting-edge tech at all levels*, ITU News, 29 June 2017. Available at: https://news.itu.int/japans-society-5-0-will-integrate-cutting-edge-tech-at-all-levels/

659 JAPANGOV, the Government of Japan, *Artificial Intelligence: A Rival for Humans, or a Partner?* Available at: https://www.japan.go.jp/tomodachi/2018/spring2018/artificial_intelligence.html

660 Intellectual Property Strategic Program 2016. Intellectual Property Strategy Headquarters. May 2016. Available at: https://www.kantei.go.jp/jp/singi/titeki2/kettei/chizaikeikaku20160509_e.pdf

661 Ueno's proposal, for example, was to introduce a small general basket clause on exceptions to copyright specifying certain factors to be considered at the end of the detailed catalogue of specific provisions on exceptions to copyright (Articles 30 to 49). This, in Ueno's own opinion, would have been a tool to foster to some extent flexibility to respond to variety and change, while maintaining legal certainty.

662 UENO, *The Flexible Copyright Exception for 'Non-Enjoyment' Purposes – Recent Amendment in Japan and Its Implication*, cit., 147.

The Subdivision on Copyright of the Council for Cultural Affairs ("*Bunka-Shingikai*")[663] published a report in April 2017[664], saying that the Japanese government had decided not to introduce a general copyright exception provision like the US one, preferring instead to the latter some *"flexible"* copyright exception provisions.

So, the plan was revised, and a different path followed. In particular, Japan eventually changed the structure of its copyright exceptions from being a "*closed-list*" structure to being a "*semi-open*" one[665]. In particular, copyright exceptions have been compartmentalised into three different types[666], depending on the purposes they serve and market considerations:

1. Harmless uses: uses that do not fall within the original purpose of uses of a copyrighted work and that can be regarded as not normally prejudicial to the interests of rightholders;
2. Minor harm uses: uses that do not fall within the original purpose of uses of a copyrighted work but which cause a lower degree of harm to the rightholder's interests;
3. Public policy uses: uses that fall within the original purpose of uses of a copyrighted work but are permitted to achieve public policy objectives such as cultural development.[667]

The bill to amend the Japanese Copyright Act was submitted to the 196th ordinary Diet session on 23 February 2018, passed in the Diet session on 18 May 2018, and came into force on 1 January 2019[668].

These "*flexible*" copyright exceptions introduced by the 2018 Amendment are found in three different provisions:

- Art. 30-4 (exploitation not for "*enjoyment*" purposes), which now allows all users to analyse and understand copyrighted works by machine learning, accessing data or information in a form in which the copyrighted expression of the works is not perceived by the user and therefore would not cause any harm to rights holders, including raw data that is fed into a computer program to perform deep learning activities, which are the basis of artificial intelligence;
- Art. 47-4 (exploitation incidental to the use of works on a computer), permits electronic incidental copies of works, recognizing that this process is necessary to carry out machine learning activities but does not harm copyright owners;
- Art. 47-5 (minor exploitation incidental to data processing on a computer and the provision of the results produced), which permits the use of copyrighted works for data verification when conducting research, recognizing that such use is important to researchers and is not detrimental to rights holders. This article

663 It is normal in the law-making process in Japan that the Government continuously consults the Councils ('Shingikai') in which experts including scholars, lawyers and sometimes stakeholders take part and basically accepts the proposal of their reports.

664 See Bunka-Shingikai Chosakuken-Bunkakai Hokokusho 2017 [The Report of the Subdivision on Copyright of the Council for Cultural Affairs 2017] (in Japanese). Available at: https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/h2904_shingi_hokokusho.pdf

665 T. HE, *Copyright Exceptions Reform and AI Data Analysis in China – A Modest Proposal*, in J. A. LEE, R. HILTY, K. C. LIU, *Artificial intelligence and Intellectual Property*, Oxford University Press, 2021, 209.

666 Agency for Cultural Affairs, "Cultural Council Copyright Subcommittee Legal Basic Issues Subcommittee Interim Summary" (2017) (in Japanese) www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/h2902_chukanmatome.pdf

667 LEE, HILTY, LIU, *Artificial intelligence and Intellectual Property*, cit., 209.

668 Act No 30 of 25 May 2018.

enables searchable databases, which are necessary to carry out data verification of the results and insights obtained through TDM[669].

It should be noted that each of these "*flexible*" exceptions to copyright contains a small basket clause (the main paragraphs of Arts. 30-4 and 47-4(1)(2), and Art. 47-5(1)(iii)), which means that such an exception can be flexibly applied not only to the specific exploitations in the prescribed articles, but also to any other equivalent exploitation. It can thus be seen as a means to foster a certain degree of flexibility while maintaining legal certainty, even though it is not a general basket clause at the end of a catalogue of specific provisions on exceptions to copyright[670]. This is for sure an attempt worthy of attention for possible future amendments to the European framework.

For the sake of brevity, the provision which interests us most is the exception provided for in Art. 30 (4), that reads as follows:

> Article 30-4 (Exploitation without the Purpose of Enjoying the Thoughts or Sentiments Expressed in a Work)
>
> *It is permissible to exploit a work, in any way and to the extent considered necessary, in any of the following cases, or in any other case in which it is not a person's purpose to personally enjoy or cause another person to enjoy the thoughts or sentiments expressed in that work; provided, however, that this does not apply if the action would unreasonably prejudice the interests of the copyright owner in light of the nature or purpose of the work or the circumstances of its exploitation:*
>
> *(i) if it is done for use in testing to develop or put into practical use technology that is connected with the recording of sounds or visuals of a work or other such exploitation;*
>
> *(ii) if it is done for use in data analysis (meaning the extraction, comparison, classification, or other statistical analysis of the constituent language, sounds, images, or other elemental data from a large number of works or a large volume of other such data; the same applies in Article 47-5, paragraph (1), item (ii));*
>
> *(iii) if it is exploited in the course of computer data processing or otherwise exploited in a way that does not involve what is expressed in the work being perceived by the human senses (for works of computer programming, such exploitation excludes the execution of the work on a computer), beyond as set forth in the preceding two items.*[671]

In practice, the Japanese exception, which is clearly based on the pre-existing exceptions and reconstructed into a multilayered style (general clause - list - catch all)[672], allows the exploitation of a work, by any means and to the extent deemed necessary, if the exploitation is aimed at neither enjoying nor causing another person to enjoy the work, unless such exploitation unreasonably prejudices the interests of the copyright holder[673].

---

669 EARE, *Japan Amends Its Copyright Legislation to Meet Future Demands in AI and Big Data*, 3 September 2018. Available at: https://eare.eu/japan-amends-tdm-exception-copyright/

670 UENO, *The Flexible Copyright Exception for 'Non-Enjoyment' Purposes – Recent Amendment in Japan and Its Implication*, cit., 147.

671 Copyright Law of Japan, CRIC (Copyright Research and Information Center). Available at: https://www.cric.or.jp/english/clj/cl2.html

672 LEE, HILTY, LIU, *Artificial intelligence and Intellectual Property*, cit., 209.

673 UENO, *The Flexible Copyright Exception for 'Non-Enjoyment' Purposes – Recent Amendment in Japan and Its Implication*, cit., 148.

Compared to the two exceptions recently introduced in EU copyright law, the Japanese TDM copyright exception is much broader, clearer and ultimately better.

First of all, it applies not only to TDM activities for a non-commercial purpose but also for commercial purposes, and not only by research organizations but also by business companies[674].

Moreover, in the Japanese TDM copyright exception there is no reference to the contested and uncertain in its meaning European requirement of "*lawful access*"[675].

A further positive element is that, unlike Article 4 of EU Directive 790/2019, even if a copyright holder expressly makes a reservation (opt-out) for an exploitation of a work for commercial TDM, the Japanese TDM copyright exception applies to the exploitation regardless of the existence of the reservation[676].

The exception has also improved the previous one, which we said had been criticised for its alleged possible rapid obsolescence resulting from the phrase "*by using a computer, it shall be permissible to make recording on a memory*". In fact, the exception now provides for the possibility "*to exploit by any means*", thus allowing both digital and analogue copying, but also ensuring that the exception will not become obsolete in case of future possibility to perform TDM by other means than computer[677].

Finally, not only the copying of a work but also its distribution and communication to the public is allowed under the Japanese TDM copyright exception, so that it is possible to copy a number of works for the purpose of machine learning as TDM and then distribute the dataset to other people who wish to make machine learning[678].

The Japanese TDM copyright exception is therefore particularly broad and any exploitation of copyrighted works for TDM purposes is permitted under Art. 30-4(ii). The new Japanese exception is also particularly important because it is a further confirmation of the possibility of seriously reforming the copyright system in the way proposed by European doctrine and analysed at the end of the last chapter. In fact, this new exception has overturned the foundations of Japanese copyright, because it introduced and is now based on the concept of the non-expressive/non-consumptive use of works. Its introduction is due to the recognition by the Japanese legislature that an exploitation for TDM is purely for the purpose of analysing data and is aimed neither at enjoying the work nor causing another person to enjoy the work. If an exploitation of a work is aimed at neither enjoying it nor causing another person to enjoy it there is no need to give the right holder a compensation and thus copyright does not need to cover such exploitation. In other words, exploitation of this kind does not prejudice the copyright holder's interests protected by a copyright law[679].

4.3. Countries that have introduced a Fair Use clause in their copyright law

Although the fair use approach is mostly associated with the United States, a similar approach can be found in many other countries around the world. These countries have thus

674 *Ibid.*, 149.
675 *Ibidem.*
676 *Ibidem.*
677 *Ibidem.*
678 *Ibidem.*
679 *Ibid.*, 150.

decided to adopt a principle of fair use and, as a result, have codified lists of criteria for determining whether or not the use is fair[680].

We will therefore briefly deal with some of the numerous countries that have decided to choose a US-like approach and therefore to include an open-ended clause on the model of the US. We will start by discussing very briefly about the copyright exception laws of Taiwan and South Korea, and we will deal with them together because they represent two similar solutions to the issue and two particularly interesting examples of civil law countries that have managed to implement a Fair Use clause modelled on the US one in their copyright laws. We will then focus our attention on the interesting case of Israel, which, in implementing the US fair use clause in its copyright legislation, has in some respects even improved on the already rather broad US treatment.

4.3.1. Text and Data mining in Taiwan and South Korea

Starting with Taiwan, an Asian civil law country, it has a hybrid system of Exceptions and Limitations, i.e. a list of specific exceptions contained in a specific section entitled "*limitations on economic rights*" (Articles 44 to 64), however accompanied and complemented by a general clause.

Taiwan, in fact, already in 1992, decided to transplant the US model of fair use to its copyright law, at Article 65, which therefore now provides for an open clause largely imitating the US fair use test that can serve as a general principle for determining all circumstances, including the listed exceptions. Under this approach, when the text and data mining process satisfy the four factors (which is highly likely as the use is transformative and normally does not harm the potential market of the copyright owners), it could then be exempted[681]. A particular feature of the Taiwanese fair use clause is that it was revised in 2003 to allow courts to take into account an agreement between copyright holders and users (Art. 65(3) and (4))[682].

Art 65 of the Taiwanese Copyright Act provides[683]:

> *Fair use of a work shall not constitute infringement on economic rights in the work.*
>
> *In determining whether the exploitation of a work complies with the reasonable scope referred to in the provisions of Articles 44 through 63, or other conditions of fair use, all circumstances shall be taken into account, and in particular the following facts shall be noted as the basis for determination:*
>
> 1. *The purposes and nature of the exploitation, including whether such exploitation is of a commercial nature or is for nonprofit educational purposes.*
> 2. *The nature of the work.*
> 3. *The amount and substantiality of the portion exploited in relation to the work as a whole.*
> 4. *Effect of the exploitation on the work's current and potential market value.*

680 M. GEIST, *Fairness Found: How Canada Quietly Shifted from Fair Dealing to Fair Use*, in *The Copyright Pentalogy: How the Supreme Court of Canada Shook the Foundations of Canadian Copyright Law* [online], Ottawa: Les Presses de l'Université d'Ottawa, University of Ottawa Press, 2013. Available at: http://books.openedition.org/uop/969, par. 17.

681 LEE, HILTY, LIU, *Artificial intelligence and Intellectual Property*, cit., 207.

682 UENO, *The Flexible Copyright Exception for 'Non-Enjoyment' Purposes – Recent Amendment in Japan and Its Implication*, cit., 147.

683 Taiwan Copyright Act, Laws and Regulations Database of The Republic of China. Available at: https://law.moj.gov.tw/ENG/LawClass/LawAll.aspx?pcode=J0070017

Where the copyright owner organization and the exploiter organization have formed an agreement on the scope of the fair use of a work, it may be taken as reference in the determination referred to in the preceding paragraph.
In the course of forming an agreement referred to in the preceding paragraph, advice may be sought from the specialized agency in charge of copyright matters.

For the sake of curiosity, it has been noted that Taiwan has adopted a new approach in the 2014 revision of its copyright law. The four fair use factors are now examined by courts in two different contexts. First, when applying Articles 44 to 63, the court must consider the four factors when the statute explicitly requires the particular use to be made within a "*reasonable/equal scope*". Second, the four factors in s. 65(2) are still an independent and open-ended clause for the court to determine fair use, as before[684].

A similar approach to Taiwan is also followed by Korea, which, also being a civil law country, used to adopt a closed list copyright exception model in its copyright law. The articles and exceptions that might have been relevant to TDM at the time were: Article 25 (use for school education), which allows the publication of works made public in school textbooks for primary and secondary schools and the use of part of the works made public by designated educational institutions for teaching purposes under certain conditions; Article 28 (citation), which allows anyone to cite works made public for reporting, criticism, education and research, etc., provided the citation is within a reasonable limit and consistent with fair practice; Article 31 (use by libraries or archives), according to which libraries and archives are exempted from liability for the use of books held by them under certain conditions and circumstances[685].

Among these exceptions, it has been noted, the most widely used was the quotation exception. The purpose of the quotation and the reasonable extent of the quotation have both been interpreted very broadly by the courts. In a number of cases, the possibility of invoking the quotation exception has been held to depend on several factors: the purpose of the quotation, the nature of the quoted work, the content and quantity of the quoted part, the relationship between the quoted work and the quoting work, the general notion of readers, and the possibility of displacement of the market demand for the original work. The factors just mentioned, which apply to the quotation exception, are quite clearly reminiscent of the factors we have already listed when discussing the fair use doctrine in the US. Seen in this way, there seems to be no apparent difference between the quotation exception and the fair use doctrine. However, it is clear that judges in a civil law country such as Korea have been uncomfortable and hesitant to expand the quotation exception too much. For this reason, the quotation exception has instead been denied in many other cases, apparently borderline. Given the fact that the citation exception could not be considered sufficient to strike a good balance between the interests of creators and users, Korea finally decided to introduce a general fair use clause in the Korea Copyright Act[686].

The main change to this situation came with the amendment of the Copyright Act in 2011, when the Copyright Act was amended to comply with the Korea-U.S. Free Trade Agreement (FTA) obligations of 2007. Kor-Us FTA § 18.4.10.(a) provided that*:*

---

684 LEE, HILTY, LIU, *Artificial intelligence and Intellectual Property*, cit., 207.
685 Organization for Economic Co-operation and Development (OECD) Report: Chapter 7 - *Legal Aspects of Open Access To Publicly Funded Research*, in *Enquiries into Intellectual Property's Economic Impact*, 2015, 397. Available at: https://www.oecd.org/sti/ieconomy/Chapter7-KBC2-IP.pdf
686 S. J. JONG, *Fair Use in Korea*, infojustice, 27 february 2017. Available at: https://infojustice.org/archives/37819

"With respect to this Article and Articles 18.5 and 18.6, each Party shall confine limitations or exceptions to exclusive rights to certain special cases that do not conflict with a normal exploitation of the work, performance, or phonogram, and do not unreasonably prejudice the legitimate interests of the right holder".

As a result, therefore, the South Korean legislature introduced a US style fair use test in Article 35-3, which provides:

Article 35-3 (Fair Use of Works, etc.)[687]

(1) Except as provided in Articles 23 through 35-2 and 101-3 through 101-5, where a person does not unreasonably prejudice an author's legitimate interest without conflicting with the normal exploitation of works, he or she may use such works.

(2) In determining whether an act of using works, etc. falls under paragraph (1), the following shall be considered:

1. Purposes and characters of use including whether such use is for or not for nonprofit;

2. Types and natures of works, etc.;

3. Amount and substantiality of portion used in relation to the whole works, etc.;

4. Effect of the use of works, etc. on the current or potential market for or value of such work etc.

According to this article, it is therefore possible to make use of protected works for reporting, criticism, education and research, etc., when such use does not conflict with the normal exploitation of the works and does not unreasonably prejudice the author's legitimate interests[688]. It has been seen as a combination of the three-step test and fair use[689].

As opposed to Taiwan's approach, however, Korean copyright law seems to want to distinguish the listed exceptions (Articles 23 to 35-2 and 101-3 to 101-5) with its fair use test, so that if an act is covered by one of the listed exceptions, then there would be no room for application of the fair use test in Article 35-3. On the other hand, similar to the situation in Taiwan, none of the exceptions listed in the Korean copyright law can completely cover the use of TDM[690].

Following the amendment, although most commentators acknowledged the necessity of the general copyright exception provision, some noted that there is, however, a major drawback to introducing such a general clause into the law of a civil law country like South Korea: the fair use doctrine has been established on the basis of a huge number of precedents for a long time in the United States. This, the commentators feared, may lead to uncontrolled situations due to the lack of judicial precedents. Indeed, the Korean Copyright Act enumerates exceptions and applies the provisions to cases of copyright infringement on a

687 Translation provided by: Korea Law Translation Center – Korea Legislation Research Institute. Available at: https://elaw.klri.re.kr/eng_service/lawView.do?hseq=42726&lang=ENG

688 Organization for Economic Co-operation and Development (OECD) Report: Chapter 7 - *Legal Aspects of Open Access to Publicly Funded Research*, cit., 397.

689 Y. CHANG, *Debates on Introduction of "Fair use" to the Copyright Act of Japan and Korea. Do Japan and Korea need Fair use?,* Comparative IP Academic Workshop Working Paper No. 2, 2009, 288. See also: P. K. YU, *Customizing Fair Use Transplants*, in Laws, Vol. 7, Issue 1, Article 9, 2018, Texas A&M University School of Law Legal Studies Research Paper No. 17-78, 6.

690 LEE, HILTY, LIU, *Artificial intelligence and Intellectual Property*, cit., 208.

case-by-case basis. These two legal systems are so different that fair use cannot be consistent with the Korean legal system[691].

In conclusion, the use of copyrighted works in TDM process will be evaluated by the parties to the fair use test within these two jurisdictions. However, it is now really difficult to predict how courts in Taiwan and South Korea will decide those cases involving TDM of copyrighted works based on their open fair use provisions, as the four-factor test is a legal transplant in itself, and the interpretation of the four factors will need to serve local needs, and thus may vary from country to country. We have seen how in the US, the transformative nature of non-expressive uses of copyrighted works in google books project and search engines have been considered fair use. Following this trend, also in these two countries text and data mining could be regarded as fair use, as they are presumptively "*non-expressive*" and transformative, and therefore do not harm the potential market of the copyright owner[692].

4.3.2. Text and Data Mining in Israel

At the establishment of the State of Israel in May 1948, the Law and Administration Ordinance declared that all previous legislation would become part of Israeli law, subject to the changes required by the creation of the state. Thus, the Israeli legal system is now composed of remnants of Ottoman law (in force until 1917), laws of the British Mandate, incorporating a large body of English common law, elements of Jewish religious law and some elements of other systems.

Israeli copyright law is thus strongly based on the British construction of copyright law and its principles with considerable and different influences over the last century. Initially in the 1980s, when the first continental influences led to the recognition of moral rights; then in the 1990s, with the increasing Americanisation of the law. It was during this period that English copyright rules were reinterpreted with an American mindset by Israeli courts[693].

Being inspired by the English fair dealing legislation, which we have already examined in the third chapter, Section 2(1) of the Israeli 1911 Act permitted fair dealings with a work. In order to qualify for the defense, the defendant had to be able to convince the court of two cumulative requirements:

1. that the use fell within one of the enumerated purposes listed in sec. 2(1) - that is: private study, research, criticism, review or newspaper summary- being consequently a dealing not fitting one of these purposes surely not fair;
2. the use needed to be fair, a term left undefined in the Act[694].

The key difference between US fair use and UK-inspired fair dealing lies in the limited purposes of fair dealing. Unlike the open fair use model, fair dealing models typically identify specific categories of purposes for which fair dealing is allowed. The model creates, therefore, a two-stage analysis: first, whether the intended use qualifies for one of the permitted purposes, and second, whether the use itself meets the fair dealing criteria. On the contrary, fair use raises only the second stage of the analysis, since there are no statutory limitations on permitted purposes. Given the possibility that some uses may fall outside the fair dealing area of the circumscribed purposes, some fair dealing countries have begun to consider

691 CHANG, *Debates on Introduction of "Fair use" to the Copyright Act of Japan and Korea. Do Japan and Korea need Fair use?*, cit., 288.
692 LEE, HILTY, LIU, *Artificial intelligence and Intellectual Property*, cit., 208.
693 M. D. BIRNHACK, *Mandatory Copyright: From Pre-Palestine to Israel, 1910-2007*, In U. SUTHERSANEN, Y. GENDREAU (eds.), *A Shifting Empire: 100 Years of The Copyright Act 1911*, Edward Elgar, 2012, 16.
694 BIRNHACK, *Mandatory Copyright: From Pre-Palestine to Israel, 1910-2007*, cit., 25.

whether to adopt fair use provisions or expand their fair dealing criteria. Among these countries we can include Israel[695].

In fact, in 2007, the State of Israel approved a reform to the previous Israeli Copyright Act of 1911 (as set via the British Copyright Act of 1911) which, with regard to the possibility of legally conducting activities in derogation of copyright (including today's TDM activities), introduced an extensive number of additional exceptions, covering a number of different uses of works, none of which are directly applicable to TDM activities[696], but, most importantly, shifted its regulation from an approach inspired by the British system of "*fair dealing*" to another approach referring to the American system of "*fair use*".

In fact, Israel's new copyright statute completed the move from fair dealing toward fair use that the Israeli Supreme Court had already initiated in 1993 in its ruling in *Geva v. Walt Disney Co*[697].

In this case, very important as it marks the transition from the European influence on Israeli copyright to the American one, the use of the work successfully passed the first fair dealing test, but in facing the second component of the two-step test (i.e. fairness) the Court found direct inspiration in the U.S. four factors of evaluation of the fair use[698]. The Court engaged in a detailed analysis of the four factors in the case at hand, ultimately finding that the use was commercial, thus weighing against its qualification as fair dealing. In proceeding with the analysis, the Court relied heavily on American case law and fair use literature, as they were back in 1993[699], explicitly stating its preference for the American fair use model over the English fair dealing model[700].

In a subsequent case involving the character of Charlie Chaplin, the Supreme Court underlined that the test of fairness was more important than the first one concerning the purpose[701].

The ultimate result of this operation was the emergence of an artificial hybrid between English fair dealing and American fair use. Under *Geva v. Walt Disney Co*, for instance, the defendant had to pass both the English barrier demonstrating that his use fell within the list of enumerated purposes and the American four-factor barrier, so that the system showed both the rigidity of a permitted use rule and the uncertainty of a four-factor standard[702].

The reform of 2007 resulted at the end in the adoption of an open rule, invocable in a wide range of cases and situations.

Article 19 broadly reproduces Article 107 of the corresponding US Copyright Act, adding some modifications that have been considered by many commentators[703] as improvements of the US fair use clause.

---

695 GEIST, *Fairness Found: How Canada Quietly Shifted from Fair Dealing to Fair Use*, cit.

696 Organization for Economic Co-operation and Development (OECD) Report: Chapter 7 - *Legal Aspects of Open Access to Publicly Funded Research*, cit., 395.

697 C.A. 2687/92 Geva v. Walt Disney Co., 48(1) P.D. 251 (1993).
In 1993, the Supreme Court of Israel faced the issue of "*fair dealing*" referring to the work of an artist, David Geva, who had modeled a new character (Moby Duck) upon Disney's famous Donald Duck, in order to carry out ironic criticism of Israeli society. Geva relied upon exceptions to copyright as the means by which freedom of expression could be upheld. He argued that his use of Disney's character was consistent with the American regime of fair use.

698 Geva, ibid, at 275-83.

699 The Court cited *Sony Corp. v. Universal City Studios, Inc.*, 464 U.S. 417 (1984); *American Geophysical Union v. Texaco*, 802 F.Supp. 1 (S.D.N.Y., 1992); *Walt Disney Productions v. Air Pirates*, 581 F.2d 751 (9th Cir. 1978).

700 *Geva*, at 270.

701 C.A. 8393/96 *Mifal Ha'Pais v. The Roy Export Establishing Co.*, 54(1) P.D. 577, 597 (2000) [Heb.].

702 BIRNHACK, *Mandatory Copyright: From Pre-Palestine to Israel, 1910-2007,* cit., 26.

703 Z. EFRONI, *Israel's Fair Use*, Blog of The Center for Internet and Society at Stanford Law School, 30 January 2008. Available at: http://cyberlaw.stanford.edu/blog/2008/01/israel's-fair-use

On the basis of this ruling, the purposes for which TDM actions may be legally carried out are no longer exhaustively listed in the previously current closed enumeration of allowable purposes, enlisted as follows: "*private study, research, criticism, review or newspaper summary*". They must be identified based on a new "*open"* provision that generally outlines the characteristics, purposes and effects the activity (including TDM) must have in order to be legally carried out independently of the will of the rightsholder, establishing as follows:

> *(a) Fair use in a work is permitted, amongst other things, for these purposes: private study, research, criticism, review, journalistic reporting, quotation, or instruction and examination by an educational institution.*
>
> *(b) In determining whether a use made of a work is fair within the meaning of this paragraph the factors to be considered shall include, inter alia:*
>
> *a. The purpose and character of the use;*
> *b. The character of the work used;*
> *c. The scope of the use, quantitatively and qualitatively, in relation to the work as a whole;*
> *d. The impact of the use on the value of the work and its potential market.*
>
> *(c) The Minister may make regulations prescribing conditions under which a use shall be deemed a fair use*[704].

The latter are, therefore, the purposes that a Court should take into consideration in order to determine whether a dealing is fair or infringing. It is generally recognized that the list is not meant to be exhaustive[705], so that, for example, courts have interpretatively added a further important factor such as the attribution to the author or copyright owner[706].

As to the description of the first fair use-factor (*purpose and character of the use*), the Israeli text does not include the sentence we can find in the homologue American provision "[…] *including whether such use is of a commercial nature or is for nonprofit educational purposes*". This different and more succinct formulation meant to avoid the interpretative embarrassments that had arisen, we have seen, within American Courts which, faced with this apparently restrictive regulatory provision, were inclined to recognise fair use even outside the circle of non-profit educational uses[707]. Therefore, the court should merely consider "*[t]he purpose and character of the use*" a much more open evaluation than a strict transformativeness examination, where transformativeness is not, at the moment, a requirement Israeli Courts are literally bound to[708].

We must also point out the addition of a further subsection (c) which gives to the Minister the possibility to introduce further and new conditions in the presence of which a certain use of the techniques of TDM should certainly be considered "*fair*"[709]. It has been

---

[704] Israel Copyright Law. Available at: https://www.tau.ac.il/law/members/birnhack/IsraeliCopyrightAct2007.pdf

[705] Organization for Economic Co-operation and Development (OECD) Report: Chapter 7 - *Legal Aspects of Open Access to Publicly Funded Research*, cit., 395.

[706] *Ibid.*

[707] EFRONI, *Israel's Fair Use*, cit.: "Courts now should be able to bring powerful normative-qualitative considerations into their fair use analysis without feeling guilty, which is great".

[708] *Ibidem.*

[709] Israel Copyright Act, article 19 Section (c): "The Minister may make regulations prescribing conditions under which a use shall be deemed a fair use […]".

argued that this important provision tends to reduce the margins of uncertainty inherent in a system of fair use, in which those who intend to undertake TDM activities must assess their legality by an *ex ante* prognosis about the possible outcome of lawsuit for infringement hypothetically brought against them by the rightsholder, basing their decision on an approximate forecast about the interpretations of Courts. Subsection (c) grants the minister - who acts by means of essentially administrative and/or high administrative acts (definitely easier to adopt and implement than an act with the force of law) - the possibility of introducing irrefutable written presumptions that lower the margins of uncertainty[710].

4.4. TDM in Australia and Canada. The fair dealing approach.

Let us now analyse the particular cases of two common law countries, Australia and Canada, whose copyright laws do not yet have a specific exception for text and data mining and which, unlike Israel (a mixed legal system, combining aspects of civil law and some of common law), which has moved from fair dealing to fair use without any particular problems, have still remained faithful to the fair dealing approach.

We will see, however, how both countries have been studying and thinking about the introduction and transition to a fair use clause for some time, and how the relationship and approach of the two countries to fair dealing clauses differs considerably, with Australia seeming to be less inclined to extensive interpretations of its fair dealing clauses, and Canada, on the contrary, being more attentive and open to innovation and considering an increasing number of uses as fair dealing, as well as, recently, even considering the possibility of introducing a specific exception for TDM activities.

4.4.1. Text and Data Mining in Australia

Let us therefore start with Australia, a country belonging to the common law tradition which has not yet equipped itself with a specific exception for TDM, but which has been studying the issue and working on a solution for a long time. Australia has often been compared to Canada when it comes to copyright exceptions, as the approach of the two is partly similar, due to the presence, because of English influence, of fair dealing clauses in its copyright legislation, but, on the other hand, at the same time quite different, due to the alleged greater reluctance and closure, especially by the Courts, to consider uses as fair dealing. In fact, unlike the United States and Canada, where, as we have seen and will see, for different reasons, fair use and fair dealing are considered flexible instruments - the Australian fair dealing exception has not received such a broad interpretation from the courts[711][712].

However, Australia has also shown at least some interest in finding more flexible solutions for the regulation of copyright exceptions. Indeed, it has the peculiarity of having tried to introduce the fair use clause in its copyright legislation since 1998. In particular, since that date, the governments that have succeeded one another have asked for opinions on whether such a clause should be introduced eight different times, most of the times obtaining

[710] EFRONI, *Israel's Fair Use*, cit.
[711] See, for example, *National Rugby League Investments Pty Limited v. Singtel Optus Pty Ltd* [2012] FCAFC 59, 27 April 2012.
[712] Organization for Economic Co-operation and Development (OECD) Report: Chapter 7 - *Legal Aspects of Open Access to Publicly Funded Research*, cit., 394.

a positive opinion and a recommendation to introduce it[713]. The last two consultations are of particular interest to us, as they also directly addressed the issue of TDM.

The first occasion on which this occurred was during the consultation that began in 2012, in which the Australian Law Reform Commission received the "*Terms of Reference*"[714] for the Copyright Inquiry and was asked to consider whether the statutory exceptions and licences in the Copyright Act 1968 were adequate and appropriate in the digital environment and whether further exceptions should be recommended. The final report was submitted to Parliament on 13 February 2014. The report discusses at length the comparative advantages and disadvantages of introducing a fair use defence or modifying the fair dealing defence, containing 30 recommendations for reform. The key recommendation is for the introduction of a fair use exception to Australian copyright law[715].

Indeed, the ALRC recommended at that time to introduce into the Copyright Act an exception for fair use, complete with an explicit statement that fair use of copyrighted material does not infringe copyright, and a non-exhaustive list of factors to be considered in determining whether the use is fair use ("*the fairness factors*"). In the event that this proposal was rejected, as an alternative, it was proposed to amend the fair dealing provisions to cover additional and different purposes permitted by it, including, in (h), incidental or technical use, which included text and data mining[716][717].

This proposal for the introduction of a fair use clause in Australian copyright also received, in the subsequent 2016 consultation, the approval and support of the Australian Productivity Commission, which essentially confirmed the need and appropriateness of doing so, saying that "*the Australian Government should accept and implement the Australian Law Reform Commission's final recommendations regarding a fair use exception in Australia*"[718].

However, this has not yet happened, and the landscape of exceptions to copyright in Australia, as outlined in the Australian Copyright Act[719], is therefore currently still composed of a number of provisions, the fair dealing provisions, which render certain specific uses of copyright works, that must nevertheless pursue particular purposes, as non-infringing.

These purposes are five:

1. research or study (Sec. 40(1A))
2. review or criticism (Sec. 41)
3. parody or satire (Sec. 41 A)
4. news reporting (Sec. 42)

---

[713] In particular, six of the consultations eventually recommended the adoption of the "fair use" model of copyright exceptions: two inquiries specifically into the Copyright Act (1998, 2014); and four broader reviews (both 2004, 2013, 2016). Only one, the one in 2000, recommended not to introduce the fair use clause, and another, the one in 2005, did not reach a final report.

[714]Copyright and the Digital Economy (ALRC 122) - Terms of Reference. 29.06.2012. Available at: https://www.alrc.gov.au/inquiry/copyright-and-the-digital-economy/terms-of-reference-13/

[715] Organization for Economic Co-operation and Development (OECD) Report: Chapter 7 - *Legal Aspects of Open Access To Publicly Funded Research*, cit., 394.

[716] See: Copyright and the Digital Economy (ALRC 122) - Incidental or Technical Use and Data and Text Mininghttps://www.alrc.gov.au/publication/copyright-and-the-digital-economy-alrc-report-122/11-incidental-or-technical-use-and-data-and-text-mining/

[717] Copyright and the Digital Economy (ALRC 122) - Recommendations. 2. 12. 2013. Available at: https://www.alrc.gov.au/publication/copyright-and-the-digital-economy-alrc-report-122/recommendations-6/

[718] See: Australian Government, Productivity Commission, Intellectual Property Arrangements – Inquiry report, 2016. Available at: https://www.pc.gov.au/inquiries/completed/intellectual-property/report

[719] Australian Copyright Act 1968. Available at: http://www5.austlii.edu.au/au/legis/cth/consol_act/ca1968133/

5. judicial proceedings or professional legal advice (Sec. 43)

In addition to these fair dealing cases, the Australian Copyright Act also contains other specific exceptions to copyright ("*Acts not constituting infringement of copyright*") in Divisions 3 to 5, which include and cover various uses that are exempted from the application of copyright, ranging from copies for private use and copies made by educational establishments, to temporary reproductions made in the course of communications[720].

Despite the fact that there is not yet a specific exception for TDM or a fair use clause that can with certainty cover the reproduction activities carried out in the TDM process, the doctrine has tried, not dissimilarly to what we tried to do in the previous chapter and to what European academics did in the years preceding the insertion of the two exceptions in the copyright acquis communautaire, to subsume TDM activities within the existing fair dealing provisions or other specific exceptions[721].

In particular, the exception for purposes of research or study contained in Sec. 40(1A), which is homologous to the European exceptions for purposes of research, has been found to be one of the most usable exceptions to render lawful the copies that occur in the TDM process.

It reads as follows:

> *A fair dealing with a literary work (other than lecture notes) does not constitute an infringement of the copyright in the work if it is for the purpose of, or associated with, an approved course of study or research by an enrolled external student of an educational institution.*

In order to determine whether the reproduction of all or part of a work constitutes fair dealing for research or study purposes, subsection 2 then provides a set of criteria, or guidelines, similar to those contained in the US fair use clause, namely:

1. *the purpose and character of the dealing;*
2. *the nature of the work or adaptation;*
3. *the possibility of obtaining the work or adaptation within a reasonable time at an ordinary commercial price;*
4. *the effect of the dealing upon the potential market for, or value of, the work or adaptation; and*
5. *in a case where part only of the work or adaptation is reproduced, the amount and substantiality of the part copied taken in relation to the whole work or adaptation.*

Studying the applicability of this exception for the typical scenario of machine learning, an activity that can be assimilated to TDM, it has been argued[722] that in particular factors (1), (3) and (4) are the factors that would mostly push towards the qualification of this activity as fair dealing, while on the other hand factors (2) and (5) would tend to the contrary.

As to factor 1, according to the transformativeness test which was developed by US courts, the nature of the use of TDM, which seems to be of a transformative rather than

[720] Organization for Economic Co-operation and Development (OECD) Report: Chapter 7 - *Legal Aspects of Open Access to Publicly Funded Research*, cit.,

[721] The analysis presented below about the current legitimacy of text and data mining activities in Australian copyright law is taken from: R. MATULIONYTE, *Australian Copyright Law Impedes the Development of Artificial Intelligence: What Are the Options?*, in International Review of Intellectual Property and Competition Law (IIC) 52, 2021, 417–443.

[722] R. MATULIONYTE, *Australian Copyright Law Impedes the Development of Artificial Intelligence: What Are the Options?*, cit.

consumptive nature, Australian commentators have suggested this should also be acknowledged in Australia. They would also suggest that TDM carried out for research and non-commercial purposes should be qualified as fair.

Factor 3 (the possibility of obtaining the work within a reasonable time at an ordinary commercial price) would also support the qualification of TDM as fair dealing, given the likely perceived impossibility of seeking rightholders for each work to be included in the dataset and mined and of obtaining licences for each of them.

The fourth factor would also be inclined to the qualification of fair use because it is hardly possible to think and argue that the activities of TDM can have any negative impact on the market of the works used.

More problems would be created by factor 2, as fair dealing is more likely to be recognised for activities carried out on material containing factual information rather than on highly creative content, which would perhaps exclude most TDM projects, which make use of academic and scientific publications.

However, the factor that seems to create more problems to the qualification of such activity as fair dealing for the purposes of research or study is the fifth, first of all because of the need by TDM, already stressed several times in this work, to use large amounts of whole works. The fifth factor, in fact, although it does not explicitly state that the unauthorized copying of an entire copyrighted work would not *a priopri* constitute fair dealing, nevertheless, provides that it can be considered as acceptable for the purposes of fair dealing the copying, for example, of a single article from a periodical publication, provided that other articles from the same publication are not also copied[723]; or when only a "*reasonable portion*" is copied, which is then defined and identified by the same article as 10% of the entire publication or one chapter[724]. This suggests that generally only copying of a small amount of material is acceptable for qualification of use as fair dealing for research and study purposes.

Potentially, it has been said, also the other fair dealing provisions could be used, but probably limited to situations where the output of the project constitutes criticism, review, parody or satire; probably excluding however some cases for which there is a requirement of acknowledgement of authors whose works have been used, such as for the provision on fair dealing for news reporting, which requires that "*a sufficient acknowledgement of the work is made*", as well as, similarly, also for the fair dealing provision for purpose of criticism and review.

More interesting, on the other hand, might be the specific exceptions for temporary copies contained in Sec. 43A and Sec. 43B[725].

Section 43A of the Copyright Act 1968 covers temporary reproductions made "*as part of the technical process of making or receiving a communication*" and was introduced to cover copies that occur during online communication of protected material (such as "*cache copies*").

Section 43B covers temporary reproductions "*if the reproduction is incidentally made as a necessary part of a technical process of using a copy of the work*" and was intended to cover inter alia RAM (Random Access Memory) copies that are created on users' computers or other end-user devices when users access a particular piece of online content.

These exceptions might cover some copies made during a TDM process but would not cover the whole process. In fact, as explained above, after initially creating a dataset, this dataset is executed through software that requires multiple copies and re-copies of the content. Here too, indeed, the problem arises that we have already addressed with regard to the same exception provided by EU copyright law. That is, these kind of exceptions are fine

[723] Sec. 40(3) and (4) Copyright Act 1968.
[724] Sec. 40(5) Copyright Act 1968.
[725] R. MATULIONYTE, *Australian Copyright Law Impedes the Development of Artificial Intelligence: What Are the Options?*, cit.

for the analysis phase, but they are not so fine with regards to those phases preceding the analysis, i.e. those in which the dataset to be mined is created, which requires the making of lasting copies of works or parts of them, which almost certainly would not satisfy the requirement of the temporariness of the copies, being furthermore difficult to argue that the reproduction of a work in a dataset is "*incidental*", given the fact that the content to be included in the dataset is often carefully selected, cleaned, and often classified and labelled before being mined. Not to mention the uncertainty created by some of the conditions sets by the provisions: for example, Section 43(A)(1), by express statutory provision, applies to those temporary and technical copies that are made "*as part of the technical process of making or receiving a communication*", and it is really unclear whether a TDM process would necessarily include communication of copies.

4.4.2. Text and Data Mining in Canada

A particular case is Canada, where a fair use clause modelled on the U.S. one has not been introduced, mainly for reasons of cultural discrepancies[726]; instead, the application of the fair dealing provisions characteristic of the Commonwealth countries (notably the UK and Australia) has remained. Here, however, unlike Australia, they have been subject for some years to very broad interpretations by the Courts, to the point of being assimilated to a fair use clause. In addition, a consultation about the introduction of an exception to copyright for text and data mining activities was recently launched.

The Canadian Copyright Act[727] regulates copyright at the federal level since its original enactment in 1921, and with respect to copyright exceptions, the Canadian Copyright Act contained a fair dealing exception from its first enactment, in section 29. The text of the fair dealing exception current until 2012, provided that fair dealing applied to the following purposes: research, private study, criticism, review, and news reporting. The Act also provides, as far as we are concerned, for a specific exception on temporary reproductions for technological processes at Article 30.71[728] and other exceptions that might apply to TDM but would likely apply in more limited situations and to a smaller subset of users.

We will focus our attention here exclusively on Canada's approach to fair dealing provisions, as with regard to the exception for temporary reproductions the considerations made when dealing with EU law and Australia also hold true here[729].

Canada's approach to the particular copyright exceptions that are the fair dealing exceptions has varied considerably over the last few decades. Until recently, Canada's fair dealing provisions were considered to be quite restrictive, both in terms of the purposes for which fair dealing could be invoked, and in terms of how the courts interpreted them.

---

[726] M. NAIR, *Canada and Israel: Cultivating Fairness of Use,* PIJIP Research Paper no. 2012-04 American University Washington College of Law, Washington, D.C., 2012.

[727] Copyright Act (R.S.C., 1985, c. C-42). Available at: https://laws-lois.justice.gc.ca/eng/acts/C-42/Index.html

[728] Copyright Act (R.S.C., 1985, c. C-42), Temporary reproductions, 30.71:
It is not an infringement of copyright to make a reproduction of a work or other subject-matter if
(a) the reproduction forms an essential part of a technological process;
(b) the reproduction's only purpose is to facilitate a use that is not an infringement of copyright; and
(c) the reproduction exists only for the duration of the technological process.

[729] Copyright Board of Canada interpreted this provision as "intended to capture copies that happen automatically, or without the direct control of the user", and that are automatically deleted once the technological process is completed. While some TDM activity may require making ephemeral copies, other TDM may require copies of works to be stored indefinitely. See: SOCAN, Re: Sound, CSI, Connect/SOPROQ, Artisti – Tariff for Commercial Radio, 2011-2017 (2016), at para 175 to 186, online: Copyright Board https://decisions.cb-cda.gc.ca/cb-cda/decisions/en/366778/1/document.do.

Examples of this tendency are two very important pronouncements of Canada's highest court, the first in 1990, *Bishop v Stevens*[730], in which the court held that because the Copyright Act was based on UK law, it was adopted with only one object: "*namely, the benefit of authors of all kinds, whether the works were literary, dramatic or musical*"; the second, in 1997, *Michelin v CAW Canada*[731], in which the plaintiff argued that the use of a logo was a parody and thus qualified as criticism under the fair dealing exception, an argument that was rejected by the Federal Court, which instead emphasised the need to interpret the fair use clauses restrictively, while maintaining that parody was not an enumerated exception within the Copyright Act and that further, it was not synonymous with criticism[732].

This particularly restrictive approach by Canadian courts lasted until 2002, when a change within the courts began. This change is evident in the 2002 case *Théberge v Galerie d'Art du Petit Champlain inc.* [733], in which the court gave explicit support for a copyright balance and particular consideration for copyright's effect on innovation[734]. It pointed out here the dangers that a copyright that is too unbalanced towards rightsholders can create both to the public and to the innovation process[735][736].

The real turning point in the process of changing the attitude of the Courts towards fair dealing provisions, however, came in the case of *CCH Canadian Ltd. v Law Society of Upper Canada*[737], a landmark case dating March 2004, in which the court again reaffirmed its support for a balanced copyright, giving a definitive change and new life to the fair dealing provisions. The Court, which was asked to decide upon the application of the fair dealing defence for purposes of research and private study[738], provided a detailed discussion of the fair dealing exception[739].

At the beginning of the judgement and shortly before focusing on the analysis of the case, it first of all stated, as a demonstration of the substantial change of approach towards the fair dealing provision, that the exceptions to copyright infringement are real users' rights that must be balanced against the rights of copyright owners and creators[740].

---

730 *Bishop v. Stevens*, 1990 CanLII 75 (SCC), [1990] 2 SCR 467. Available at: http://canlii.ca/t/1fsv7

731 *Michelin v. Can,* 1996 CanLII 11755 (FC), [1997] 2 FC 306. Available at: https://canlii.ca/t/4g4v

732 GEIST, *Fairness Found: How Canada Quietly Shifted from Fair Dealing to Fair Use*, cit.

733 *Théberge v. Galerie d'Art du Petit Champlain inc.*, 2002 SCC 34 (CanLII), [2002] 2 SCR 336. Available at: https://canlii.ca/t/51tn.

734 Ibid, § 31: "[T]he proper balance among these and other public policy objectives lies not only in recognizing the creator's rights but in giving due weight to their limited nature…. Once an authorized copy of a work is sold to a member of the public, it is generally for the purchaser, not the author, to determine what happens to it".

735 Ibid., § 32: "[e]xcessive control by holders of copyrights and other forms of intellectual property may unduly limit the ability of the public domain to incorporate and embellish creative innovation in the long-term interests of society as a whole, or create practical obstacles to proper utilization".

736 GEIST, *Fairness Found: How Canada Quietly Shifted from Fair Dealing to Fair Use*, cit.

737 *CCH Canadian Ltd. v. Law Society of Upper Canada*, [2004] 1 S.C.R. 339, 2004 SCC 13. Available at: https://decisions.scc-csc.ca/scc-csc/scc-csc/en/item/2125/index.do

738 The case involved a dispute between the Law Society of Upper Canada and several legal publishers. The Law Society, which maintains the Great Library, a leading law library in Toronto, provided the profession with two methods of copying cases and other legal materials. First, it ran a service whereby lawyers could request a copy of a particular case or article. Second, it maintained several standalone photocopiers that could be used by library patrons. The legal publishers objected to the Law Society's copying practices and sued for copyright infringement. They maintained that the materials being copied were entitled to copyright protection and that the Law Society was authorizing others to infringe on their copyright.

739 *Ibid.* at §§ 48-60.

740 *Ibid.* at §§ 12 and 48. § 48: "Before reviewing the scope of the fair dealing exception under the Copyright Act, it is important to clarify some general considerations about exceptions to copyright infringement. Procedurally, a defendant is required to prove that his or her dealing with a work has been fair; however, the fair dealing exception is perhaps more properly under- stood as an integral part of the Copyright Act than

The Court then proceeded to analyse the case in the light of the so-called "*two step test*" to determine whether or not a given activity can be considered as fair dealing under section 29 of the Copyright Act, and held that, in order to do so, a defendant must first prove[741] that the activity falls within one of the permissible purposes listed in the Act. However, surprisingly, the Court held here that the purpose "*research*" "*must be given a broad and liberal interpretation in order to ensure that the rights of users are not unduly restricted*". It also held at the same point that the term "*research*" is not limited to non-commercial or private contexts[742].

Secondly, the court proceeded to discuss the second step, according to which the dealing must be fair so that, in addition to the purpose of the use, a series of criteria set out by the courts have to be assessed on a case-by-case basis. This ruling is of particular importance, as it finally defined which criteria are to be used in assessing the fairness of the use[743]. They include:

- The purpose of the dealing, arguing for this point the court that "*allowable purposes should not be given a restrictive interpretation or this could result in the undue restriction of users' rights*"[744].
- The character of the dealing, according to which, attention should be paid to assessing whether a single copy or several copies have been made[745].
- The amount of the dealing, according to which "*both the amount of the dealing and importance of the work allegedly infringed should be considered in assessing fairness*"[746].
- The existence of any alternatives to the dealing, for which the courts said that "*alternatives to dealing with the infringed work may affect the determination of fairness*"[747].
- The nature of the work, stating on this point the court that *"[I]f a work has not been published, the dealing may be more fair in that its reproduction with acknowledgement could lead to a wider public dissemination of the work—one of the goals of copyright law. If, however, the work in question was confidential, this may tip the scales towards finding that the dealing was unfair*"[748].
- The effect of the dealing on the market of the work, the court finally helding that "*although the effect of the dealing on the market of the copyright owner is an important factor, it is neither the only factor nor the most important factor that a court must consider in deciding if the dealing is fair*"[749].

The changes brought to the Canadian copyright system by the *CCH* decision, and the newly found innovative spirit, triggered a great debate about the purpose of fair dealing provisions. From this moment on, Canadian copyright academics began to question and

simply a defence. Any act falling within the fair dealing exception will not be an infringement of copyright. The fair dealing exception, like other exceptions in the Copyright Act, is a user's right. In order to maintain the proper balance between the rights of a copyright owner and users' interests, it must not be interpreted restrictively".

741 See: *Alberta Education v. Access Copyright*, 2010 F.C.A. 198 §§ 18-19 (2010). Available at http://www.canlii.org/en/ca/fca/doc/2010/2010fca198/2010fca198.html.

742 CCH, § 51.

743 It is noteworthy that the Linden six-factor test was itself influenced by Hubbard v Vosper, a United Kingdom decision authored by Lord Denning.

744 CCH, § 54.

745 *Ibid.*, § 55.

746 *Ibid.*, § 56.

747 *Ibid.*, § 57.

748 *Ibid.*, §58.

749 *Ibid.*, §59.

consider the possibility and benefits of expanding fair dealing provisions[750] which, at the time, were limited to five purposes: research, private study, criticism, reporting and review[751]. Indeed, some scholars called for the adoption of a pro-innovative fair-use-like provision based on longstanding fair dealing jurisprudence[752], for example by making the current list of fair dealing categories "*illustrative*" rather than exhaustive, simply by adding the words "*such as*" to the current provision[753].

However, the introduction of Bill C-32 (later Bill C-11)[754] in mid-2010 disappointed those who hoped that statutory reform would introduce a flexible fair dealing provision, as the government chose instead to introduce new purposes, but refused to make the provision more open to allow courts to identify new purposes in appropriate circumstances. In fact, with the copyright reform bill, that received royal assent in June 2012, the Canadian government added three additional purposes to the law: parody, satire and education[755].

The precedent set by the Supreme Court of Canada in the *CCH* case, however, did not cease to have effects any time soon, as Canadian courts a few years later affirmed that not only research purposes, but all fair dealing purposes, should be given the same broad and liberal interpretation[756].

These two events, namely the Bill C-11 reform, and the changed interpretation of the Canadian courts, have led some commentators to see this as "*a subtle move towards an American-style fair use defence*"[757]. In fact, it has been noted that, although Canadian copyright has formally maintained the two-stage analysis of fair dealing, the first stage has been modified over time, through the introduction of additional purposes and the broad interpretation given to them by the courts, to become so easy to meet that future Canadian analyses of fair dealing will reasonably include only a perfunctory assessment of the first-stage purposes test, and,

---

750 See e.g. C. CRAIG, *The Changing Face of Fair Dealing in Canadian Copyright Law: A Proposal for Legislative Reform* in M. GEIST, ed, *In the Public Interest: The Future of Canadian Copyright Law* (Toronto: Irwin Law, 2005) 437; C. CRAIG, *Locking Out Lawful Users: Fair Dealing and Anti-Circumvention in Bill C-32*. Available at: http://www.irwinlaw.com/content/assets/content-commons/666/CCDA%2007%20Craig.pdf ; and M. A. WILKINSON, *Copyright, Collectives, and Contracts: New Math for Educational Institutions and Libraries*, http://www.irwinlaw.com/pages/content-commons/copyright-collectives-and-contracts--newmath-for-educational-institutions-and-libraries---margaret-ann-wilkinson in M. GEIST, ed, *From "Radical Extremism" to "Balanced Copyright": Canadian Copyright and the Digital Agenda*, Irwin Law, Toronto, 2010, 177, 503.

751 GEIST, *Fairness Found: How Canada Quietly Shifted from Fair Dealing to Fair Use*, cit.

752 See also e.g. M. GEIST, *Copyright Consultation Submission*, 2 *Osgoode Hall Rev L Pol'y* (2009), 59; S. E. TROSOW, *Copyright Consultation Submission* 2 Osgoode Hall Rev L Pol'y (2009), 169. For an opposing perspective, see B. SOOKMAN, D. GLOVER, *Why Canada Should Not Adopt Fair Use: A Joint Submission to the Copyright Consultations,* 2 Osgoode Hall Rev L Pol'y (2009), 55.

753 M. GEIST, *How Copyright Reform Could Support Canada's Supercluster Investment*, Centre for International Governance Innovation (CIGI), 19 March 2019. Available at: https://www.cigionline.org/articles/how-copyright-reform-could-support-canadas-supercluster-investment/

754 Bill C-32, *An Act to amend the Copyright Act*, 3d Sess, 40th Parl, 2010 (First Reading 2 June 2010). Available at: http://www.parl.gc.ca/HousePublications/Publication.aspx?Pub=Bill&Doc=C-32_1&Language=&Mode=1&Parl=40&Ses=3 .

755 GEIST, *Fairness Found: How Canada Quietly Shifted from Fair Dealing to Fair Use*, cit.

756 *Warman v. Fournier*, 2012 FC 803 (CanLII), § 31: "The SCC stated in CCH, at paragraph 51, that the fair dealing purposes (in that case, research) "must be given a large and liberal interpretation in order to ensure that users' rights are not unduly constrained." Applying this large and liberal interpretation to news reporting, I find that the respondents' dealing in respect of the Kay Work falls within this purpose. They posted the excerpts of the Kay Work on Free Dominion to promulgate the facts recounted in that article. Thus, the first criterion for fair dealing is met. The news reporting exception also requires that the source and author be mentioned, which is also satisfied in this case".

757 GEIST, *Fairness Found: How Canada Quietly Shifted from Fair Dealing to Fair Use*, cit.

instead, a much more rigorous analysis in the second-stage, the six-factor assessment, thus increasingly resembling the analysis carried out in American fair use[758].

Nonetheless, it has been stressed[759], the Canadian system still corresponds to an essentially "*fair dealing*" model. In fact, the current list of eight fair dealing purposes identified in the Act literally is meant to be not "*illustrative*" but "*exhaustive*", preventing therefore courts from adding interpretatively further purposes. On the contrary, in the US model, the exception is basically open to any purpose and the fair balance whereby a use can be deemed as fair depends just upon the use of the work and not upon the purpose of the copying. Some commentators argued that more could have been done than simply reaffirming that fair dealing was a user's right, especially when, when coupled with Bill C-11's statutory reforms, there was the possibility of having effectively turned the Canadian fair dealing clause into a fair use provision[760].

Coming finally to the specific subject of our research, we must conclude that, according to the current Canadian law provisions, TDM can be qualified as fair dealing depending on the specific characteristics of each individual case, which have to be assessed on the basis of the above-mentioned conditions, so that TDM would need to fall under one of the enumerated purposes and be "*fair*". The Supreme Court, over the years, more than once has underlined the importance of allowing fair dealing for purposes of research and private study, claiming that such activities could, under the appropriate factual circumstances, be qualified as fair dealing under the new Canadian copyright regime, since machine learning does not harm the primary purposes of the original work.

Something, however, seems to have changed in the last period, and in the right direction, as the Canadian government seems to have recognised the value of TDM in research, problem solving, and AI development and understood the TDM requirement for a large-scale reproduction of copyrighted works[761].

In fact, a parliamentary review of the Copyright Act (the Act) was held during 2018-19. During this consultation, stakeholders raised a number of issues relating to AI and IoT. In particular, with regard to TDM, they provided a number of recommendations for legislative amendments to provide greater legal certainty for those wishing to avail themselves of the use of TDM techniques. Stakeholder recommendations included: expanding the permissible purposes of the fair dealing exception to include TDM[762]; amending the fair dealing exception to make it open-ended, such as the fair use provision in the United States[763]; amending the exception for temporary reproductions for technological processes in section

---

[758] *Ibidem.*

[759] GEIST, *How Copyright Reform Could Support Canada's Supercluster Investment*, cit.

[760] GEIST, *Fairness Found: How Canada Quietly Shifted from Fair Dealing to Fair Use*, cit.

[761] Canada's federal government has released the report "*Building a Nation of Innovators*" in which are summarized the efforts undertook over the past three years, emphasizing the increased investments made in strategic sectors of innovation policy. Available at: https://www.ic.gc.ca/eic/site/062.nsf/eng/h_00105.html

[762] See for example, Untitled, Creative Commons (brief) (15 May 2018), online: House of Commons, INDU https://www.ourcommons.ca/Content/Committee/421/INDU/Brief/BR9887146/br-external/CreativeCommons-e.pdf .

[763] See for example, Brief – Statutory Review of the Copyright Act submitted by Pascale Chapdelaine, on behalf of Canadian intellectual property law scholars, Pascale Chapdelaine & Myra Tawfik (brief) (22 October 2018), online: House of Commons, INDU https://www.ourcommons.ca/Content/Committee/421/INDU/Brief/BR10166923/br-external/ChapdelainePascale01-e.pdf .

30.71 of the Act to cover TDM[764]; and creating a new exception dedicated specifically to TDM[765].

In its final report[766], the Standing Committee on Industry, Science and Technology (INDU) made two recommendations that apply to TDM. First, INDU recommended that the government "*amend the Copyright Act to facilitate the use of a work or other material for the purpose of information analysis*"[767]. Second, INDU recommended "*amending section 29 of the Copyright Act to make the list of purposes allowable under the fair dealing exception an illustrative list rather than an exhaustive one*"[768]. This second recommendation would have implications not only for TDM's business.

On 16 July 2021, the federal government (through the Minister for Innovation, Science and Industry and the Minister for Canadian Heritage) announced[769] its public consultation[770] on adapting the Canadian copyright framework to ongoing developments in artificial intelligence (AI) and the Internet of Things (IoT). A consultation document has also been released that identifies some of the challenges to the copyright framework posed by these topics, some options and approaches considered to address these challenges, and questions to guide stakeholder input. Participants will have until 17 September 2021 to share their input, and the responses received will be made publicly available after the consultation period, helping to inform the government's policy development process to ensure that Canada's copyright framework for AI and IoT reflects the evolving digital world.

In its consultation paper, the Canadian government justified the consultation as follows:

"*Copyright law is now faced with another major technological evolution: artificial intelligence (AI), a general-purpose technology with widespread and increasing implications throughout the economy and society. AI is raising questions as to whether the copyright framework can adequately address new situations that are arising and increasingly will arise, and whether changes to the copyright framework are needed. Examples of new situations that are arising include the use of copyrighted works as part of text and data mining (TDM) to train and develop AI applications, and the use of AI to create, produce and distribute literary, musical and other kinds of works*".

---

764 See for example, Untitled, Mark Hayes (brief) (2 November 2018), online: House of Commons, INDU https://www.ourcommons.ca/Content/Committee/421/INDU/Brief/BR10166884/br-external/HayesMark-e.pdf .

765 See for example, The Software Alliance to the Standing Committee on Industry, Science and Technology Regarding the 2018 Statutory Review of the Copyright Act, BSA: the Software Alliance (brief) (26 September 2018), online: House of Commons, INDU https://www.ourcommons.ca/Content/Committee/421/INDU/Brief/BR10057455/br-external/BSATheSoftwareAlliance-e.pdf .

766 Canada, Parliament, House of Commons, Statutory Review of the Copyright Act: Report of the Standing Committee on Industry, Science and Technology, 42nd Parliament, 1st Session, No 16 (June 2019) [INDU Report], online: House of Commons https://www.ourcommons.ca/DocumentViewer/en/42-1/INDU/report-16 .

767 *Ibid*, Recommendation 23.

768 *Ibid*, Recommendation 18.

769 Government of Canada, News Release: The Government of Canada Launches Consultation on a Modern Copyright Framework for AI and the Internet of Things, 16 July 2021. Available at: https://www.canada.ca/en/innovation-science-economic-development/news/2021/07/the-government-of-canada-launches-consultation-on-a-modern-copyright-framework-for-ai-and-the-internet-of-things.html

770 Government of Canada, A Consultation on a Modern Copyright Framework for Artificial Intelligence and the Internet of Things, 16 July 2021. Available at: https://www.ic.gc.ca/eic/site/693.nsf/eng/00316.html#s21

As the Canadian government has therefore shown a strong willingness to address the issue of TDM and to further investigate its relationship with copyright[771], it is likely that there will be new legislative developments in the coming years.

4.5. Text and Data Mining in China

Our work could not be complete and exhaustive if we don't try, as far as possible, given the limited sources (rather rare and not easily accessible[772]), to trace a brief overview of the approach that one of the major world powers is nowadays trying to adopt. Therefore, we shall now, at the conclusion of our work, deal with China, since it is certainly one of the countries that is most affirming itself in the field of artificial intelligence and which has, for some time now, showed great ambitions in this respect.

The Chinese Copyright Law[773] (hereinafter CLC) was enacted in 1990 and has been revised three times over the past three decades: first in 2001, then in 2010, and finally a third time, recently, in 2020. In fact, on 11 November 2020, Chinese President Xi Jinping, in a Presidential Decree No.55, announced The Decision of the Standing Committee of the National People's Congress on Amending the Copyright Law of the People's Republic of China, which was approved at the 23rd session of Standing Committee of the 13th National People's Congress. The CLC amendment recently came into force on 1 June 2021.

Like the neighbouring Asian countries already analysed, Taiwan, South Korea and Japan, China also chose to adopt for its legislation the model of civil law countries, and also in regulating copyright it has therefore followed the model of the European droit d'auteur countries, choosing, consistently with this choice, to implement a closed list of exceptions instead of an open-ended clause as in the United States.

China's copyright law includes an article, Article 24, containing a list of exceptions, which was originally introduced in 1990 as Article 22 and which remained unchanged following the amendments to the CLC in 2001, 2010, and, as argued by the doctrine, substantially in the 2020 revision. This article now lists 13 circumstances in which a copyrighted work may be used without the permission of the rightholders and without the need to pay any remuneration.

Article 24 of the CLC now provides:

> *In the following cases, a work may be exploited without the permission from, and without payment of remuneration to, the copyright owner, provided that the name or appellation of the author and the title of the work are mentioned, and the normal use of the work is not prejudiced and the legitimate rights and interests of the copyright owner are not unreasonably prejudiced:*
> *(1) use of another person's published work for the purposes of the user's own private study, research or appreciation;*
> *(2) appropriate quotation from another person's published work in one's own work for the purposes of introducing or commenting on a certain work, or explaining a certain point;*

---

[771] *Ibid*, "The central policy issue to be explored in this part of the consultation is whether amendments should be introduced in the Act to clarify how the copyright framework applies to TDM activity, and if so, what should those amendments be. There is currently some uncertainty regarding the extent to which existing exceptions in the copyright framework apply to TDM activity".

[772] The analysis below of Chinese legislation and jurisprudence is in fact based almost entirely on the work: T. HE, *Copyright Exceptions Reform and AI Data Analysis in China – A Modest Proposal*, in J. A. LEE, R. HILTY, K. C. LIU, *Artificial intelligence and Intellectual Property,* cit.

[773] Copyright Law of the People's Republic of China (as amended up to the Decision of November 11, 2020, of the Standing Committee of the National People's Congress on Amending the Copyright Law of the People's Republic of China). Available at (Chinese only): https://wipolex.wipo.int/en/legislation/details/21065

(3) *unavoidable inclusion or quotation of a published work in the media, such as newspaper, periodical and radio and television program, for the purpose of reporting current events;*
(4) *publishing or rebroadcasting by the media, such as a newspaper, periodical, radio station and television station, of an article published by another newspaper or periodical, or broadcast by another radio station or television station, etc. on current political, economic or religious topics, except where the declares that such publishing or rebroadcasting is not permitted;*
(5) *publishing or broadcasting by the media, such as a newspaper, periodical, radio station and television station of a speech delivered at a public gathering, except where the author declares that such publishing or broadcasting is not permitted;*
(6) *translation, adaptation, compilation, playing, or reproduction in a small quantity of copies of a published work by teachers or scientific researchers for use in classroom teaching or scientific research, provided that the translation or the reproductions are not published for distribution;*
(7) *use of a published work by a State organ to a justifiable extent for the purpose of fulfilling its official duties;*
(8) *reproduction of a work in its collections by a library, archive, memorial hall, museum, art gallery, cultural halls etc. for the purpose of display, or preservation of a copy, of the work;*
(9) *gratuitous live performance of a published work, for which no fees are charged to the public, nor payments are made to the performers and there is no aim to profit;*
(10) *copying, drawing, photographing or video-recording of a work of art put up or displayed in an outdoor public place;*
(11) *translation of a published work of a Chinese citizen, legal entity, or other organization from the standard written Chinese language into minority nationality languages for publication and distribution in the country; and*
(12) *providing published works in an accessible fashion that can be perceived by people with print disabilities;*
(13) *Other circumstances provided for by laws and administrative regulations.*

The provisions of the preceding paragraph apply to restrictions on copyright-related rights.[774]

Even before the amendment of Article 22 (now renumbered as Article 24), which introduced the provision in the same wording, Article 21 of the Regulations for the Implementation of the Copyright Law of the People's Republic of China (hereinafter RICL) provided for the so-called "*two-step test*", stating as follows:

> *The exploitation of a published work which may be exploited without permission from the copyright owner in accordance with the relevant provisions of the Copyright Law shall not impair the normal exploitation of the work concerned, nor unreasonably prejudice the legitimate interests of the copyright owner.*

This provision was seen as a complement to the list of specific exceptions by the doctrine, meant as a limitation similar to the three-step test of the Berne Convention. Thus, according to this rule uses of protected material enlisted in Article 24 as exceptions to the

774 Unofficial translation provided by: HE, *Copyright Exceptions Reform and AI Data Analysis in China – A Modest Proposal*, cit., 210.
For alternative translations see: https://www.lexology.com/library/detail.aspx?g=e71cbb95-9c32-4129-9da8-92b990e09b24;

rights conferred by copyright must not violate the two requirements of the test: "*not impair the normal exploitation*" and not "*unreasonably prejudice the legitimate interests of the copyright owner*".

This is with respect to the Chinese Copyright exception system in general. We have already understood from what has been said above that China, contrary to what we could believe, has not yet included a specific exception for TDM. Let us now see whether, within the framework of the exceptions outlined above, TDM activities are currently allowed or not, whether there are particular trends pointing towards the affirmative solution and, at the end, some proposals put forward by the doctrine for the amendment of the CLC in order to accommodate text and data mining activities in the Chinese legal system.

The doctrine notes[775] that the only two provisions that can be considered applicable to TDM are contained in Article 24(1) and (6). In particular, paragraph 1 provides for the "u*se of a published work for the purposes of the user's own private study, research or self-entertainment*". This exception, therefore, could easily include "*private research*". However, according to the doctrine, this would not be enough, because TDM activities carried out by those who have the greatest interest in using TDM techniques would still be excluded, i.e. companies, which carry out research that could not be qualified as private. Moreover, they note that in any case, even if a private party wished to take advantage of the exception to carry out TDM activities, it would still be limited by the already mentioned "*two-step test*" provided for in Article 24 CLC and Article 21 of the RICL, which would almost certainly affect the amount of material usable by the private user, which, requiring a lot of material to perform the TDM, would certainly go beyond the quantity considered tolerable.

Another possible foothold for considering TDM activities legitimate in China is Article 24(6), which provides an exception for:

> *translation, adaptation, compilation, broadcasting or reproduction in a small quantity of copies, of a published work for use by teachers or scientific researchers in classroom teaching or scientific research of schools, provided that the translation adaptation, compilation, broadcasting or reproduction is not published or distributed.*

Even with regard to this paragraph, perplexities have emerged. Although "*scientific research*" is included in the exception, which seems to be widely understood, the doctrine considers that the exception is not sufficiently broad to be used as an exception for the TDM. In fact, it sets limits that seem to be rather incompatible with the requirements we have said are proper to TDM, including the amount of material to be used, which is limited here by the term "*reproduction in a small quantity of copies*". Moreover, it is not yet clear whether or not this exception covers scientific research activities carried out for commercial purposes.

In conclusion, beyond these two exceptions, which may only partially cover TDM activities, in particular exclusively those carried out by private parties and, to a limited extent, scientific research, it can be safely argued that there is currently no strong exception that can guarantee a safe zone for those uses related to text and data mining and the activities deriving therefrom, such as the development of machine learning and artificial intelligence.

Let us now look at the trends in case law and the proposals made by the doctrine to try to fill this gap.

With regard to the Chinese jurisprudence, as often pointed out by the doctrine[776], it has always been rather inclined to support and encourage the use of new technologies[777],

---

[775] See: HE, *Copyright Exceptions Reform and AI Data Analysis in China – A Modest Proposal,* cit., 205.
[776] T. HE, *Copyright Exceptions Reform and AI Data Analysis in China – A Modest Proposal,* cit., 211.
[777] See for example the case: *Wang Shen v Google Inc. et al.* (2012) Beijing 1st Intermediary People's Court, YZMCZ No. 1321; (2013) Beijing Higher People's Court, GMZZ No. 1221.

listening to the fair use requests of defendants and supporting them when they consider it necessary, often going outside the framework outlined in article 24 containing the list of possible exceptions, and supporting their decision using different justifications. An example of such different justifications has been the use of the two-step test of article 21 of RICL as a general fair use clause (while instead its rationale seems to be that of an additional control to the cases of exception already explicitly provided for), even by directly introducing the four factors of fair use. Other examples could be the Supreme People's Court use of its jurisprudential interpretation in order to create new exceptions (for example for caches and thumbnails) outside the CLC[778] or to introduce the four factors of fair use directly in certain cases of judgment[779]. However, this has been considered not only insufficient, but also unconstitutional by the doctrine, according to which if the Supreme People's Court were to adopt the expansive and creative interpretations given by the lower courts, it would be contrary to the law, which provides that even the interpretation of the SPC must still refer primarily to the law and conform to the objectives, principles and original meaning attributed to the law, according to the Legislation Law of China[780]. This is not to mention the result that this would lead to arbitrary application of the law and judgments with contradictory interpretations.

Faced with the lack of a specific exception that could be used as a basis for TDM activities and faced with the impossibility of following the approach relying on judicial interpretation, the Chinese doctrine has tried to propose alternative solutions to resolve the issue.

First of all, it expressed its disappointment[781] with the proposed legislative change on which the new version of the CLC that has just entered into force was based, namely the NCAC draft for review of 30 April 2020[782]. Indeed, it was described as significantly less ambitious than the previous draft of 2014[783].

Article 43 (the counterpart of the current Article 24) of the version released on 6 June 2014 in fact, instead of the opening words "*the normal use of the work is not prejudiced and the legitimate rights and interests of the copyright owner are not unreasonably prejudiced*" of the current Article 24, simply contained the phrase "*the other rights enjoyed by the copyright owner in accordance with this Law are not prejudiced*". In addition, Section 43(13) provided, instead of the exception contained in the current Section 24(13) "*other circumstances provided for by laws and administrative regulations", an "other circumstances*" exception, which was seen as a rather bold change as it made the clause appear as a small general basket clause at the end of the catalogue of specific provisions on copyright exceptions[784], so similar to those of the neighboring Asian countries

---

On this topic, see: W. CHENGUANG, *Law-making functions of the Chinese courts: Judicial activism in a country of rapid social changes*, 1 *Frontiers of Law in China*, 2006, 524-549.

778 Provisions of the Supreme People's Court on Several Issues concerning the Application of Law in Hearing Civil Dispute Cases Involving Infringement of the Right of Dissemination on Information Networks (2012) (PRC), Art. 5 (2).

779 Several Opinions of the Supreme People's Court on Some Issues in Fully Giving Rein to the Funcion of Intellectual Property Rights Adjudication in Promoting the Great Development and Flourishing of Socialist Culture and Stimulating the Indigenous and Coordinated Development of Economy (2011) (PRC), para 8.

780 Legislation law (2015) (PRC), Art. 104.

781 T. HE, *Copyright Exceptions Reform and AI Data Analysis in China – A Modest Proposal,* cit., 213.

782 Revision proposal for the Copyright Law of China (2020). Available at: http://www.iprdaily.cn/article_24614.html

783 Legislative Affairs Office of the State Council P.R. China, *Notice About Circular of the Legislative Affairs Office of the State Council on Promulgating the Copyright Law of the People's Republic of China (Draft Revision for Review) for Public Consultation* (2014). Available at: http://www.gov.cn/xinwen/2014-06/10/content_2697701.htm

784 UENO, *The Flexible Copyright Exception for 'Non-Enjoyment' Purposes – Recent Amendment in Japan and Its Implication*, cit., 145.

already analyzed. It also inserted at the end of article 43, after the list of 13 exceptions, the content of article 21 of RICL "*the use of a published work pursuant to the above provisions may not prejudice the normal use of the work and may not unreasonably prejudice the lawful rights and interests of the copyright owner*", thus making the closed list an open one and opening the list not only to TDM but also to other future innovative uses not provided for in the list of exceptions.

In practice, however, the version that became law is the 2020 version, did not change the wording of the former Article 22, now 24, very much. In fact, apart from the addition of the two-step test of Article 21 of the RICL at the beginning of Article 24 and the new exception contained in point 13, which, however, is not immediately operative but only opens up the possibility of the insertion of further exceptions in the future by the law or administrative regulations, it did not bring any new elements, thus leaving currently open also the question about the insertion of a new exception for text and data mining, which could instead easily be included, if the 2014 version had passed, in point 13 of Article 43 "*other circumstances*".

The Chinese doctrine has thus sought alternative solutions and proposals.

A first solution proposed was a general fair use clause on the American model. Although the possibility and appropriateness of introducing such a clause in an Asian civil law country has already been amply demonstrated by the introduction of general clauses in the legislation of both South Korea and Taiwan, the doctrine considered this solution unfeasible for several reasons. First of all, because of the different relationship between those countries and the US and the lack of sympathy between China and the US; moreover, it has been pointed out that the Chinese legislator, when considering the transposition of laws from foreign rights, is rather careful to check that the values behind the foreign state accord with Chinese tradition and policy. In the case of fair use, the US Supreme Court argued in *Eldred v. Ashcroft*[785] that the "*fair use defence affords considerable latitude for scholarship and comment [...] even for parody*", which suggests that values such as freedom of speech might not be viewed favourably by Chinese government officials.

The second proposed solution, putting aside the first one, should instead be based on the model of the new Japanese exception already analysed, which could certainly be more feasible.

[785] *Eldred v Ashcroft*, 537 US 186, 219-20 (2003).

CONCLUDING REMARKS

In this work we have extensively discussed the topic of text and data mining from its very beginning, underlining the fundamental importance that data and their exploitation have nowadays and how TDM techniques are the tools useful for this purpose.

Given the increasing attention on the topic, which is still rather a relatively niche one (despite its great importance), and the interest that governments and experts in the field are showing, there have been many new developments on the subject during the drafting of this paper. Just to name a few: the national implementation process of the EU directive, whose deadline was June 7, 2021; the UK's exit from the EU and the prompt attempt to address the issue by freeing itself from the constraints of the EU copyright exception framework; the new artificial intelligence programmes of various countries (EU, Canada, UK) and the several consultations that discussed the relationship between artificial intelligence and intellectual property rights; the amendments to the Chinese copyright.

The numerous applications that TDM can have both in the scientific and business fields made us realize that the opportunity to exploit data is extremely tempting and could be a chance to relaunch an economy that has been stagnating for some time and even now is struggling to recover. The need for economic revitalisation and renewal is increasingly perceived as urgent today, in a world gripped by a pandemic crisis that has severely tested the world's healthcare as well as production and trading systems. A new challenge to the ability of states and international organisations to deal with a never before experienced emergency, capable of affecting all countries in the world equally and simultaneously.

Today more than ever a high-profile technical and scientific approach is needed that, starting from the all the knowledge available worldwide and from all the data collected could formulate effective solutions in a tight timeframe, where the timeliness of intervention is, unfortunately, measured in number of lives.

Nevertheless, there must be maximum awareness of the tensions that TDM activities, although useful, can generate above all in the legal intellectual property field, in particular copyright and database protection.

We have deduced, through historical investigation, how this clash originated from the uncontrolled expansion of rights that took place in the last century in correspondence with the three different waves of technological innovation, in an attempt to maintain the control of rights holders over the protected material. This has resulted in an unbalanced regulation of interests between rights holders and users.

We have noticed the increasing attention that some European governments are paying to the exploitation of data as a useful tool for the creation of new knowledge and value, even if slightly later than in some extra UE countries, such as Japan and the US.

We then saw the different approaches followed by the European Union to the issue. However, we could not help but notice how the solution finally enacted, i.e. the introduction of the two exceptions to Articles 3 and 4 of Directive 790/2019 failed to achieve the expected results and above all failed to fully achieve the most hoped-for objectives, i.e. legal certainty and greater freedom of manoeuvre, identifying as main criticalities of the two exceptions the narrowness of the beneficiaries and the possibility for rights holders to inhibit unquestionably the application of the broader exception.

During the legislative process that led to the directive, rather broad and complex as it includes several other debated topics, the lobbying activities of some stakeholders have certainly prevailed over the choices that experts in the field suggested and recommended as more beneficial and balanced.

The main non-EU countries have already addressed this issue, directly or indirectly, and in most cases resolved it quite easily.

It must be clear to us that we are now in the midst of a real technological revolution based on data, which has and will have repercussions on a number of cutting-edge activities, such as the use of machine learning for the development of the artificial intelligence industry. It is therefore of paramount importance that the legal framework for the exploitation of data be adapted to this massive revolution, not sacrificing the undoubtedly greater benefits that data can bring in terms of scientific research and economic prosperity for the benefit of few.

Since the early interest shown by the European Union in the issue, in the "*Licences for Europe*" project, the reluctance of the legislator to really resolve the issue and to "*sacrifice*" (although there is really no sacrifice) the interests of rightholders for the public interest has been quite evident. This attitude could only culminate in the adoption of the two exceptions, which, instead of properly solving the issue, as advised for years by most stakeholders, have finally led to a compromise that will not really benefit anyone and will rather continue to leave that uncertainty which, for almost a decade now, was intended to be removed.

If Europe wishes to avoid falling behind the other world powers, which have already been equipping themselves for some time and are already reaping the benefits of this work (see, for example, the case of the United States, Japan or Israel), or are equipping themselves now (Canada and China), having witnessed the essentially unsuccessful experience of the European Union, it will have to re-open the issue as soon as possible and, probably, find an alternative solution to the current exceptions on text and data mining.

BIBLIOGRAPHY

BOOKS AND JOURNALS

AGGARWAL C.C., *Data Mining: The Textbook*, Springer International Publishing, 2015.

ANANIADOU S., KELL D.B., TSUJII J., *Text mining and its potential applications in systems biology,* in *Trends in biotechnology* 24(12), 2007.

ARMSTRONG T. K., *Two Comparative Perspectives on Copyright's Past and Future in the Digital Age*, 15 *J. Marshall Rev. Intell. Prop. L.* 698 (2015-2016).

ARNOLD Sir R., *An Overview of European Harmonization Measures in Intellectual Property Law*, in OHLY A., PILA J., *The Europeanization of Intellectual Property Law*, Oxford University Press, 2013.

ARNOLD R., BENTLY L., DERCLAYE E., DINWOODIE G., *IP Law Post-Brexit*, 101(2) Judicature 65 (2017). Available at: https://judicature.duke.edu/articles/ip-law-post-brexit/

ATKINSON B., FITZGERALD B., *A Short History of Copyright: The Genie of Information*, Springer International Publishing, Berlin, 2014.

BALDWIN P., *The Copyright Wars: Three Centuries of Trans—Atlantic Battle*, Princeton University Press, 2014.

BEEBE B., A*n Empirical Study of U.S. Copyright Fair Use Opinions,1978-2005,* 156 U. Pa. L. Rev. 549, 610-11 (2008).

BETTIG R. V., *Copyrighting Culture: The Political Economy of Intellectual Property*, Taylor & Francis, New York, 1996, 12.

BHATIA P., *Data mining and data warehousing: Principles and Practical Techniques,* Cambridge University Press*,* 2019.

BIRNHACK M. D., *Mandatory Copyright: From Pre-Palestine to Israel, 1910-2007*, In SUTHERSANEN U., GENDREAU Y. (eds.), *A Shifting Empire: 100 Years of The Copyright Act 1911*, Edward Elgar, 2012.

BORGHI, M., AND KARAPAPA, S., *Copyright and Mass Digitization: a Cross-Jurisdictional Perspective*, Oxford University Press, 2013.

BORGMAN, C. L. . *Whose text, whose mining, and to whose benefit?* in *Quantitative Science Studies*, 2020, 1(3), 993-1000.

BOTTIS M., PAPADOPOULOS M., ZAMPAKOLAS C., GANATSIOU P., *Text and Data Mining in the EU 'Acquis Communautaire' Tinkering with TDM & Digital Legal Deposit,* in Erasmus Law Review vol. 12 n. 2, 2019.

BOYLE J., *(When) Is Copyright Reform Possible? Lessons from the Hargreaves Review*, in OKEDIJI R. L., *Copyright Law in an Age of Limitations and Exceptions*, Cambridge University Press, 2017.

BREMERT B., *Legal aspects of text mining publicly available data*, 1. Available at: https://www.datenschutzzentrum.de/uploads/projekte/itesa/Bremert-Legal_aspects_of_text_mining_publicly_available_data-paper.pdf

BRESNAHAN T. F., TRAJTENBERG M., *General purpose technologies 'Engines of growth'?* in *Journal of Econometrics*, 65(1), 1995, 83-108.

BROOKS, M., MURRAY-RUST, P., OPPENHEIM, C., *The Social, Political and Legal Aspects of Text and Data Mining (TDM)* in *D-Lib Magazine*, 20(11/12), December 2014. Available at: http://www.dlib.org/dlib/november14/brook/11brook.html.

BRYNJOLFSSON E., HITT L. M., KIM H. H., *Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?,* 2011.

BURRELL R., COLEMAN A., *Copyright Exceptions: The Digital Impact*, Cambridge University Press, 2005.

CANTATORE F., *Authors, Copyright, and Publishing in the Digital Era*, Information Science Reference, 2014, 13.

CALABRESE B., *Sulla dimensione imprenditoriale e societaria degli organismi di ricerca ai fini di text and data mining*, in *Contratto e impresa*, 2020, 4.

CARROLL, M. W., *Copyright and the Progress of Science: Why Text and Data Mining Is Lawful*, 53 *UC Davis L. Rev.* 893 (2019), 893–964.

CASO R., *Alle origini del copyright e del droit d'auteur: spunti in chiave di diritto e tecnologia*, in Trento Law and Technology Research Group, 2010. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2254259

CASO, R.*, Il conflitto tra diritto d'autore e ricerca scientifica nella disciplina del text and data mining della direttiva sul mercato unico digitale* in *Trento LawTech Research, Paper* nr. 38, 2020.

CASPERS M., GUIBAULT L., *A right to "read" for machines: assessing a black-box analysis exception for data mining*, in P*roceedings of the Association for Information Science and Technology* 53(1), 2016. Available at: https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/pra2.2016.14505301017

CHANG Y., *Dabates on Introduction of "Fair use" to the Copyright Act of Japan and Korea. Do Japan and Korea need Fair use?*, Comparative IP Academic Workshop Working Paper No. 2, 2009. Available at: http://www.win-cls.sakura.ne.jp/pdf/22/16.pdf

CHARLESWORTH, A., TONKIN, E. L. *If You Find Yourself in a Hole, Stop Digging: Legal and Ethical Issues of Text/Data Mining in Research*. in *Working with Text*, 2016, 3, 61–88.

CHAPMAN P., *CRISP-DM 1.0 Step-by-step data mining guide*, 2000. Available at: https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2015-16/kdd/files/CRISPWP-0800.pdf

CHENGUANG W., *Law-making functions of the Chinese courts: Judicial activism in a country of rapid social changes*, 1 *Frontiers of Law in China*, 2006, 524-549.

COOK J. S., COOK L. L., *Social, Ethical and Legal Issues of Data Mining*, in *Data Mining: Opportunities and Challenges,* Idea Group Publishing, 2002.

COURTNEY K. K., SAMBERG R., VOLLMER T., *Big data gets big help: Law and policy literacies for text data mining*, in *College & Research Libraries News 81 (4) Scholarly Communication*, 2020. Available at: https://crln.acrl.org/index.php/crlnews/article/view/24383

COX K. L., *Text and Data Mining and Fair Use in the United States*, Association of Research Libraries, 2015. Available at: https://www.arl.org/resources/text-and-data-mining-and-fair-use/

CRAIG C., *Locking Out Lawful Users: Fair Dealing and Anti-Circumvention in Bill C-32*. Available at: http://www.irwinlaw.com/content/assets/content-commons/666/CCDA%2007%20Craig.pdf.

CRAIG C., *The Changing Face of Fair Dealing in Canadian Copyright Law: A Proposal for Legislative Reform* in GEIST M., ed, *In the Public Interest: The Future of Canadian Copyright Law*, Irwin Law, Toronto, 2005.

D. ASAY C., *Is Transformative Use Eating the World*, 60 *B.C. L. Rev.* (2020).

DE FILIPPI P., GRACZ K., *Resolving the Crisis of Copyright Law in the Digital Environment: Reforming the "Copy-Right" into a "Reuse-Right"*, in Proceedings of the 7th International Conference on the Interaction of Knowledge Rights, Data Protection and Communication, KnowRight 20, 2013.
Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2318016

DERCLAY E., *The Database Directive*, in I. STAMATOUDI, P. TORREMANS, *EU Copyright Law*, 9.31.

DUCATO R., STROWEL A., *Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to "Machine Legibility*", CRIDES Working Paper Series, 2018.

EGER T., SCHEUFEN M., *The Past and The Future of Copyright Law: Technological Change and Beyond*, 5, in J. DE MOT, Liber Amicorum Boudewijin Bouckaert, die Keure, 2012.

FALCE V., *Direttiva Copyright 2019: fair use ed eccezioni al copyright tra esigenze di «apertura» e necessità di indirizzo*. Available at: https://www.filodiritto.com/direttiva-copyright-2019-fair-use-ed-eccezioni-al-copyright-tra-esigenze-di-apertura-e-necessita-di-indirizzo

FAYYAD U., PIATETSKY-SHAPIRO G., SMYTH P., *From Data Mining to Knowledge Discovery in Databases*, in *AI Magazine*, 1996, 17(3), 43.

FILIPPOV S., *Mapping Text and Data Mining in Academic and Research Communities* in *Europe* in *Lisbon Council Special Briefing: Text and Data Mining*, 2014.

FISCHMAN AFORI O., *The Evolution of Copyright Law and Inductive Speculations as to Its Future*, 19 *J. Intell. Prop. L.* 231 (2012).

FISHER W., *Theories of Intellectual Property*, 1, in S. MUNZER, *New Essays in the Legal and Political Theory of Property,* Cambridge University Press, 2001.

GARZA BARBOSA R., *Revisiting International Copyright Law*, 8 Barry L Rev 43 (2007).

GEIGER, C., FROSIO, G., BULAYENKO, O., *Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/Eu., Centre for International Intellectual Property Studies (CEIPI) Research Paper No. 2019-08,* 2019.

GEIGER C., FROSIO G., BULAYENKO O., *Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?,* in IIC 2018, Vol. 49, No. 7.

GEIGER C., *The Future of Copyright in Europe: Striking a Fair Balance between Protection and Access to Information*, Report for the Committee on Culture, Science and Education – Parliamentary Assembly, Council of Europe.

GEIGER C., SCHÖNHERR F., *The Information Society Directive*, in I. STAMATOUDI, P. TORREMANS, EU Copyright Law, 11.64.

GEIGER C., *The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Fundamental Right to Research and the failed Text and Data Mining Limitations in the CSDM Directive*, PIJIP/TLS Research Paper Series 2021, No. 66.

M. GEIST, *Copyright Consultation Submission*, 2 *Osgoode Hall Rev L Pol'y* (2009).

GEIST M., *Fairness Found: How Canada Quietly Shifted from Fair Dealing to Fair Use*, in *The Copyright Pentalogy: How the Supreme Court of Canada Shook the Foundations of Canadian Copyright Law* [online], Ottawa: Les Presses de l'Université d'Ottawa, University of Ottawa Press, 2013. Available at: http://books.openedition.org/uop/969

GERVAIS, D. *Exploring the interfaces between big data and intellectual property law*. In *Journal of Intellectual Property, Information Technology and E-Commerce Law,* 2019, 10(1).

GERVAIS D. J., *Making Copyright Whole: A Principled Approach to Copyright Exceptions and Limitations*, 5 University of Ottawa Law & Technology Journal. 1 (2008).

GERVAIS D., *The 1909 Copyright Act in International Context*, 26 Santa Clara High Tech. L.J. 185 (2012).

GHIDINI, G., & BANTERLE, F. *A Critical View on the European Commission's Proposal for a Directive on Copyright in the Digital Single Market*. in *Giurisprudenza Commerciale, 2018, 6, 921-933*.

GHIDINI G., *Exclusion and Access in Copyright Law: The unbalanced features of the European Directive "on Information Society"*, in *Rivista di Diritto Industriale*, 2013.

GIMBEL M., *Some Thoughts on the Implications of Trusted Systems for Intellectual Property Law*, in 50 Stan. L. Rev. (1998), 1671–1687. Available at: www.jstor.org/stable/1229309.

GINSBURG J. C., *Overview of Copyright Law*. in DREYFUSS R., PILA J. (eds.), *Oxford Handbook of Intellectual Property Law*, Oxford University Press, 2018.

GRACZ K., Brid*ging the Gaps Between Social and Legal Norms Concerning Protection of Intellectual and Artistic Creations: On the Crisis of Copyright Law in the Digital Era*, in The Journal of World Intellectual Property, 2013, 16(1-2).

GRIFFITHS J., Infopaq, BSA and the 'Europeanisation' of United Kingdom Copyright Law, in Media & Arts Law Review, Vol. 16, 2011. Available at SSRN: https://ssrn.com/abstract=1777027

GUIBAULT L., *Copyright Limitations and Contracts, An Analysis of the Contractual Overridability of Limitations on Copyright*, Kluwer Law International, The Hague, 2002.

GUIBAULT L., *Why Cherry-Picking Never Leads to Harmonisation: The Case of the Limitations on Copyright under Directive 2001/29/EC,* in 1 JIPITEC 55 (2010).

GUPTA G.K., *Introduction to Data Mining with Case Studies*, 3rd edition, PHI learning Private Limited, 2014, I.

HAN J., KAMBER J., PEI J., *Data Mining: Concepts and Techniques, 3rd Edition*, Morgan Kaufmann, 2011.

HANDKE, C., GUIBAULT, L., & VALLBÉ, J.-J.*, Is Europe Falling Behind in Data Mining? Copyright's Impact on Data Mining in Academic Research.* in *New Avenues for Electronic Publishing in the Age of Infinite Collections and Citizen Science: Scale, Openness and Trust - Proceedings of the 19th International Conference on Electronic Publishing, Elpub 2015, 120-130.*

HANNAY, W. M., STRAUCH, B., CARSON, B. M., & MONTGOMERY, J., *Legally Speaking: Of Mindfields and Minefields: Legal Issues in Text and Data Mining.* in *Against the Grain*, 2014, 26(1).

HE T., *Copyright Exceptions Reform and AI Data Analysis in China – A Modest Proposal*, in LEE J. A., HILTY R., LIU K. C., *Artificial intelligence and Intellectual Property*, Oxford University Press, 2021.

HESSE C., *The rise of Intellectual Property, 700 B.C. – A.D. 2000: An Idea in the Balance*, Daedalus Vol. 131, No. 2, On Intellectual Property, Spring 2002.

HILTY R. M., RICHTER H., in HILTY R. M., MOSCON V., *Modernisation of the EU Copyright Rules, Position Statement*, MPI for Innovation and Competition Research Paper No. 17-12.

HUA J. J., *Toward A More Balanced Approach: Rethinking and Readjusting Copyright Systems in the Digital Network Era*, Springer-Verlag, Berlin Heidelberg, 2014.

HUGENHOLTZ B., SENFTLEBEN M. R.F, *Fair Use in Europe: In Search of Flexibilities*, in Amsterdam Law School Legal Studies Research Paper No. 2012-39 Institute for Information Law Research Paper No. 2012-33, 2011.

HUGENHOLTZ P.B. (ed.), *The Future of Copyright in a Digital Environment*, Den Haag, Kluwer, 1996.

HUGENHOLTZ P. B., W*hy the copyright directive is unimportant and possibly invalid,* in EIPR, 2000.

JONDET N., *The Text and Data Mining Exception in the Proposal for a Directive on Copyright: Why the European Union Needs to Go Further than the Laws of Member States*' in *Propriété Intellectuelles* no.67, 2018, 25-35.

KANTARDZIC M. , *Data Mining: Concepts, Models, Methods, and Algorithms, 3rd Edition*, Wiley-IEEE Press, 2019, I.

KAMOCKI P., KETZAN E., WILDGANS J., WITT A., *New exceptions for Text and Data Mining and their possible impact on the CLARIN infrastructure*, Selected papers from the CLARIN Annual Conference 2018, Linköping Electronic Conference Proceedings 159.

KELLI A. TAVAST A. PISUKE H., *Copyright and Constitutional Aspects of Digital Language Resources: The Estonian Approach* in *Juridica International* XIX, 2012, 40-48. Available at: https://www.researchgate.net/publication/288558711_Copyright_and_constitutional_aspects_of_digital_language_resources_The_estonian_approach

KELLI A., *The conceptual bases for codifying Estonia's IP law and the main legislative changes: From the comparative approach to embedding drafted law into the socio-economic context*, in International Comparative Jurisprudence 1 (2015).

KEMP D., *Copyright on Steroids: In Search of an End to over Protection*, 41 McGeorge L. Rev. (2017).

KU R.S.R., *The Creative Destruction of Copyright: Napster and the New Economics of Digital Technology*, 69 U Chi L Rev 263 (2002). Available at: https://chicagounbound.uchicago.edu/uclrev/vol69/iss1/7/.

LEAFFER M. A., *Understanding Copyright Law*, Fifth Edition, LEXISNEXIS, 2010.

LEE Y.H., *United Kingdom Copyright Decisions and Legislative Developments 2014*, In IIC International Review of Intellectual Property and Competition Law, 46 (2), 2015.

LEVAL P. N., *Toward a Fair Use Standard*, 103 *Harv. L. Rev.* 1105, 1111 (1990), 1105-1136.

LIBERTINI M., *La nuova disciplina degli aiuti a favore di ricerca, sviluppo e innovazione*, in RAFFAELLI E. A., *Antitrust between EU and national law*, Milano-Bruxelles, 2015, 460 ss.

LOCKE J. P., *Two Treatises on Government,* 1689.

MARGONI T., *Text and Data Mining in Intellectual Property Law: Towards an Autonomous Classification of Computational Legal Methods*, CREATe Working Paper 2020/2021. Available at http://eprints.gla.ac.uk/215611/.

MARGONI T., *The Harmonisation of EU Copyright Law: The Originality Standard*, June 2016. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2802327

MATULIONYTE R., *Australian Copyright Law Impedes the Development of Artificial Intelligence: What Are the Options?*, in International Review of Intellectual Property and Competition Law (IIC) 52, 2021, 417–443.

MAUREL L., *L'exception TDM dans la loi numérique: mérites, limites et perspectives*, S.I.Lex, 11 November 2016.

MOORE G. E., *Cramming more components onto integrated circuits*, in Electronics, Volume 38, Number 8, April 19, 1965. Available at: https://newsroom.intel.com/wp-content/uploads/sites/11/2018/05/moores-law-electronics.pdf

M. NAIR, *Canada and Israel: Cultivating Fairness of Use,* PIJIP Research Paper no. 2012-04 American University Washington College of Law, Washington, D.C., 2012.

NETANEL N. W., *Making Sense of Fair Use*, 15 *Lewis & Clark L. Rev*. 715 (2011).

PASCUZZI G., *Il diritto dell'era digitale*, 5th edition, Il Mulino, Bologna, 2020.

POWELL W. W., SNELLMAN K., *The Knowledge Economy*, in *Annual Review of Sociology*, 30(1), 2004.

PROSSER W. L., *Privacy*, 48 Calif. L. Rev. 383 (1960), 383-423.

QUINTAIS J., *The New Copyright in the Digital Single Market Directive: A Critical Look*, in European Intellectual Property Review 2020(1).

REICHMANN J., OKEDIJI R. L., *When Copyright Law and Science Collide: Empowering Digitally Integrated Research Methods on a Global Scale*, in *96 Minnesota Law Review* 1362 *(2012).*

RENDAS T., *Destereotiping the Copyright Wars: The "fair use v. closed list" debate in the EU*. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2657482.

RICKETSON S., GINSBURG J., *International Copyright and Neighbouring Rights*, *The Berne Convention and Beyond*, Oxford University Press, 2006.

RODOTÀ S., *Repertorio di fine secolo*, Laterza, 1999.

O.-A. ROGNSTAD, *Restructuring the Economic Rights in Copyright – Some Reflections on an "Alternative Model"*, Journal of the Copyright Society of the U.S.A 62(4).

ROGNSTAD O.-A., POORT J., *The Right to a Reasonable Exploitation Concretized. An Incentive Based Approach* in HUGENHOLTZ B., *Copyright Reconstructed: Rethinking Copyright's Economic Rights in a Time of Highly Dynamic Technological and Economic Change*, Information Law Series.

ROSATI E., *Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and its Role in the Development of AI Creativity* in *Asia Pacific Law Review,* 2019, 27(2), 198-217.

ROSATI E., *Copyright in the EU: In Search of (In)Flexibilities*, in Journal of Intellectual Property Law & Practice, Volume 9, Issue 7, July 2014.

ROSATI E., *The Construction of Economic Rights in the Infosoc Directive*, in *Copyright and the Court of Justice of the European Union*, Oxford University Press, 2019.

ROSATI E., *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, 8. Available at: https://op.europa.eu/en/publication-detail/-/publication/fdb4ecaa-20f1-11e8-ac73-01aa75ed71a1/language-en

SAG M., *The New Legal Landscape for Text Mining and Machine Learning* in *Journal of the Copyright Society of the USA,* 2019, 66.

SAMUELSON P., *Regulating Technology Through Copyright Law: A Comparative Perspective*, 42 European Intellectual Property Review (2020).

SCHWARTZ E., *An Overview of the International Treatment of Exceptions*, PIJIP Research Paper no. 2014-02, American University Washington College of Law, Washington, D.C.

SCHWARTZ P., *The Value of Privacy Federalism,* in *Social Dimension of Privacy: Interdisciplinary Perspectives,* Cambridge University Press, 2015.

SENFTLEBEN M., *Bridging the Differences between Copyright's Legal Traditions – The Emerging EC Fair Use Doctrine*, in Journal of the Copyright Society of the U.S.A., 57(3), Spring 2010.

SENFTLEBEN M.R.F., *Copyright, Limitations and the Three-step test. An analysis of the Three-step test in International and EC Copyright Law,* Kluwer Law International, The Hague, 2004.

SENFTLEBEN M., *The International Three-Step Test: A Model Provision for EC Fair Use Legislation*, in 1 (2010) JIPITEC 67. Available at: https://www.jipitec.eu/issues/jipitec-1-2-2010/2605.

SGANGA C., *The right of reproduction*, in E. ROSATI (ed.), *The Routledge Handbook of EU Copyright Law*, Routledge, 2021. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3803999).

SOBEL B.L.W., *Artificial Intelligence's Fair Use Crisis*, 41 *Colum. J.L. & Arts* 45 (2017).

SOOKMAN B., GLOVER D., *Why Canada Should Not Adopt Fair Use: A Joint Submission to the Copyright Consultations,* 2 Osgoode Hall Rev L Pol'y (2009).

SPEDICATO G., *Principi di diritto d'autore*, Il Mulino, Bologna, 2020.

STEFIK M., *Shifting the Possible: How Trusted Systems and Digital Property Rights Challenge Us to Rethink DigitalPublishing,*12 Berkeley Tech L J 137 (1997).

STRAHILEVITZ L. J., *Toward a Positive Theory of Privacy Law*, 126 Harv. L. Rev. 2010 (2013).

SWANSON D., SMALHEISER N., *Assessing a gap in thebiomedical literature: magnesium deficiency and neurologic disease*. In *Neurosci. Res. Commun.15*, 1–9.

TAMURA Y., *Rethinking Copyright Institution*, in World Intellectual Property Organization Journal (W.I.P.O.J.) No. 1, 2009.

TROSOW S. E., *Copyright Consultation Submission* 2 Osgoode Hall Rev L Pol'y (2009).

TRUYENS M., VAN EEECKE P., *Legal aspects of text mining,* in *Computer Law & Security Review 30 (2)*, 2014.

TŪBAITĖ-STALAUSKIENĖ A., *EU Copyright Law: Developing Exceptions and Limitations Systematically – An Analysis of Recent Legislative Proposals*, in Baltic Journal of Law & Politics 11:2 (2018). Available at: https://sciendo.com/article/10.2478/bjlp-2018-0014

UENO T., *Rethinking the Provisions on Limitations of Rights in the Japanese Copyright Act, Toward a Japanese-style "Fair Use" Clause*, in AIPPI Journal, July 2009. Available at: http://www.f.waseda.jp/uenot/Japanese-style%20fair%20use%20AIPPI%202009.pdf

UENO T., *The Flexible Copyright Exception for "Non-Enjoyment" Purposes – Recent Amendment in Japan and Its Implication*, in GRUR International, 70(2), 2021.

WARE M., MABE M., *The STM Report: An overview of scientific and scholarly journal publishing, International Association of Scientific, Technical and Medical Publishers*, 2015.

WARREN S., BRANDEIS L., *The Right to Privacy*, 4 *Harv. L. Rev.* 193 (1890).

WILKINSON M. A., *Copyright, Collectives, and Contracts: New Math for Educational Institutions and Libraries*, in GEIST M., ed, *From "Radical Extremism" to "Balanced Copyright": Canadian Copyright and the Digital Agenda*, Irwin Law, Toronto, 2010.

YU P. K., *Customizing Fair Use Transplants*, in Laws, Vol. 7, Issue 1, Article 9, 2018, Texas A&M University School of Law Legal Studies Research Paper No. 17-78.

LEGISLATION

Australian Copyright Act 1968. Available at: http://www5.austlii.edu.au/au/legis/cth/consol_act/ca1968133/

Berne Convention for the Protection of Literary and Artistic Works (1886). For a brief summary on the contents provided by the Convention see: https://www.wipo.int/treaties/en/ip/berne/summary_berne.html

Bill C-32, *An Act to amend the Copyright Act*, 3d Sess, 40th Parl, 2010 (First Reading 2 June 2010). Available at: http://www.parl.gc.ca/HousePublications/Publication.aspx?Pub=Bill&Doc=C-32_1&Language=&Mode=1&Parl=40&Ses=3 .

Canada - Copyright Act (R.S.C., 1985, c. C-42). Available at: https://laws-lois.justice.gc.ca/eng/acts/C-42/Index.html

China - Copyright Law of the People's Republic of China (as amended up to the Decision of November 11, 2020, of the Standing Committee of the National People's Congress on Amending the Copyright Law of the People's Republic of China). Available at (Chinese only): https://wipolex.wipo.int/en/legislation/details/21065

China - Revision proposal for the Copyright Law of China (2020). Available at: http://www.iprdaily.cn/article_24614.html

Decreto Legislativo, n. 196 del 30 giugno 2003 "Codice in materia di protezione dei dati personali".

Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

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.

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

Estonia - Autoriõiguse ja autoriõigusega kaasnevate õiguste seaduse eelnõu (The Estonian draft Copyright and Related Rights Act), 2014.

Estonia - Estonian Copyright Act art. 19(3).

France - French Intellectual Property Code (Code de la Propriété Intellectuelle, CPI) AMENDEMENT N°180. https://www.assemblee-nationale.fr/14/amendements/3399/AN/180.asp.

Germany - Act on Copyrights and Related Rights (Urheberrechtsgesetz).

Green Paper on Copyright and the Challenge of Technology - Copyright Issues Requiring Immediate Action. COM (88) 172 final, 7 June 1988. Available at: http://aei.pitt.edu/1209/

Ireland - Copyright and Other Intellectual Property Law Provisions Bill 2018 (Bill 31 of 2018). Available at: https://www.oireachtas.ie/en/bills/bill/2018/31/

Ireland: Copyright and Other Intellectual Property Law Provisions Act 2019 (Number 19 of 2019). Available at: https://data.oireachtas.ie/ie/oireachtas/act/2019/19/eng/enacted/a1919.pdf

Irish Government, '*General Scheme of a Copyright Bill Approved by Government*', Department of Business, Enterprise and Innovation, 4 August 2016.

Israel Copyright Law. Available at: https://www.tau.ac.il/law/members/birnhack/IsraeliCopyrightAct2007.pdf

Japan - Copyright Law of Japan.

Legge n. 675 del 31 dicembre 1996 "Tutela delle persone e di altri soggetti rispetto al trattamento dei dati personali".

Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market - COM(2016) 593 final, European Commission, 14 September 2016 [COM(2016) 593 final].

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

Regulation (Eu) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

Taiwan Copyright Act, Laws and Regulations Database of The Republic of China. Available at: https://law.moj.gov.tw/ENG/LawClass/LawAll.aspx?pcode=J0070017

The Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) (1994). Full text available at: https://www.wto.org/english/docs_e/legal_e/trips_e.htm#art1

UK Copyright, Designs and Patents Act 1988. Available at: https://www.legislation.gov.uk/ukpga/1988/48/contents

United States - 17 U.S. Code Title 17. Available at: https://www.copyright.gov/title17/

WIPO Copyright Treaty (WCT). Available at: https://wipolex.wipo.int/en/text/295166

WIPO Performances and Phonograms Treaty (WPPT) (1996). Available at: https://wipolex.wipo.int/en/text/295578

REPORTS, LETTERS AND OFFICIAL DOCUMENTS

Agency for Cultural Affairs, "Cultural Council Copyright Subcommittee Legal Basic Issues Subcommittee Interim Summary" (2017) (in Japanese) www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/h2902_chukanmatome.pdf

Academics Against Press Publishers' Right. Academics launch final appeal to European Parliament. Available at: https://www.ivir.nl/academics-against-press-publishers-right/

ANTONIPILLAI J., LEE M. K., *Report: Intellectual Property and the U.S. Economy, Industries in Focus, 2016 Update*, report co-authored by the US Economics & Statistics Administration and the United States Patent and Trademark Office. Available at: https://www.uspto.gov/learning-and-resources/ip-motion/intellectual-property-and-us-economy.

Australian Government, Productivity Commission, Intellectual Property Arrangements – Inquiry report, 2016. Available at: https://www.pc.gov.au/inquiries/completed/intellectual-property/report

Brief – Statutory Review of the Copyright Act submitted by Pascale Chapdelaine, on behalf of Canadian intellectual property law scholars, Pascale Chapdelaine & Myra Tawfik (brief) (22 October 2018), online: House of Commons, INDU https://www.ourcommons.ca/Content/Committee/421/INDU/Brief/BR10166923/br-external/ChapdelainePascale01-e.pdf

BUCHHOLTZ S., BUKOWSKI M., ŚNIEGOCKI A., *Big and open data in Europe: A growth engine or a missed opportunity?*, the Warsaw Institute for Economic Studies (WISE Institute), report commissioned by demosEUROPA – Centre for European Strategy Foundation within the "Innovation and entrepreneurship" programme, 2014. Available at: http://wise-europa.eu/en/2014/02/26/big-open-data-in-europe-a-growth-engine-or-a-missed-opportunity/

Canada - "*Building a Nation of Innovators",* 2019. Available at: https://www.ic.gc.ca/eic/site/062.nsf/eng/h_00105.html

Bunka-Shingikai Chosakuken-Bunkakai Hōkokusho 2017 [The Report of the Subdivision on Copyright of the Council for Cultural Affairs 2017] (in Japanese). Available at: https://www.bunka.go.jp/seisaku/bunkashingikai/chosakuken/pdf/h2904_shingi_hokokusho.pdf

Call to Members of the European Parliament – Europe's citizens, startups, human rights organisations, publishers, creators, educators, cultural heritage professionals, librarians, and researchers ask for your support, 2 July 2018. Available at: https://copybuzz.com/wp-content/uploads/2018/07/Copyright-Open-Letter-on-EP-Plenary-Vote-on-Negotiation-Mandate.pdf

Canada, Parliament, House of Commons, Statutory Review of the Copyright Act: Report of the Standing Committee on Industry, Science and Technology, 42nd Parliament, 1st Session, No 16 (June 2019) [INDU Report], online: House of Commons https://www.ourcommons.ca/DocumentViewer/en/42-1/INDU/report-16 .

CARRE, S., GEIGER, C., LAPOUSTERLE, J., MACREZ, F., HASSLER, T., SEUBA, X., SCHHNHERR, F., & HEMMERLE-ZEMP, M. *Response of the CEIPI to the Public Consultation of the European Commission on the Review of the European Union Copyright Rules,* 2017.

CASPERS M., GUIBAULT L., *Baseline report of policies and barriers of TDM in Europe*, FutureTDM report, 2016. Available at: https://project.futuretdm.eu/publications/

CLARK, J., *Text Mining and Scholarly Publishing*, *Report Commissioned by the Publishing Research Consortium (PRC)*, Amsterdam, 2013.

Communication from the Commission of the European Communities, "Green Paper on copyright and the challenge of technology – Copyright issues requiring immediate action", COM(88) 172 final.

C*ommunication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A Digital Single Market Strategy for Europe*, COM(2015) 192 final, 6.5.2015. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52015DC0192

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Promoting a fair, efficient and competitive European copyright-based economy in the Digital Single Market, COM(2016)592. Available at: https://digital-strategy.ec.europa.eu/en/library/promoting-fair-efficient-and-competitive-european-copyright-based-economy-digital-single-market

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Towards a modern, more European copyright framework, COM(2015) 626 final, 2.12.2015. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2015%3A626%3AFIN

Communication from the Commission, "A European strategy for data", Brussels, 19.02.2020, COM(2020) 66 final.

Communication from the Commission, "Making the most of the EU's innovative potential, An intellectual property action plan to support the EU's recovery and resilience", 25 Nov. 2020, COM(2020) 760 final.

Communication from the Commission, "Shaping Europe's digital future", Brussels, 19.02.2020, COM(2020) 67 final.

Copyright and the Digital Economy (ALRC 122) - Incidental or Technical Use and Data and Text Mininghttps://www.alrc.gov.au/publication/copyright-and-the-digital-economy-alrc-report-122/11-incidental-or-technical-use-and-data-and-text-mining/

Copyright and the Digital Economy (ALRC 122) - Recommendations. 2. 12. 2013. Available at: https://www.alrc.gov.au/publication/copyright-and-the-digital-economy-alrc-report-122/recommendations-6/

Copyright and the Digital Economy (ALRC 122) - Terms of Reference. 29.06.2012. Available at: https://www.alrc.gov.au/inquiry/copyright-and-the-digital-economy/terms-of-reference-13/

Copyright Review Committee, "Copyright and Innovation: a Consultation Paper", for the Minister for Jobs, Enterprise and Innovation, Dublin 2012.

Do not block Europe from becoming a leader in Artificial Intelligence (AI): Adopt a broad and mandatory Text and Data Mining (TDM) exception. Available at: https://cdt.org/wp-content/uploads/2019/01/2019-01-15-Open-Letter-TDM-Copyright.pdf

Draft compromise amendments on Text and data mining. Version 1. 24 May 2018. Available at: https://www.communia-association.org/wp-content/uploads/2018/05/24052018-Draft-CA-on-Art.-3-v1.pdf

Draft compromise amendments on Text and data mining, 24.05.2018. Available at: https://www.communia-association.org/wp-content/uploads/2018/05/24052018-Draft-CA-on-Art.-3-v1.pdf

Draft Report on the proposal for a directive of the European Parliament and of the Council on copyright in the Digital Single Market (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)). Committee on Legal Affairs Rapporteur: Therese Comodini Cachia. Available at: https://www.europarl.europa.eu/doceo/document/JURI-PR-601094_EN.pdf?redirect

EARE's position on the JURI Committee Report on the Copyright Directive, 26 June 2018. Available at: https://eare.eu/assets/uploads/2018/06/EARE_Statement_JURI-Vote_Copyright-Directive.pdf

Eare's position on the results of the trilogue discussion on the copyright, Feb 22, 2019. Available at: https://eare.eu/eares-position-on-the-results-of-the-trilogue-discussions-on-the-copyright-directive/

Eurodoc Open Letter to European Parliament on Copyright Directive, 2 July 2018. Available at: http://eurodoc.net/news/2018/eurodoc-writes-open-letter-to-european-parliament-on-copyright-directive

European Alliance for Research Excellence, *24 Organisations urge Rapporteur Axel Voss MEP to strike a more ambitious deal on TDM*, 8 June 2018. Available at: https://eare.eu/24-organisations-urge-rapporteur-voss-reach-ambitious-deal-tdm/

European Alliance for Research Excellence, *The European Parliament's Position on the Copyright Directive Will Hold Back European Research And Innovation*, 13 September 2018. Available at: https://eare.eu/european-parliaments-position-copyright-hold-back-research-innovation/

European Commission. Commission agrees way forward for modernising copyright in the digital economy, MEMO/12/950. Brussels: 05.12.2012. Available at: http://europa.eu/rapid/press-release_MEMO-12-950_en.htm

European Commission. Communication from the Commission on content in the Digital Single Market, COM(2012) 789 final. Brussels: 18.12.2012. Available at: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52012DC0789

European Commission - Directorate General Internal Market and Services*, Report on the responses to the Public Consultation on the Review of the EU Copyright Rules,* July 2014. Available at: https://digital-strategy.ec.europa.eu/en/library/modernisation-eu-copyright-rules-useful-documents

European Commission (2013). "Licenses For Europe: Structured stakeholder dialogue 2013". Available at: https://ec.europa.eu/licences-for-europe-dialogue/

European Commission, "White Paper on Artificial Intelligence - A European approach to excellence and trust", Brussels, 19.02.2020, COM(2020) 65 final.

EU Copyright Reform Proposals Unfit for the Digital Age, (Open Letter to Members of the European Parliament and the Council of the European Union), 24 February 2017. Available at: https://www.create.ac.uk/wp-content/uploads/2017/02/OpenLetter_EU_Copyright_Reform_24_02_2017.pdf

European University Association, *EU Copyright Reform: European Parliament vote is a missed opportunity to bolster R&I*, 13 September 2018. Available at: https://eua.eu/news/150:eu-copyright-reform-european-parliament-vote-is-a-missed-opportunity-to-bolster-r-i.html

Government of Canada, A Consultation on a Modern Copyright Framework for Artificial Intelligence and the Internet of Things, 16 July 2021. Available at: https://www.ic.gc.ca/eic/site/693.nsf/eng/00316.html#s21

Government of Canada, News Release: The Government of Canada Launches Consultation on a Modern Copyright Framework for AI and the Internet of Things, 16 July 2021. Available at: https://www.canada.ca/en/innovation-science-economic-development/news/2021/07/the-government-of-canada-launches-consultation-on-a-modern-copyright-framework-for-ai-and-the-internet-of-things.html

GUIBAULT L., WESTKAMP G., RIEBER-MOHN T., *Study on the implementation and effect in Member States' laws of Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the Information Society*, Report commissioned by the European Commission's Internal Market Directorate- General, in response to the invitation to tender MARKT/2005/07/D, Amsterdam Law School Legal Studies Research Paper No. 2012-28, Institute for Information Law Research Paper No. 2012-23, 2007.

HARGREAVES, I., GUIBAULT, L., HANDKE, C., VALCKE, P., MARTENS, B., LYNCH, R., & FILIPPOV, S., *Standardisation in the area of innovation and technological development, notably in the field of Text and Data Mining: Report from the Expert Group,* 2014.

HARGRAVES I., *Digital Opportunity - A Review of Intellectual Property and Growth, An Independent Report by Professor Ian Hargreaves*, May 2011. Available at: https://www.gov.uk/government/publications/digital-opportunity-review-of-intellectual-property-and-growth

HM Government, *Modernising copyright: a modern, robust and flexible framework. Government response to consultation on copyright exceptions and clarifying copyright law*, 2012. Available at: https://www.mpaonline.org.uk/wp-content/uploads/2017/09/Modernising_Copyright_-_a_modern_robust_and_flexible_framework_-_Government_response.pdf

HM Government. *The Government Response to the Hargreaves Review of Intellectual Property and Growth*. August 2011. Available at: http://www.ipo.gov.uk/ipresponse-full.pdf

H.R. Rep. No. 94-1476. Available at: https://www.copyright.gov/history/law/clrev_94-1476.pdf

"*IDC's Global DataSphere Forecast Shows Continued Steady Growth in the Creation and Consumption of Data*", May 2020.
Available at: https://www.idc.com/getdoc.jsp?containerId=prUS46286020

Intellectual property rights intensive industries and economic performance in the European Union, Industry-Level Analysis Report, September 2019, Third edition. A joint project between the European Patent Office and the European Union Intellectual Property Office. Available at: https://www.epo.org/service-support/publications.html?pubid=201#tab3

Intellectual Property Strategic Program 2008. Strengthening the Intellectual Property Strategy, Targeting the World. Intellectual Property Strategy Headquarters, June 18, 2008. Available at: http://www.kantei.go.jp/jp/singi/titeki2/keikaku2008_e.pdf

Intellectual Property Strategic Program 2016. Intellectual Property Strategy Headquarters. May 2016. Available at: https://www.kantei.go.jp/jp/singi/titeki2/kettei/chizaikeikaku20160509_e.pdf

Irish Copyright Review Committee, '*Modernising Copyright'*, Department of Jobs, Enterprise and Innovation, October 2013. Available at: http://www.cearta.ie/wp-content/uploads/2013/10/CRC-Report.pdf

JEAN-CLAUDE JUNCKER, *A New Start for Europe: My Agenda for Jobs, Growth, Fairness and Democratic Change*, Political Guidelines for the next European Commission, 14 July 2014. Available at https://www.eesc.europa.eu/resources/docs/jean-claude-juncker---political-guidelines.pdf

Legislative Affairs Office of the State Council P.R. China, *Notice About Circular of the Legislative Affairs Office of the State Council on Promulgating the Copyright Law of the People's Republic of China*

(Draft Revision for Review) for Public Consultation (2014). Available at: http://www.gov.cn/xinwen/2014-06/10/content_2697701.htm

LIBER, *Europe Needs A Broad & Mandatory TDM Exception*, 13 November 2018. Available at: https://libereurope.eu/article/europe-needs-a-broad-mandatory-tdm-exception/

Letter from European Technology SMEs, Open Access Publishers and the Research Sector WG 4, May 22nd, 2013. Available at: https://libereurope.eu/wp-content/uploads/2013/05/Letter_of_withdrawalL4E_TDM_May-24_4.pdf

LIBER (Association of European Research Libraries) (2013). "Stakeholders representing the research sector, SMEs and open access publishers withdraw from Licences for Europe". Available at: https://libereurope.eu/article/stakeholders-representing-the-research-sector-smes-and-open-access-publishers-withdraw-from-licences-for-europe-2/

Licences for Europe, Working Group 4 – Text and data mining, Conclusions of the First meeting of 4 February 2013. Available at: https://ec.europa.eu/licences-for-europe-dialogue/en/content/wg4-conclusions-meetings.html

Licences for Europe, Working Group 4 – Text and data mining, Conclusions of the Second meeting of 8 March 2013. Available at: https://ec.europa.eu/licences-for-europe-dialogue/en/content/wg4-conclusions-meetings.html

Letter sent by concerned participants in response to the "Licences for Europe – A Stakeholder Dialogue" workshop on text and data mining for scientific research purposes, 26.02.2013. Available at: https://libereurope.eu/document/licences-for-europe-a-stakeholder-dialogue-text-and-data-mining-for-scientific-research-purposes-working-group/

MANYIKA J., CHUI M., BROWN B., BUGHIN J., DOBBS R., ROXBURGH C., HUNG BYERS A., *Big data: The next frontier for innovation, competition and productivity*, McKinsey Global Institute, May 2011. *Report* downloadable from the website: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation#

MARTIN J., DE CARVALHO L., *Rapport de la mission sur l'exploration de données* («*Text and Data mining*»), Conseil Supérieur de la Propriété Littéraire et Artistique, July 2014 [Rapport CSPLA]. Available at: http://www.culturecommunication.gouv.fr/content/download/105451/1236408/version/1/file/Rapport%20Text%20and%20Data%20Mining%20(exploration%20de%20donn%C3%A9es).pdf

MCDONALD, D., & KELLY, U. *The Value and Benefits of Text Mining*, JISC, The Higher Education Funding Council for England, 2012.

National Research Council, The Digital Dilemma: Intellectual Property in the Information Age, 2000.

Net Jidai niokeru Chitekizaisan Senmonchôsakai, Digital Net Jidai niokeru Chizaiseido no Arikata nitsuite (Report) [Intellectual Property System for the Digital-Network Age] 11 (2008) (in Japanese).

Organization for Economic Co-operation and Development (OECD) Report: Chapter 7 - *Legal Aspects of Open Access To Publicly Funded Research*, in *Enquiries into Intellectual Property's Economic Impact*, 2015. Available at: https://www.oecd.org/sti/ieconomy/Chapter7-KBC2-IP.pdf

Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market - Presidency compromise proposal regarding Articles 2 to 9 (Malta Presidency), 8 May 2017. Available at: https://data.consilium.europa.eu/doc/document/ST-8929-2017-INIT/en/pdf

Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market - Presidency compromise proposal (consolidated version) and state of play, 13 December 2017. Available at: https://data.consilium.europa.eu/doc/document/ST-15651-2017-INIT/en/pdf

Proposal for a Directive of the European Parliament and the Council on copyright in the Digital Single Market - Presidency compromise proposal. https://www.parlament.gv.at/PAKT/EU/XXVI/EU/01/76/EU_17608/imfname_10800647.pdf

Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market - Agreed negotiating mandate, 25 May 2018. Available at: https://www.consilium.europa.eu/media/35373/st09134-en18.pdf

Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market - Update of negotiation mandate. 7 February 2019. Available at: https://data.consilium.europa.eu/doc/document/ST-5893-2019-ADD-1/en/pdf

Provisions of the Supreme People's Court on Several Issues concerning the Application of Law in Hearing Civil Dispute Cases Involving Infringement of the Right of Dissemination on Information Networks (2012) (PRC), Art. 5 (2).

Public Consultation on the Review of the EU Copyright Rules, 2013. Available at: https://digital-strategy.ec.europa.eu/en/library/modernisation-eu-copyright-rules-useful-documents

Report on the Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)) (Rapporteur: MEP Axel Voss), Amendment 65.

Second Version of the EU parliament draft proposal. https://twitter.com/communia_eu/status/1004047868280360963/photo/1

Setting Europe in Motion: President-elect Juncker's Main Messages from his speech before the European Parliament, Statement in the European Parliament plenary session ahead of the vote on the

College, Strasbourg, 22 October 2014. Available at: https://www.eesc.europa.eu/resources/docs/jean-claude-juncker---political-guidelines.pdf

Several Opinions of the Supreme People's Court on Some Issues in Fully Giving Rein to the Funcion of Intellectual Property Rights Adjudication in Promoting the Great Development and Flourishing of Socialist Culture and Stimulating the Indigenous and Coordinated Development of Economy (2011) (PRC).

STONER R., DUTRA J., *Copyright Industries in the U.S. Economy, The 2020 Report*, prepared for the International Intellectual Property Alliance (IIPA), 2020. Available at: https://www.iipa.org/files/uploads/2020/12/2020-IIPA-Report-FINAL-web.pdf.

STRYCHARZ J., *Trend analysis, future applications and economics of TDM*, Future TDM, 2014. Available at: https://www.futuretdm.eu/blog/legal-policies/trend-analysis-future-applications-and-economics-of-tdm/.

The Copyright Directive is failing (Open Letter to Members of the European Parliament and the Council of the European Union), 26 April 2018. Available at: https://www.create.ac.uk/wp-content/uploads/2018/04/OpenLetter_EU_Copyright_Research_Centres_26_04_2018.pdf

The Software Alliance to the Standing Committee on Industry, Science and Technology Regarding the 2018 Statutory Review of the Copyright Act, BSA: the Software Alliance (brief) (26 September 2018), online: House of Commons, INDU https://www.ourcommons.ca/Content/Committee/421/INDU/Brief/BR10057455/br-external/BSATheSoftwareAlliance-e.pdf .

E., TRIAILLE, J.-P., D'ARGENTEUIL, J. DE M., & FRANCQUEN, A. DE., *Study on the legal framework of text and data mining (TDM),* Directorate-General for Financial Stability, F. S. and C. M. U., Directorate-General for Internal Market Entrepreneurship and SMEs, I., Commission, 2014.

UK IPO, Artificial intelligence and intellectual property: call for views, 7 September 2020. Available at: https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views.

UK IPO, Consultation outcome - Government response to call for views on artificial intelligence and intellectual property, Updated 23 March 2021. Available at: https://www.gov.uk/government/consultations/artificial-intelligence-and-intellectual-property-call-for-views/government-response-to-call-for-views-on-artificial-intelligence-and-intellectual-property

UK IPO. 'I*mpact Assessment – Exception for Copying of Works for Use by Text and Data Analytics* (BIS0312) (Final Impact Assessement – Part 3)', 13 December 2012. http://www.legislation.gov.uk/ukia/2014/156/pdfs/ukia_20140156_en.pdf.

Understanding Copyright and Related Rights, WIPO, 2016. Available at: https://www.wipo.int/edocs/pubdocs/en/wipo_pub_895_2016.pdf

Untitled, Creative Commons (brief) (15 May 2018), online: House of Commons, INDU https://www.ourcommons.ca/Content/Committee/421/INDU/Brief/BR9887146/br-external/CreativeCommons-e.pdf .

Untitled, Mark Hayes (brief) (2 November 2018), online: House of Commons, INDU https://www.ourcommons.ca/Content/Committee/421/INDU/Brief/BR10166884/br-external/HayesMark-e.pdf .

CASES

European Union and European national courts

ECJ Apis-Hristovich v. Lakorda.

C- 393/09 *Bezpečnostní softwarová asociace v. Ministerstvo kultury* [2010] E.C.R. I-13971.

Case 304/07 Directmedia Publishing GmbH v. Albert-Ludwigs-Universität Freiburg [2008] ECLI:EU:C:2008:552.

Case C-762/19 SIA 'CV-Online Latvia' v. SIA 'Melons'.

CJEU, C-203/02, The British Horseracing Board Ltd and Others v. William Hill Organization Ltd (9 November 2004).

C-145/10 *Eva-Maria Painer v. Standard VerlagsGmbH and Others* [2011] E.C.R I-12533.

ECJ 9 November 2004, C-444/02 (Fixtures Marketing v. OPAP).

ECJ 9 November 2004, C-46/02 (Fixtures Marketing v. Veikkaus Oy).

ECJ 9 November 2004, C-338/02 (Fixtures Marketing v. Svenska Spel)

C-604/10 *Football Dataco v. Yahoo! UK and Others* [2012] EU:C:2012:115.

C-403/08 *Football Association Premier League and Others v. QC Leisure and Others*

Case C-5/08 Infopaq International A/S v Danske Dagblades Forening [2009] ECR I-6569

CJEU, C-467/08, Padawan v SGAE (21 October 2008) ECLI:EU:C:2010:620

CJEU Innoweb v. Wegener.

C-429/08 *Karen Murphy v. Media Protection Services* [2011] E.C.R. I-09083

Tribunal de Grande Instance, Paris, 3me ch, 30 April 2004, confirmed by Cour de Cassation, Paris, 1rech, 28 February 2006

Tribunale di Milano, Section for intellectual property, 1 July 2009, no. 8787.

United States

Am. Geophysical Union v. Texaco, Inc., 60 F.3d 913, 930 (2d Cir. 1994).

Authors Guild v. Google, 770 F.Supp.2d 666 (S.D.N.Y. 2011).

Authors Guild v Google No. 13-4829-cv (2d Cir. Oct. 16, 2015).

Authors Guild v. HathiTrust, 755 F.3d 87 (2d Cir. 2014).

Authors Guild v. HathiTrust, 902 F.Supp.2d 445 (SDNY October 10, 2012).

Bill Graham Archives v. Dorling Kindersley Ltd., 448 F.3d 605, 612 (2d Cir. 2006).

Blanch v. Koons, 467 F.3d 244, 256 (2d Cir. 2006).

Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994).

A.V. v. iParadigms, LLC (4th Cir. 2009).

Eldred v Ashcroft, 537 US 186, 219-20 (2003).

Feist Publ'ns Inc v Rural Tel Serv Co 499 US 340, 345 (1991).

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

Fox v. TVEyes (S.D.N.Y. 2014).

Katz v. United States, 389 U.S. 347 (1967).

Kelly v. Arriba Soft, 336 F.3d 811 (9th Cir. 2003).

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

Sony Corporation of America v. Universal City Studios, Inc. 464 U.S. 417 (1984).

Sundeman v. Seajay Soc'y, Inc., 142 F.3d 194, 206 (4th Cir. 1998).

White v. West (S.D.N.Y. 2014).

Israel

C.A. 2687/92 Geva v. Walt Disney Co., 48(1) P.D. 251 (1993).

C.A. 8393/96 *Mifal Ha'Pais v. The Roy Export Establishing Co.*, 54(1) P.D. 577, 597 (2000) [Heb.].

Australia

National Rugby League Investments Pty Limited v. Singtel Optus Pty Ltd [2012] FCAFC 59, 27 April 2012.

Canada

Alberta Education v. Access Copyright, 2010 F.C.A. 198 (2010). Available at http://www.canlii.org/en/ca/fca/doc/2010/2010fca198/2010fca198.html.

Bishop v. Stevens, 1990 CanLII 75 (SCC), [1990] 2 SCR 467. Available at: http://canlii.ca/t/1fsv7

CCH Canadian Ltd. v. Law Society of Upper Canada, [2004] 1 S.C.R. 339, 2004 SCC 13. Available at: https://decisions.scc-csc.ca/scc-csc/scc-csc/en/item/2125/index.do

Michelin v. Caw, 1996 CanLII 11755 (FC), [1997] 2 FC 306. Available at: https://canlii.ca/t/4g4v

SOCAN, Re: Sound, CSI, Connect/SOPROQ, Artisti – Tariff for Commercial Radio, 2011-2017 (2016).

Théberge v. Galerie d'Art du Petit Champlain inc., 2002 SCC 34 (CanLII), [2002] 2 SCR 336. Available at: https://canlii.ca/t/51tn.

Warman v. Fournier, 2012 FC 803 (CanLII).

China

Beijing Higher People's Court, GMZZ No. 1221.

Wang Shen v Google Inc. et al. (2012) Beijing 1st Intermediary People's Court, YZMCZ No. 1321; (2013).

BLOGS AND WEBSITES

BURANYI S., I*s the staggeringly profitable business of scientific publishing bad for science?*, The Guardian, 27 June 2017. Available at: https://www.theguardian.com/science/2017/jun/27/profitable-business-scientific-publishing-bad-for-science

CHEE F. Y., *Commission starts legal action against 23 EU countries over copyright rules*, Reuters, 26 June 2021. Available at: https://www.reuters.com/world/europe/commission-starts-legal-action-against-23-eu-countries-over-copyright-rules-2021-07-26/

COMMUNIA DSM Directive Implementation Portal. Available at: https://www.notion.so/DSM-Directive-Implementation-Portal-97518afab71247cfa27f0ddeee770673.

COMMUNIA Eurovision DSM Contest. Available at: https://eurovision.communia-association.org.

Communia Association, *Can Voss salvage the text and data mining exception*?, 4 June 2018. Available at: https://www.communia-association.org/2018/06/04/can-voss-salvage-text-data-mining-exception/

Communia Association, *JURI rapporteur proposes to fix most egregious flaws of the copyright reform proposal*, 20 March 2017. Available at: https://www.communia-association.org/2017/03/20/lead-mep-proposes-fixes-egregious-flaws-copyright-reform-proposal/

CREATe, EU Copyright Reform – Evidence on the Copyright in the Digital Single Market Directive. Available at: https://www.create.ac.uk/policy-responses/eu-copyright-reform/#ert_pane4-0

Discours d'ouverture du Congrès littéraire international Victor Hugo 7 juin 1878. Available at: https://fr.wikisource.org/wiki/Discours_d%27ouverture_du_Congrès_littéraire_international

EARE, Japan Amends Its Copyright Legislation to Meet Future Demands In Ai And Big Data, *3 September, 2018*. Available at: https://eare.eu/japan-amends-tdm-exception-copyright/

EFRONI Z., *Israel's Fair Use*, Blog of The Center for Internet and Society at Stanford Law School, 30 January 2008. Available at: http://cyberlaw.stanford.edu/blog/2008/01/israel's-fair-use

"Estonian president delights in country's high proportion of unicorns", The Guardian. Availbale at: https://www.theguardian.com/world/2018/jun/29/estonia-unicorns-president-kersti-kaljulaid-delight

EU Parliament, The Ordinary Legislative Procedure. Available at: https://www.europarl.europa.eu/olp/en/ordinary-legislative-procedure/overview

European Parliament, A "rapporteur" - the person who presents reports to Parliament. Available at: https://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+IM-PRESS+20060725STO09938+0+DOC+XML+V0//EN

Facebook profile privacy policies. Available at: https://www.facebook.com/about/basics/manage-your-privacy/profile

Facebook Terms of Service. Available at: https://www.facebook.com/terms.php

FURNAS A., *Everything You Wanted to Know About Data Mining but Were Afraid to Ask*, The *Athlantic*, 2012. Available at: https://www.theatlantic.com/technology/archive/2012/04/everything-you-wanted-to-know-about-data-mining-but-were-afraid-to-ask/255388/

GEIST M., *How Copyright Reform Could Support Canada's Supercluster Investment*, Centre for International Governance Innovation (CIGI), 19 March 2019. Available at: https://www.cigionline.org/articles/how-copyright-reform-could-support-canadas-supercluster-investment/

How Estonia became Europe's tech hotspot, The New Economy. Available at: https://www.theneweconomy.com/technology/how-estonia-became-europes-tech-hotspot

"*How Tesco became Britain's top supermarket*", MoneyWeek, 2007. Available at: https://moneyweek.com/31267/how-tesco-became-britains-top-supermarket

HUGENHOLTZ B., *The New Copyright Directive: Text and Data Mining (Articles 3 and 4)*, Kluwer Copyright Blog, July 24 2019. Available at: http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/

JAO N., *Japan's 'Society 5.0' will integrate cutting-edge tech at all levels*, ITU News, 29 June 2017. Available at: https://news.itu.int/japans-society-5-0-will-integrate-cutting-edge-tech-at-all-levels/

JAPANGOV, the Government of Japan, *Artificial Intelligence: A Rival for Humans, or a Partner?* Available at: https://www.japan.go.jp/tomodachi/2018/spring2018/artificial_intelligence.html

JONG S. J., *Fair Use in Korea*, infojustice, 27 february 2017. Available at: https://infojustice.org/archives/37819

LI R., *History of Data Mining*, KDnuggets. Available at: https://www.kdnuggets.com/2016/06/rayli-history-data-mining.html

MAYO M., "*Where Analytics, Data Science, Machine Learning Were Applied: Trends and Analysis*", KDnuggets. Available at: https://www.kdnuggets.com/2019/03/poll-analytics-data-science-ml-applied-2018.html.

New lead MEP could shift talks on contentious copyright bill, Euractiv. Available at: https://www.euractiv.com/section/digital/news/new-lead-mep-could-shift-talks-on-contentious-copyright-bill/

PASCUZZI G., *Le valutazioni degli studenti le fa l'algoritmo (a proposito di Educational Data Mining)*, in *L'Adige*, 13 ottobre 2020. Available at: https://www.giovannipascuzzi.eu/2020/10/13/le-valutazioni-degli-studenti-le-fa-lalgoritmo-a-proposito-di-educational-data-mining/

PIATETSKY G., "*What Data You Analyzed – KDnuggets Poll Results and Trends*", KDnuggets. Available at: https://www.kdnuggets.com/2017/04/poll-results-data-analyzed.html

PIEVATOLO M. C., *L'età del privilegio*, Rivista Il Mulino. Available at: https://www.rivistailmulino.it/a/l-et-del-privilegio

ROSATI E., KRETCHMER M., *EU copyright reform: quo vadis*?, in The IPKat, 11 May 2017. Available at: https://ipkitten.blogspot.com/2017/05/eu-copyright-reform-quo-vadis.html

ROSATI E., *An EU text and data mining exception: will it deliver what the Digital Single Market Strategy promised?*, The IPKat, 22 May 2017. Available at: https://ipkitten.blogspot.com/2017/05/an-eu-text-and-data-mining-exception.html

ROSATI E., *Irish Government proposes introduction of Irish text and data mining exception*, The IPKat, March 2018. Available at: https://ipkitten.blogspot.com/2018/03/irish-government-proposes-introduction.html

TOONDERS J., *Data Is the New Oil of the Digital Economy*, Wired. Available at: https://www.wired.com/insights/2014/07/data-new-oil-digital-economy/

VOLLMER T., *European Parliament legal affairs committee pushes for strong exception for text and data mining,* Communia, 21 March 2017. Available at: https://www.communia-association.org/2017/03/21/european-parliament-legal-affairs-committee-pushes-strong-exception-text-data-mining/

VOLLMER T., *The wandering saga of the text and data mining exception in the EU copyright reform*, 5 June 2017. Available at: https://www.communia-association.org/2017/06/05/wandering-saga-text-data-mining-exception-eu-copyright-reform/

VOLLMER T., *The Estonian Presidency's new (and optional) TDM exception: small potatoes*, Communia Association, 4 December 2017. Available at: https://www.communia-association.org/2017/12/04/estonian-presidencys-new-optional-tdm-exception-small-potatoes/

VOLLMER T., *Proposed Council compromise on TDM: still not good enough*, 6 April 2018. Available at: https://www.communia-association.org/2018/04/06/proposed-council-compromise-tdm-still-not-good-enough/

WATSON I., *How Alan Turing Invented the Computer Age*, 2012. Available at: https://blogs.scientificamerican.com/guest-blog/how-alan-turing-invented-the-computer-age/

"*What main methodology are you using for your analytics, data mining, or data science projects? Poll*", KDnuggets. Available at: https://www.kdnuggets.com/polls/2014/analytics-data-mining-data-science-methodology.html

What role for the European Parliament's committees and how do they work? Available at: https://epthinktank.eu/2020/05/06/what-role-for-the-european-parliaments-committees-and-how-do-they-work/

The Student Paper Series of the Trento LawTech Research Group is published since 2010

http://lawtechnew.jus.unitn.it/main-menu/paper-series/student-paper-series-of-the-trento-lawtech-research-group/2/

Freely downloadable papers already published:

STUDENT PAPER N. 73

Risarcimento del danno da violazione dei diritti di proprietà intellettuale e retroversione degli utili. Un'analisi comparata

BRUNO FEDERICO (2022), Risarcimento del danno da violazione dei diritti di proprietà intellettuale e retroversione degli utili. Un'analisi comparata, Trento Law and Technology Research Group, Student Paper Series; 73. Trento: Università degli Studi di Trento. DOI: 10.5281/zenodo.5878282

STUDENT PAPER N. 72

Eccezioni e limitazioni al diritto d'autore nell'Unione Europea: profili critici e spunti comparatistici applicati al settore GLAM alla luce dell'emergenza COVID-19

MARONI ELEONORA (2022), Eccezioni e limitazioni al diritto d'autore nell'Unione Europea: profili critici e spunti comparatistici applicati al settore GLAM alla luce dell'emergenza COVID-19, Trento Law and Technology Research Group, Student Paper Series; 72. Trento: Università degli Studi di Trento. DOI:10.5281/zenodo.5878211

STUDENT PAPER N. 71

L'animal welfare nelle filiere alimentari: etichettatura e certificazioni

ZANON MIRIANA (2021), L'animal welfare nelle filiere alimentari: etichettatura e certificazioni, Trento Law and Technology Research Group, Student Paper Series; 71. Trento: Università degli Studi di Trento. 978-88-8443-959-8

STUDENT PAPER N. 70

Aggiornamenti di diritto agroalimentare nella riflessione dottrinale angloamericana

ANADOTTI, ELENA; DI GIOVANNI, SILVIA; FREZZA, ANNA CAROLINA; HOSSU, LORENA PATRICIA; MARCONATO, ELENA; NOSCHESE, ANGELA; PENDENZA, ALICE; PEPE, FRANCESCO; PIEROBON, VALERIA; POLI, ELISA; PURITA, CLAUDIA;

RAFFA, DJAMILA; ROTONDI, SERGIO ANDREA; SANTOLIN, GAIA – a cura di IZZO, UMBERTO; FERRARI, MATTEO (2021), Aggiornamenti di diritto agroalimentare nella riflessione dottrinale angloamericana, Trento Law and Technology Research Group, Student Paper Series; 70. Trento: Università degli Studi di Trento. 978-88-8443-958-1

STUDENT PAPER N. 69

Diritto del turismo e Covid-19: cosa è cambiato nella seconda estate pandemica

ANGIARI, YOUSSEF; ARZARELLO, ANDREA; AZILI, FEDERICO; BONOMELLI, CHIARA; BUBBOLA, IRENE; CADAMURO, CLAUDIA; CARRETTA, ANNA; CONDOTTA, ALESSANDRO; DA PRATO, MARIKA; DAL TOSO, VIRGINIA; DE AGOSTINI, FILIPPO; DE FRANCESCHI, SERENA; DELL'EVA, MARTINA; DELMARCO, MARTINA; DELLA MURA, MARCO; DI MASCIO, FRANCESCA; FIUTEM, LORENZO; GENNARA, GIULIA; INNOCENTI, ALBERTO; LORIERI, ANNA; MAFFEI, BEATRICE; MARCOLINI, ALESSIA; MANZO, ARIANNA; MINERVINI, MONICA MARIA; MURESAN, ANAMARIA ELENA; NARDIN, NICOLÒ; PAISSAN, FILIPPO; PAISSAN, INGMAR; PANERO, MARTINA; PAVALEANU, CRISTIAN; RIZ, FRANCESCA; SCARSELLA, ALESSIA; SCODANIBBIO, GIULIA; SORRENTINO, MARIAROSA; TUCCI, GIULIANA; VIGNOLI, MARTINA; ZACCARIN, STEPHANIE; ZUCAL, SARA; IZZO, UMBERTO (a cura di) (2021), Diritto del turismo e Covid-19: cosa è cambiato nella seconda estate pandemica, Trento Law and Technology Research Group, Student Paper Series; 69. Trento: Università degli Studi di Trento. 978-88-8443-954-3

STUDENT PAPER N. 68

La protezione dei dati relativi alla salute nell'era dei Big Data. Un'analisi sulla sanità digitale in dialogo tra diritto e tecnologia.

LIEVORE ANNA (2021), La protezione dei dati relativi alla salute nell'era dei Big Data. Un'analisi sulla sanità digitale in dialogo tra diritto e tecnologia, Trento Law and Technology Research Group, Student Paper Series; 68. Trento: Università degli Studi di Trento. 978-88-8443-903-1

STUDENT PAPER N. 67

«Cuius commoda, eius et incommoda»: l'art. 2049 del codice civile nella gig economy

PILZER LARA (2021), «Cuius commoda, eius et incommoda»: l'art. 2049 del codice civile nella gig economy, Trento Law and Technology Research Group, Student Paper Series; 67. Trento: Università degli Studi di Trento. 978-88-8443-946-8

STUDENT PAPER N. 66

La responsabilità sanitaria nel post covid-19: scenari e proposte per affrontare il contenzioso

PRIMICERI GIORGIA (2021), La responsabilità sanitaria nel post Covid-19: scenari e proposte per affrontare il contenzioso, Trento Law and Technology Research Group, Student Paper Series; 66. Trento: Università degli Studi di Trento. 978-88-8443-945-1

STUDENT PAPER N. 65

Legal design e sanità digitale: un innovativo approccio per favorire la tutela dei dati personali

FRANCESCO TRAVERSO (2021), Legal design e sanità digitale: un innovativo approccio per favorire la tutela dei dati personali, Trento Law and Technology Research Group, Student Paper Series; 65. Trento: Università degli Studi di Trento. 978-88-8443-943-7

STUDENT PAPER N. 64

L'appropriazionismo artistico nell'arte visuale: una comparazione tra Italia e Stati Uniti
LAURA DI NICOLA (2021), L'appropriazionismo artistico nell'arte visuale: una comparazione tra Italia e Stati Uniti, Trento Law and Technology Research Group, Student Paper Series; 64. Trento: Università degli Studi di Trento. 978-88-8443-942-0

STUDENT PAPER N. 63

Il disegno industriale e la moda tra disciplina dei disegni e modelli e normativa sul diritto d'autore

MARGHERITA RUDIAN (2021), Il disegno industriale e la moda tra disciplina dei disegni e modelli e normativa sul diritto d'autore, Trento Law and Technology Research Group, Student Paper Series; 63. Trento: Università degli Studi di Trento. 978-88-8443-941-3

STUDENT PAPER N. 62

Sistemi decisionali automatizzati e tutela dei diritti: tra carenza di trasparenza ed esigenze di bilanciamento

IRENE TERENGHI (2021), Sistemi decisionali automatizzati e tutela dei diritti: tra carenza di trasparenza ed esigenze di bilanciamento, Trento Law and Technology Research Group, Student Paper Series; 62. Trento: Università degli Studi di Trento. 978-88-8443-940-6

STUDENT PAPER N. 61

Unfair trading practices in the business-to-business food supply chain between public and private regulation

MARIA VITTORIA BORGHETTO (2020), Unfair Trading Practices in the Business-to-Business Food Supply Chain Between Public and Private Regulation, Trento Law and Technology Research Group, Student Paper Series; 61. Trento: Università degli Studi di Trento. 978-88-8443-933-8

STUDENT PAPER N. 60

PFAS e inquinamento delle falde acquifere venete: la tutela civilistica fra danno ambientale e azioni risarcitorie collettive

VERONICA RAISA (2020), PFAS e inquinamento delle falde acquifere venete: la tutela civilistica fra danno ambientale e azioni risarcitorie collettive, Trento Law and Technology Research Group. Student Paper Series; 60. Trento: Università degli Studi di Trento. ISBN: 978 88 8443 927 7

STUDENT PAPER N. 59

Il turismo alla prova del covid-19: una ricerca interdisciplinare: da quali dati partire e quali risposte dare alla più grande crisi che il comparto turistico abbia mai affrontato

UMBERTO IZZO (a cura di), Autori: ANDREATTA, GIULIA; ANDREOLI, ELISA; ARDU, SIMONE; BORTOLOTTO, FABIO; BRUZZO, PIERLUIGI; CALZOLARI, GIULIA; CAMPOS SANTOS, DIEGO; CARLINO, PIETRO; CAVALLERA, LORENZO; CEPPAROTTI, GIACOMO; CIABRELLI, ANTONIA; DALLE PALLE, GIORGIA; DAPRÀ, VALENTINA; DE SANTIS, DIEGO; FAVARO, SILVIA; FAVERO, ELEONORA; FERRARI, LAURA; GATTI, VERONICA; GAZZI, CHRISTIAN; GISMONDO, MARIANNA; GIUDICEANDREA, ANNA; GUIDA, GIOVANNI; INCARNATO, ANDREA; MARANER, ROBERTA; MICHELI, MARTA; ELENA MORARASU, LAURA; CHIARA NARDELLI, MARIA; PALLOTTA, EMANUELE; PANICHI, NICCOLÒ; PELLIZZARI, LAURA; PLAKSII, ANDRII; RANIERO, SAMANTHA; REGNO SIMONCINI, EMANUELE; RUSSO, SARA; SCHIAVONE, SARA; SERAFINO, ANTONIO; SILENZI, LUCA; TIRONZELLI, ELENA; PEGGY TSAFACK, CYNTHIA; VIGLIOTTI, AYLA; ZINETTI, GIULIA, Il turismo alla prova del Covid-19: una ricerca interdisciplinare: da quali dati partire e quali risposte dare alla più grande crisi che il comparto turistico abbia mai affrontato, Trento Law and Technology Research Group, Student Paper Series; 59. Trento: Università degli Studi di Trento. 978-88-8443-903-1

STUDENT PAPER N. 58

La responsabilità dell'internet service provider alla luce della nuova direttiva sul diritto d'autore nel mercato unico digitale

LAURA CAMARELLA (2020), La responsabilità dell'Internet Service Provider alla luce della nuova direttiva sul diritto d'autore nel mercato unico digitale, Student Paper Series; 58. Trento: Università degli Studi di Trento. 978-88-8443-893-5

STUDENT PAPER N. 57

Rischio idrogeologico e responsabilità civile

CATERINA ROBERTI (2020), Rischio idrogeologico e responsabilità civile, Trento Law and Technology Research Group. Student Paper Series; 57. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-891-1

STUDENT PAPER N. 56
Assistente vocale e dati sanitari. Le sfide dell'intelligenza artificiale alla luce del Regolamento (UE) n. 2016/679

LIVIA PETRUCCI (2020), Assistente vocale e dati sanitari. Le sfide dell'intelligenza artificiale alla luce del regolamento (UE) N. 2016/679, Trento Law and Technology Research Group. Student Paper Series; 56. Trento: Università degli Studi di Trento. ISBN: 978 88 8443 888 1

STUDENT PAPER N. 55

The Legal Dimension of Energy Security in EU Law

ANDREAS SCHMIEDHOFER (2020), The legal dimensions of energy security in EU law, Trento Law and Technology Research Group. Student Paper Series; 55. Trento: Università degli Studi di Trento. ISBN: 978 88 8443 888 1

STUDENT PAPER N. 54

Macchine intelligenti che creano ed inventano. Profili e rilievi critici del nuovo rapporto tra intelligenza artificiale e diritti di proprietà intellettuale.

LAURA TREVISANELLO (2020), Trento Law and Technology Research Group. Student Paper Series; 54. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-887-4

STUDENT PAPER N. 53

La protezione delle indicazioni geografiche: il sistema europeo e il sistema cinese a confronto

MARTA COGO (2019), Trento Law and Technology Research Group. Student Paper Series; 53. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-856-0

STUDENT PAPER N. 52

Responsabilità civile e prevenzione dell'abuso interpersonale, fra molestie sessuali e bullismo

FRANCESCA PERETTI (2019), Trento Law and Technology Research Group. Student Paper Series; 52. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-856-0

STUDENT PAPER N. 51

Blockchain, Smart Contract e diritto d'autore nel campo della musica

FRANCESCO FAGLIA (2019), Trento Law and Technology Research Group. Student Paper Series; 51. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-855-3

STUDENT PAPER N. 50

Regole per l'innovazione: responsabilità civile e assicurazione di fronte all'auto a guida (progressivamente) autonoma

FILIPPO ZEMIGNANI (2019), Trento Law and Technology Research Group. Student Paper Series; 50. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-850-8

STUDENT PAPER N. 49

Unravelling the nexus between food systems and climate change: a legal analysis. A Plea for smart agriculture, a "new" organic agriculture and a wiser use of biotechnologies in the name of human rights protection

ALESSANDRA TELCH (2019), Trento Law and Technology Research Group. Student Paper Series; 49. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-842-3

STUDENT PAPER N. 48

Wireless community networks e responsabilità extracontrattuale

CHIARA VIDORNI (2019), Trento Law and Technology Research Group. Student Paper Series; 48. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-841-6

STUDENT PAPER N. 47

Proprietà intellettuale e scienza aperta: il caso studio del Montreal Neurological Institute

GIOVANNA CASSIN (2019), Trento Law and Technology Research Group. Student Paper Series; 47. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-835-5

STUDENT PAPER N. 46

Il "ciclista previdente" che si scontrò due volte: con un'auto e col principio indennitario applicato all'assicurazione infortuni,

CHRISTOPH SIMON THUN HOHENSTEIN WELSPERG (2019), Trento Law and Technology Research Group. Student Paper Series; 46. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-834 8

STUDENT PAPER N. 45

«Errare humanum est». L'errore nel diritto tra intenzionalità, razionalità ed emozioni

LEILA BENSALAH (2018), Trento Law and Technology Research Group. Student Paper Series; 45. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-829-4

STUDENT PAPER N. 44

La gestione del rischio fitosanitario nel diritto agroalimentare europeo ed italiano: il caso Xylella

Marina De Nobili (2018), Trento Law and Technology Research Group. Student Paper Series; 44. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-828-7

STUDENT PAPER N. 43

Mercato agroalimentare e disintermediazione: la dimensione giuridica della filiera corta

Orlandi Riccardo (2018), Trento Law and Technology Research Group. Student Paper Series; 43. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-827-0

STUDENT PAPER N. 42

Causa, meritevolezza degli interessi ed equilibrio contrattuale

PULEJO, CARLO ALBERTO (2018), Trento Law and Technology Research Group. Student Paper Series; 42. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-810-2

STUDENT PAPER N. 41

Graffiti, street art e diritto d'autore: un'analisi comparata

GIORDANI, LORENZA (2018), Trento Law and Technology Research Group. Student Paper Series; 41. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-809-6

STUDENT PAPER N. 40

Volo da diporto o sportivo e responsabilità civile per l'esercizio di attività pericolose

MAESTRINI, MATTIA (2018), Trento Law and Technology Research Group. Student Paper Series; 40. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-784-6

STUDENT PAPER N. 39

"Attorno al cibo". Profili giuridici e sfide tecnologiche dello Smart Packaging in campo alimentare

BORDETTO, MATTEO (2018), Trento Law and Technology Research Group. Student Paper Series; 39. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-795-2

STUDENT PAPER N. 38

Kitesurf e responsabilità civile

RUGGIERO, MARIA (2018), Trento Law and Technology Research Group. Student Paper Series; 38. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-793-8

STUDENT PAPER N. 37

Giudicare e rispondere. La responsabilità civile per l'esercizio della giurisdizione in Italia, Israele e Spagna

MENEGHETTI HISKENS, SARA (2017), Giudicare e rispondere. La responsabilità civile per l'esercizio della giurisdizione in Italia, Israele e Spagna, Trento Law and Technology Research Group. Student Paper Series; 37. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-778-5

STUDENT PAPER N. 36

Il diritto in immersione: regole di sicurezza e responsabilità civile nella subacquea

CAPUZZO, MARTINA (2017), Il diritto in immersione: regole di sicurezza e responsabilità civile nella subacquea, Trento Law and Technology Research Group. Student Paper Series; 36. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-775-4

STUDENT PAPER N. 35

La privacy by design: un'analisi comparata nell'era digitale

BINCOLETTO, GIORGIA (2017), La privacy by design: un'analisi comparata nell'era digitale, Trento Law and Technology Research Group. Student Paper Series; 35. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-733-4

STUDENT PAPER N. 34

La dimensione giuridica del Terroir

BERTINATO, MATTEO (2017), La dimensione giuridica del Terroir, Trento Law and Technology Research Group. Student Paper Series; 34. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-728-0

STUDENT PAPER N. 33

La gravità del fatto nella commisurazione del danno non patrimoniale: un'indagine (anche) nella giurisprudenza di merito

MARISELLI, DAVIDE (2017), La gravità del fatto nella commisurazione del danno non patrimoniale: un'indagine (anche) nella giurisprudenza di merito, Trento Law and Technology Research Group. Student Paper Series; 33. Trento: Università degli Studi di Trento. ISBN: 978-88-8443-727-3

STUDENT PAPER N. 32

«Edible insects». L'Entomofagia nel quadro delle nuove regole europee sui novel foods

TASINI, FEDERICO (2016), «Edible insects». L'Entomofagia nel quadro delle nuove regole europee sui novel foods = «Edible Insects»: Entomophagy in light of the new European Legislation on novel Foods, Trento Law and Technology Research Group. Student Paper Series; 32. Trento: Università degli Studi di Trento. ISBN 978-88-8443-709-9

STUDENT PAPER N. 31

L'insegnamento dello sci: responsabilità civile e assicurazione per danni ad allievi o a terzi

TAUFER FRANCESCO (2016), L'insegnamento dello sci: responsabilità civile e assicurazione per danni ad allievi o a terzi, Trento Law and Technology Research Group. Student Paper Series; 31. Trento: Università degli Studi di Trento. ISBN 978-88-8443-697-9

STUDENT PAPER N. 30

Incrocio tra Contratti e Proprietà Intellettuale nella Innovazione Scientifica e tecnologica: il Modello del Consortium Agreement europeo

MAGGIOLO ANNA (2016), Incrocio tra Contratti e Proprietà Intellettuale nella Innovazione Scientifica e tecnologica: il Modello del Consortium Agreement europeo, Trento Law and Technology Research Group. Student Paper Series; 30. Trento: Università degli Studi di Trento. ISBN 978-88-8443-696-2

STUDENT PAPER N. 29

La neutralità della rete

BIASIN, ELISABETTA (2016) La neutralità della rete, Trento Law and Technology Research Group. Student Paper Series; 29. Trento: Università degli Studi di Trento. ISBN 978-88-8443-693-1

STUDENT PAPER N. 28

Negotiation Bases and Application Perspectives of TTIP with Reference to Food Law

ACERBI, GIOVANNI (2016) Negotiation Bases and Application Perspectives of TTIP with Reference to Food Law. The Trento Law and Technology Research Group. Student Paper Series; 28. Trento: Università degli Studi di Trento. ISBN 978-88-8443-563-7

STUDENT PAPER N. 27

Privacy and Health Data: A Comparative analysis

FOGLIA, CAROLINA (2016) Privacy and Health Data: A Comparative analysis. The Trento Law and Technology Research Group. Student Paper Series; 27. Trento: Università degli Studi di Trento. ISBN 978-88-8443-546-0

STUDENT PAPER N. 26

Big Data: Privacy and Intellectual Property in a Comparative Perspective

SARTORE, FEDERICO (2016) Big Data: Privacy and Intellectual Property in a Comparative Perspective. The Trento Law and Technology Research Group. Student Paper Series; 26. Trento: Università degli Studi di Trento. ISBN 978-88-8443-534-7

STUDENT PAPER N. 25

Leggere (nel)la giurisprudenza: 53 sentenze inedite in tema di responsabilità civile nelle analisi di 53 annotatori in formazione = Reading (in) the caselaw: 53 unpublished judgemnts dealing with civil liability law analyzed with annotations and comments by 53 students during their civil law course

REMO ANDREOLLI, DALILA MACCIONI, ALBERTO MANTOVANI, CHIARA MARCHETTO, MARIASOLE MASCHIO, GIULIA MASSIMO, ALICE MATTEOTTI, MICHELE MAZZETTI, PIERA MIGNEMI, CHIARA MILANESE, GIACOMO MINGARDO, ANNA LAURA MOGETTA, AMEDEO MONTI, SARA MORANDI, BENEDETTA MUNARI, EDOARDO NADALINI, SERENA NANNI, VANIA ODORIZZI, ANTONIA PALOMBELLA, EMANUELE PASTORINO, JULIA PAU, TOMMASO PEDRAZZANI, PATRIZIA PEDRETTI, VERA PERRICONE, BEATRICE PEVARELLO, LARA PIASERE, MARTA PILOTTO, MARCO POLI, ANNA POLITO, CARLO ALBERTO PULEJO, SILVIA RICCAMBONI, ROBERTA RICCHIUTI, LORENZO RICCO, ELEONORA RIGHI, FRANCESCA RIGO, CHIARA ROMANO, ANTONIO ROSSI, ELEONORA ROTOLA, ALESSANDRO RUFFINI, DENISE SACCO, GIULIA SAKEZI, CHIARA SALATI, MATTEO SANTOMAURO, SILVIA SARTORI, ANGELA SETTE, BIANCA STELZER, GIORGIA TRENTINI, SILVIA TROVATO, GIULIA URBANIS, MARIA CRISTINA URBANO, NICOL VECCARO, VERONICA VILLOTTI, GIULIA VISENTINI, LETIZIA ZAVATTI, ELENA ZUCCHI (2016) Leggere (nel)la giurisprudenza: 53 sentenze inedite in tema di responsabilità civile nelle analisi di 53 annotatori in formazione = Reading (in) the caselaw: 53 unpublished judgemnts dealing with civil liability law analyzed with annotations and comments by 53 students during their civil law course. The Trento Law and Technology Research Group. Student Paper Series; 25. Trento: Università degli Studi di Trento. ISBN 978-88-8443-626-9

STUDENT PAPER N. 24

La digitalizzazione del prodotto difettoso: stampa 3D e responsabilità civile= The Digital Defective Product: 3D Product and Civil Liability

CAERAN, MIRCO (2016) La digitalizzazione del prodotto difettoso: stampa 3D e responsabilità civile = The Digital Defective Product: 3D Product and Civil Liability. The Trento Law and Technology Research Group. Student Paper Series; 24. Trento: Università degli Studi di Trento. ISBN 978-88-8443-663-4

STUDENT PAPER N. 23

La gestione della proprietà intellettuale nelle università australiane = Intellectual Property Management in Australian Universities

CHIARUTTINI, MARIA OTTAVIA (2015) La gestione della proprietà intellettuale nelle università australiane = Intellectual Property Management in Australian Universities. The Trento Law and Technology Research Group. Student Paper Series; 23. Trento: Università degli Studi di Trento. ISBN 978-88-8443-626-9

STUDENT PAPER N. 22

Trasferimento tecnologico e realtà locale: vecchie problematiche e nuove prospettive per una collaborazione tra università, industria e territorio = Technology Transfer and Regional Context: Old Problems and New Perspectives for a Sustainable Co-operation among Universitiy, Entrepreneurship and Local Economy

CALGARO, GIOVANNI (2013) Trasferimento tecnologico e realtà locale: vecchie problematiche e nuove prospettive per una collaborazione tra università, industria e territorio. The Trento Law and Technology Research Group. Student Paper Series; 22. Trento: Università degli Studi di Trento. ISBN 978-88-8443-525-5

STUDENT PAPER N. 21

La responsabilità dell'Internet Service Provider per violazione del diritto d'autore: un'analisi comparata = Internet Service Provider liability and copyright infringement: a comparative analysis.

IMPERADORI, ROSSELLA (2014) *La responsabilità dell'Internet Service Provider per violazione del diritto d'autore: un'analisi comparata.* Trento Law and Technology Research Group. Student Paper; 21. Trento: Università degli Studi di Trento. ISBN 978-88-8443-572-9

STUDENT PAPER N. 20

Open innovation e patent: un'analisi comparata = Open innovation and patent: a comparative analysis

PONTI, STEFANIA (2014) *Open innovation e patent: un'analisi comparata.* The Trento Law and Technology Research Group. Student Paper Series; 20. Trento: Università degli Studi di Trento. ISBN 978-88-8443-573-6

STUDENT PAPER N. 19

La responsabilità civile nell'attività sciistica

CAPPA, MARISA (2014) La responsabilità civile nell'attività sciistica = Ski accidents and civil liability. Trento Law and Technology Research Group. Student Paper Series, 19. Trento: Università degli Studi di Trento.

STUDENT PAPER N. 18

Biodiversità agricola e tutela degli agricoltori dall'Hold-Up brevettuale: il caso degli OGM

TEBANO, GIANLUIGI (2014) Biodiversità agricola e tutela degli agricoltori dall'Hold-Up brevettuale: il caso degli OGM = Agricultural Biodiversity and the Protection of Farmers from patent Hold-Up: the case of GMOs. Trento Law and Technology Research Group. Student Paper Series; 18. Trento: Università degli Studi di Trento.

STUDENT PAPER N. 17

Produrre e nutrirsi "bio": analisi comparata del diritto degli alimenti biologici

MAFFEI, STEPHANIE (2013) Produrre e nutrirsi "bio" : analisi comparata del diritto degli alimenti biologici = Producing and Eating "Bio": A Comparative Analisys of the Law of Organic Food. Trento Law and Technology Research Group. Student Paper Series; 17. Trento: Università degli Studi di Trento.

STUDENT PAPER N. 16

La tutela delle indicazioni geografiche nel settore vitivinicolo: un'analisi comparata = The Protection of Geographical Indications in the Wine Sector: A Comparative Analisys

SIMONI, CHIARA (2013) La tutela delle indicazioni geografiche nel settore vitivinicolo: un'analisi comparata. The Trento Law and Technology Research Group. Student Papers Series; 16. Trento: Università degli Studi di Trento. Facoltà di Giurisprudenza.

STUDENT PAPER N. 15

Regole di sicurezza e responsabilità civile nelle attività di mountain biking e downhill montano

SALVADORI, IVAN (2013) Regole di sicurezza e responsabilità civile nelle attività di mountain biking e downhill montano. Trento Law and Technology Research Group. Student Paper; 15. Trento: Università degli Studi di Trento.

STUDENT PAPER N. 14

Plagio, proprietà intellettuale e musica: un'analisi interdisciplinare

VIZZIELLO, VIVIANA (2013) Plagio, proprietà intellettuale e musica: un'analisi interdisciplinare. Trento Law and Technology Research Group. Student Paper; 14. Trento: Università degli Studi di Trento.

STUDENT PAPER N.13

The Intellectual Property and Open Source Approaches to Biological Material

CARVALHO, ALEXANDRA (2013) The Intellectual Property and Open Source Approaches to Biological Material. Trento Law and Technology Research Group. Student Paper Series; 13. Trento: Università degli Studi di Trento.

STUDENT PAPER N.12

Per un'archeologia del diritto alimentare: 54 anni di repertori giurisprudenziali sulla sicurezza e qualità del cibo (1876-1930)

TRESTINI, SILVIA (2012) Per un'archeologia del diritto alimentare: 54 anni di repertori giurisprudenziali sulla sicurezza e qualità del cibo (1876-1930) = For an Archeology of Food Law: 54 Years of Case Law Collections Concerning the Safety and Quality of Food (1876-1930). The Trento Law and Technology Research Group. Student Papers Series, 12.

STUDENT PAPER N.11

Dalle Alpi ai Pirenei: analisi comparata della responsabilità civile per attività turistico-ricreative legate alla montagna nel diritto italiano e spagnolo

PICCIN, CHIARA (2012) Dalle Alpi ai Pirenei: analisi comparata della responsabilità civile per attività turistico-ricreative legate alla montagna nel diritto italiano e spagnolo = From the Alps to the Pyrenees: Comparative Analysis of Civil Liability for Mountain Sport Activities in Italian and Spanish Law. The Trento Law and Technology Research Group. Student Papers Series, 11.

STUDENT PAPER N.10

Copynorms: Norme Sociali e Diritto d'Autore

PERRI, THOMAS (2012) Copynorms: Norme Sociali e Diritto d'Autore = Copynorms: Social Norms and Copyright. Trento Law and Technology Research Group. Students Paper Series, 10

STUDENT PAPER N. 9

L'export vitivinicolo negli Stati Uniti: regole di settore e prassi contrattuali con particolare riferimento al caso del Prosecco

ALESSANDRA ZUCCATO (2012), L'export vitivinicolo negli Stati Uniti: regole di settore e prassi contrattuali con particolare riferimento al caso del Prosecco = Exporting Wines to the United States: Rules and Contractual Practices with Specific Reference to the Case of Prosecco. Trento: Università degli Studi di Trento (Trento Law and Technology Research Group. Students Paper Series 9)

STUDENT PAPER N.8

Equo compenso e diritto d'autore: un'analisi comparata = Fair Compensation and Author's Rights: a Comparative Analysis.

RUGGERO, BROGI (2011) Equo compenso e diritto d'autore: un'analisi comparata = Fair Compensation and Author's Rights: a Comparative Analysis. Trento: Università degli Studi di Trento (TrentoLawand Technology Research Group. Student Papers Series, 8)

STUDENT PAPER N.7

Evoluzione tecnologica e mutamento del concetto di plagio nella musica

TREVISAN, ANDREA (2012) Evoluzione tecnologica e mutamento del concetto di plagio nella musica = Technological evolution and change of the notion of plagiarism in music Trento: Università degli Studi di Trento (Trento Law and Technology Research Group. Students Paper Series 7)

STUDENT PAPER N.6

Il trasferimento tecnologico università-imprese: profili giuridici ed economici

SIRAGNA, SARA (2011) Il trasferimento tecnologico università-imprese: profili giuridici ed economici = University-Enterprises Technological Transfer: Legal and Economic issues Trento: Università degli Studi di Trento (Trento Law and Technology Research Group. Students Paper Series 6)

STUDENT PAPER N.5

Conciliare la responsabilità medica: il modello "generalista" italiano a confronto col modello "specializzato" francese

GUERRINI, SUSANNA (2011) Conciliare la responsabilità medica: il modello "generalista" italiano a confronto col modello "specializzato" francese = Mediation & Medical Liability: The Italian "General Approach" Compared to the Specialized Model Applied in France. Trento: Università degli Studi di Trento (Trento Law and Technology Research Group. Students Paper Series 5)

STUDENT PAPER N.4

"Gun Control" e Responsabilità Civile: una comparazione fra Stati Uniti e Italia

PODETTI, MASSIMILIANO (2011) "Gun Control" e Responsabilità Civile: una comparazione fra Stati Uniti e Italia = Gun Control and Tort Liability: A Comparison between the U.S. and Italy Trento: Università degli Studi di Trento. (Trento Law and Technology Research Group. Students Paper Series 4)

STUDENT PAPER N.3

Smart Foods e Integratori Alimentari: Profili di Regolamentazione e Responsabilità in una comparazione tra Europa e Stati Uniti

TOGNI, ENRICO (2011) Smart Foods e Integratori Alimentari: Profili di Regolamentazione e Responsabilità in una comparazione tra Europa e Stati Uniti = Smart Foods and Dietary Supplements: Regulatory and Civil Liability Issues in a Comparison between Europe and United States Trento: Università degli Studi di Trento - (Trento Law and Technology Research Group. Students Paper Series; 3)

STUDENT PAPER N.2

Il ruolo della responsabilità civile nella famiglia: una comparazione tra Italia e Francia

SARTOR, MARTA (2010) Il ruolo della responsabilità civile nella famiglia: una comparazione tra Italia e Francia = The Role of Tort Law within the Family: A Comparison between Italy and France Trento: Università degli Studi di Trento - (Trento Law and Technology Research Group. Students Paper Series; 2)

STUDENT PAPER N.1

Tecnologie belliche e danno al proprio combattente: il ruolo della responsabilità civile in una comparazione fra il caso statunitense dell'Agent Orange e il caso italiano dell'uranio impoverito

RIZZETTO, FEDERICO (2010) Tecnologie belliche e danno al proprio combattente: il ruolo della responsabilità civile in una comparazione fra il caso statunitense dell'Agent Orange e il caso italiano dell'uranio impoverito = War Technologies and Home Soldiers Injuries: The Role of Tort Law in a Comparison between the American "Agent Orange" and the Italian "Depleted Uranium" Litigations Trento: Università degli Studi di Trento - (Trento Law and Technology Research Group. Students Paper Series; 1)