

# The Exception of Text and Data Mining from the Academic Libraries Standpoint

Maria-Daphne Papadopoulou<sup>1</sup>, Krystallenia Kolotourou<sup>2</sup>, Maria Bottis<sup>1</sup>

<sup>1</sup>Department of Archives, Library and Museum Sciences, Ionian University, Corfu, Greece

<sup>2</sup>Athens, Greece

Email: mariadaphne@yahoo.com, k.kolotourou@gmail.com, botti@otenet.gr

**How to cite this paper:** Papadopoulou, M.-D., Kolotourou, K., & Bottis, M. (2021). The Exception of Text and Data Mining from the Academic Libraries Standpoint. *Open Journal of Social Sciences*, 9, 502-539. <https://doi.org/10.4236/jss.2021.95028>

**Received:** April 21, 2021

**Accepted:** May 22, 2021

**Published:** May 25, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Directive 2019/790/EU (DSM Directive) aims to adapt and supplement the Union copyright framework in order to address the needs and challenges that are constantly arising in the digital environment. Given the importance of text and data mining (TDM) techniques not only with respect to the digital economy but also with regard to their benefits for the research and innovation, DSM Directive introduced the exception of TDM. Academic libraries are among the beneficiaries of the mandatory TDM exception for the purposes of scientific research. This paper analyses in detail the legal infrastructure of TDM exception as provided in Article 3 of the DSM Directive, elucidating at the same time the different notions included therein and addressing certain challenges regarding its application. Further, this paper aims at shedding light on the impact of the TDM exception on libraries and the benefit pursued by them, while in the last part the outcome and results reached through questionnaires and interviews in the framework of the research project “The exception of text and data mining in copyright law regarding Academic Libraries” realized in Greece are described and analyzed thoroughly. In view of the above, concrete recommendations and conclusions are provided with respect to the new TDM exception.

## Keywords

Text and Data Mining, Scientific Research, Copyright, Exceptions, Libraries

---

## 1. Introduction

Directive 2019/790/EU<sup>1</sup> (DSM Directive) consists of a cornerstone for both EU

<sup>1</sup>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.

and national copyright law thus aiming to supplement the relevant regulatory framework in order to be adapted to the needs and address the challenges that are constantly arising in the digital environment. The introduction of mandatory exceptions and limitations especially in the fields of research, innovation, education and cultural heritage preservation, was realized as the one-way street<sup>2</sup> for the purpose of achieving such a level of harmonization within the Union that would further structure a clear legal basis for the smooth functioning of the internal market through the uniform application of EU rules<sup>3</sup>. The repulsion between the diverging rights and interests of copyright and related-rights holders and of users had signified the need to resolve the legal uncertainty<sup>4</sup> by the means of outlining the sphere of operation of the latter through the determination of the extent of the permitted uses as further encompassing new technologies and techniques.

In the light of the above, text and data mining (TDM) was acknowledged as a valuable tool not only for digital economy but also for the promotion and enhancement of scientific research and innovation. Providing for its definition in general terms as a technique enabling the “processing of large amounts of information with a view to gaining new knowledge and discovering new trends possible” or as a tool enabling the “automated computational analysis” of digital information such as text, sounds, images or data, it had been acknowledged that TDM had to be used to its full potential for the benefit of scientific community by the means of its recognition at institutional level being as such covered by the reformed *acquis communautaire*<sup>5</sup>.

The *rationale* behind this initiative is expressly provided by the Union legisla-

<sup>2</sup>Since the optional character of the relevant regime as mostly provided under the 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, was widely perceived and also admitted by the Union legislature as hindering the achievement of the desired uniformity between the copyright systems of the Member States, impinging as such *inter alia* on cross-border uses which were not in any case covered by this regime.

<sup>3</sup>As it had been steadily provided by the Court of Justice of the European Union, EU law needs to be uniformly applied and interpreted within the Union. However, the heterogeneity of national copyright rules resulted in practice to differentiated approaches and mostly to diverging copyright schemes that were not (at least) contributing to the facilitation of cross-border uses and the proper functioning of the internal market. [Sobrino-García \(2020\)](#).

<sup>4</sup>In respect of text and data mining (as the subject-matter of this study), legal uncertainty was (and is still) related to the differentiated forms through which this technique is usually provided by publishers in respect to their content; for example, mining capabilities are either offered by academic publishers as a part of their clearance-of-rights model, or they are only provided for non-commercial purposes or they are actualized through the Copyright Clearance Center (Right Find XML for mining solution) in the case of commercial uses. In general, it seems that mining solutions and their implementation in practice is subject to contractual rules and the inevitable diversification arising therefrom. See also analytically why TDM could not be covered by the existing *Acquis Communautaire* on exceptions and limitations to copyright, and especially on the exception or limitation to copyright for the purpose of scientific research in [Stamatoudi \(2016\)](#) and [Canellopoulou-Bottis, Papadopoulou, Zampakolas, and Ganatsiou \(2019\)](#).

<sup>5</sup>Recital 8 of the DSM Directive. Article 2(2) of the DSM Directive provides as a definition for TDM that “text and data mining” means 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.

ture<sup>6</sup>. Although mere facts or data are not covered by copyright protection, the performing in practice of TDM is profoundly entailing works and other subject matter protected by copyright triggering as such certain powers of the absolute and exclusive right applying at each case at issue. The extent to which these acts could be lawfully carried out by individuals but foremost by research organizations was covered by a shroud of fog. This admitted reality was impinging not only on the achievement of the goal of a high level of copyright protection but also to the advancement of scientific research and of the competitive status of the Union *per se* as a research area<sup>7</sup>. In order to properly address these issues and reverse their said impact, two mandatory exceptions were eventually established under Articles 3 and 4 of the DSM Directive. The first one (that consists of the subject-matter of this study) is devoted to scientific research for the benefit of relevant organizations and cultural heritage institutions within the realm of which libraries and foremost academic libraries hold a dominant position. The second one refers to public and private entities in general and to the TDM utilization “in different areas of life and for various purposes, including for government services, complex business decisions and the development of new applications or technologies”<sup>8</sup>.

It has to be mentioned that this article is composed in the context of a research project titled “The exception of text and data mining in copyright law regarding Academic Libraries” within the framework of the Operational Program “Human Resources Development, Education and Lifelong Learning” of NSRF-Partnership Agreement 2014-2020.

In Part 2 of this Paper the legal infrastructure of Article 3 of the DSM Directive will be presented, while the different notions included therein will be elucidated. In Part 3 certain challenges regarding the application of the TDM exception will be addressed. Light will be shed on the impact that the TDM exception has on libraries and the benefit pursued by them in Part 4, while in the last Part (Part 5) the outcome and results reached under this research project will be analyzed.

## 2. Legal Infrastructure of Article 3 of the DSM Directive

As mentioned above, TDM has been recognized at institutional level as a widespread and rather prevalent *modus operandi* for the generation of information that includes but it is not limited to patterns, trends and correlations<sup>9</sup>. The usefulness and the added value of TDM procedure had been widely perceived in practice as a “widespread access to massive, networked computing power and exponentially increasing digital data sets” especially in terms of research techniques since the value of the data to be used (for scientific research as in the case at issue) does not rely upon their isolated existence but instead upon the extrac-

<sup>6</sup>Under Recitals 8 - 10 of the DSM Directive.

<sup>7</sup>Recital 10 of the DSM Directive.

<sup>8</sup>Recital 18 of the DSM Directive.

<sup>9</sup>In accordance with the definition of the TDM technique as provided under Article 2(2) of the DSM Directive.

tion and subsequent aggregation of their individual value<sup>10</sup>. Since the concepts of “extraction” and “aggregation” were stated above, the TDM procedure *per se* should be at first analysed in order among others to highlight the implication of copyright law in respect of protected content.

Moreover, TDM technique is analysed in the following basic stages: First comes the searching of the material to be used as being further related to the works or the data collected either on an individual basis or on the grounds of a pre-existing database and other sources in general. Within this phase the sources discovered or even parts therefrom are retrieved. Then data is extracted for the purpose of creating a new targeted dataset that could entail among others the reformulation, the modification and the annotation of this content. The third stage consists of the analysis of the dataset produced that includes in technical terms its (even partial) loading in a computer’s working memory—depending on the TDM technique implied at each case at issue—as well as possible extractions from this dataset for the purpose of recombining the material targeted and recognizing specific patterns. Lastly, the results produced are either published or shared implying as such the reproduction of the content included in a given publication or sharing of dataset<sup>11</sup>.

Following this analysis of the stages in which TDM is analysed, its substantive role in scientific research is rather clear as it had been further recognized at least in certain research areas such as in the fields of computational linguistics and biomedical discovery<sup>12</sup>. Focusing on Article 3 of the DSM Directive, the opportunities provided therefrom are strictly aligned to the needs and aims to be pursued under the scope of scientific research as being conducted by research organizations and cultural heritage institutions. In addition, it also covers the members of such institutions which will now be able to proceed to TDM techniques without similarly requiring a prior authorization.

As it had been said above, for the purpose of ensuring the application of the mandatory exception of Article 3, it is expressly provided that it cannot be overridden by contract<sup>13</sup>. In addition, Member States are not allowed to provide for a compensation to rightholders since it is *expressis verbis* provided that “in view of the nature and scope of the exception, which is limited to entities carrying out scientific research, any potential harm created to rightholders through this exception would be minimal”<sup>14</sup>. Furthermore, and as provided both under Union law (under Article 7 of the DSM Directive<sup>15</sup>) with regard to the application of exceptions and limitations in their entirety, the three-step test applies also to this

<sup>10</sup>Rosati (2018).

<sup>11</sup>As these stages had been analysed within the project entitled as “FutureTDM (2017)”.

<sup>12</sup>Rosati (2019).

<sup>13</sup>According to Article 7 par.1 of the DSM Directive “Any contractual provision contrary to the exceptions provided for in Articles 3, 5 and 6 shall be unenforceable”.

<sup>14</sup>Recital 17 of the DSM Directive. See also Latreille (2020).

<sup>15</sup>Article 7(2) of the DSM Directive provides that Article 5(5) of Directive 2001/29/EC shall apply to the exception of text and data mining for scientific purposes.

new exception.

Given the long-term debate that had taken place until the implementation of the DSM Directive, it had been said that the text eventually adopted “could be nothing more than a compromise solution”. In this regard, Article 3 includes several terms and notions that could be characterized not only as general but even as rather ambiguous<sup>16</sup>. Except of the harsh negotiations that had taken place until the eventual adoption of this legal instrument, this is also due to the unprecedented lobbying that intervened and, in any case, the strong pressure exerted from different stakeholders with greatly conflicting rights and interests<sup>17</sup>. In terms of Article 3, there are some significant issues that need to be clarified as they had been already stated and will be complemented in some further aspects in this Part. This clarification is essential if considering the urgent need to provide for effective solutions for the achievement of a “high level of protection” for copyright and related-rights holders as the EU law and case-law have emphatically and repeatedly dictated. Provided that massive and in most cases unauthorized digital uses of copyright protected content had already prevailed in any imaginable field, it could be said that the reformulation of EU law and currently of the copyright systems of Member States had already come rather belatedly to deal with these crucial issues. Although the interpretative role of the CJEU is undoubtedly critical and rather determinative, the clarification of the notions provided under the DSM Directive should not be displaced for the future, as many theorists claim. Instead, clear legal rules should be established in order to implement in practice the objectives pursued under the DSM Directive. As a result, the investigation of the prerequisites set-out by Article 3 of the DSM Directive and the deepening in the concepts provided therefrom (as the subject-matter of this study) are critical for the comprehension of this provision and its proper implementation in practice.

### **2.1. Purpose of Scientific Research and Beneficiaries**

There are certain specified prerequisites that have to be cumulatively fulfilled in order for this exception to be implemented in practice. These requirements are firstly aligned to the character of the purpose to be pursued under this exception, as well as to the identification of the respective beneficiaries.

Firstly, the TDM activities undertaken have to fall within the realm of scientific research. As it is provided by the Union legislature, the term of “scientific research” should be understood in the light of Recital 12 that provides that “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”. As a result, this term shall be construed

<sup>16</sup>See also [Ferri \(2020\)](#).

<sup>17</sup>As [Quintais \(2020\)](#) mentions “what started as a legislative instrument to promote the digital single market turned into an industry policy tool, shaped more by effective lobbying than evidence and expertise. The result is a flawed piece of legislation. Despite some positive aspects, the DSM Directive includes multiple problematic provisions”.

within the meaning of this Directive as covering “both the natural sciences and the human sciences”. Except for this element, what is crucial is that this concept had not been further identified or limited. Consequently, it is conceived as entailing both commercial and non-commercial uses. However, the said inclusion of both commercial and non-commercial uses should not be confused with the character of the beneficiaries of this exception since these are explicitly limited to non-for-profit research organizations and to publicly accessible cultural heritage institutions.

As explicitly mentioned, the beneficiaries of the TDM exception provided in Article 3 are research organisations and cultural heritage institutions. In order legal certainty to be achieved the DSM Directive provides in Article 2 the definitions of both categories of the beneficiaries. According to Article 2(1) “research organization” means a university, including its libraries, a research institute or any other entity, the primary goal of which is to conduct scientific research or to carry out educational activities involving also the conduct of scientific research: 1) on a not-for-profit basis or by reinvesting all the profits in its scientific research; or 2) 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 organization”. According to the definition given to “cultural heritage institution” in Article 2(3) is a publicly accessible library or museum, an archive or a film or audio heritage institution.

In other words, no other person either natural or legal one may invoke the application of this exception even if proceeding to research activities. Nevertheless Article 4 could be invoked in those cases. In order for an entity to be qualified as a beneficiary it shall operate on a non-profit basis or if it does so, it must re-invest the profits accrued in their entirety to scientific research<sup>18</sup>. As provided under Recitals 11 and 12 of the DSM Directive, this exception is also applicable to cases where there is a co-operation between the public and private sector. However, this is mostly aligned to the need of the beneficiaries of this exception to rely upon their private partners for a specific purpose, i.e., for carrying out TDM techniques by using among others their technological tools. According to Recital 11, “While research organisations and cultural heritage institutions should continue to be the beneficiaries of that exception, they should also be able to rely on their private partners for carrying out text and data mining, including by using their technological tools”. Thus, even in cases where the research organisations rely on their private partners for performing TDM, research organisations will remain the beneficiaries of the exception<sup>19</sup>.

It had been stated that since “research does not recognise borders” it should also entail activities serving commercial purposes thus having an equal importance in social terms with the strictly-construed non-commercial research. In this regard, the institutions covered by the TDM exception may also be engaged

<sup>18</sup>Paramithiotis (2020).

<sup>19</sup>Binctin (2019).

in partnerships with the private sector for the purpose of enhancing transfer of knowledge<sup>20</sup> and the overall promotion of innovation. Notwithstanding the fact though that there is not an explicit provision on the type of research to be covered by the exception resulting as such to the perception that commercial activities could also be included in this realm as mentioned above the Union law provides that research organizations and cultural heritage institutions as the beneficiaries of this rule should not have a commercial character. As it is expressly provided under Recital 12, in order for an institution to be covered by the TDM exception it has either to be a non-profit entity or an entity pursuing a public-interest mission. In addition, cultural heritage institutions should be understood as covering publicly accessible libraries, museums, archives and film or audio heritage institutions regardless of the type of work or subject-matter of protection held in their permanent collections. Therefore, it had been respectively stated that “public broadcasting organizations and commercial research organizations” are excluded from the new mandatory exemptions. The crucial difference though is that in the case of non-for-profit and publicly accessible entities as specified above, copyright and related rights holders cannot opt out since it consists of a non-overridable by contract rule. On the contrary, they have such a possibility with entities of a commercial nature in order to be able to protect their “commercial interest” in the framework of Article 4<sup>21</sup>. The *rationale* behind this identification of beneficiaries is provided by the Union legislature that had further stressed that “organisations upon which commercial undertakings have a decisive influence allowing such undertakings to exercise control because of structural situations” should not fall within the scope of this exception since this could result to their “preferential access” to research results<sup>22</sup>.

Lastly, the purpose to be pursued under this exception is aligned with the role of beneficiaries since they bear the burden of proof that the TDM activities undertaken either by themselves and/or by their members are carried out exclusively within the framework of scientific research. Accordingly, researchers and users in general of the content provided by libraries will also have to prove their “scientific purpose” of their research. If this not the case, then the exception is not applicable and subsequently a respective authorization has to be *a priori* obtained.

## 2.2. The Principle of “Lawful Access”

Another condition that seems to consist of the *sine qua non* prerequisite for the application of the exceptions provided under Article 3 of the DSM Directive is that of “lawful access”. It should be at first stated that the Union legislature tried to shed some light on the comprehension and interpretation of this term under Recital 14 of the DSM Directive; “lawful access” should be understood as covering access to content based on an open access policy (which is strictly aligned to

<sup>20</sup>See also Bottis, Papadopoulos, Zampakolas, & Ganatsiou (2019).

<sup>21</sup>Regulating Text and Data Mining in the European Union: Issues and Challenges (2020).

<sup>22</sup>Recital 17 of the DSM Directive.



the conduct of libraries which already implement it in practice) or on the basis of contractual arrangements between rightholders and research organisations or cultural heritage institutions, such as through subscriptions, or through other lawful means such as licensing agreements concluded between the said parties. For instance, in the case where research organisations or cultural heritage institutions have proceeded to subscriptions in order to have access to the content provided thereto, the persons attached and covered by those subscriptions should be deemed to have lawful access. Of course, lawful access should also cover access to content that is freely available online meaning where no subscriptions or other measures (such as technological protection measures (TPMs)) are applicable<sup>23</sup>.

However, the legitimacy of other related uses remains unclear; such uses refer among others to the potential covering of the content of a lawfully accessible public library that falls within the category of its submitted legal deposits. An additional example cited is related to the dissemination of the content produced meaning whether or not it is permitted under the TDM exception to “transfer a lawfully acquired database to a research partner” for mining in another state even if both states are providing for this exception. Moreover, a concern has been raised on the grounds of Recital 9 of the DSM Directive; It is provided therein that TDM may either not involve, otherwise trigger the reproduction right at all or that there could be cases where the mandatory exception provided under Article 5(1) of the Directive 2001/29/EC concerning temporary acts of reproduction<sup>24</sup> should be applicable. Notwithstanding the fact that the exception concerning temporary acts of reproduction is only applicable to TDM techniques that do not involve the making of copies beyond the scope of that exception, it had been argued that entities falling outside the scope of the TDM exception “might still be able to rely on pre-existing law as a fallback argument”<sup>25</sup>. Lastly, it had been argued that in the cases where lawful access would be subject to licensing agreements, the scientific sector would be heavily affected in comparison to commercial TDM thus relying upon other tools<sup>26</sup>.

Nonetheless, it should be stated that irrespective of the criticism paid (to some extent at least) to the introduction of the prerequisite of “lawful access” thus be-

<sup>23</sup>In this regard see also VG Bild-Kunst, C-392/19, where the CJEU in its conclusion found that if the rightholder did authorize the publication of their work explicitly and without reservations or without otherwise resorting to technological measures limiting access/use of their work, then a link to such work would not fall under Article 3 of the InfoSoc Directive; if, instead, the rightholder imposed or set up technological measures restricting access to/use of their work, a link that circumvents such measures would fall within the scope of application of Article 3.

<sup>24</sup>The exception provided under Article 5(5) of the Directive 2001/29/EC as it had been transposed into the Greek law under Article 28B of the Law No. 2121/1993.

<sup>25</sup>Gerrish & Skavlan (2019).

<sup>26</sup>As it had been respectively stated “commercial TDM is often focused in some areas of online analytics (such as retail analytics) (which) are often related to consumer movements and trends gained through the use of cookies, plug-ins or social media”. This part of literature had also argued that the scope of application of the TDM exception provided under Article 3 of the DSM Directive is rather narrow thus excluding “commercially-backed” research organizations such as private universities, profit-making entities and start-ups. *Ibid*, pp. 26-27, 45-63.



ing considered as giving a significant level of power to copyright and related-rights holders (foremost publishers)<sup>27</sup>, this principle works as an essential safeguard for the latter. It shall be recalled that copyright and related-rights holders will lose control over their works and subject-matters of protection without any compensation or reward on the grounds of the equilibrium established under the DSM Directive between their exclusive rights and the enhancement of research and innovation.

### 2.3. Exception from the Right of Reproduction and from the Extraction Right

Article 3 of the DSM Directive provides that “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, TDM of works or other subject matter to which they have lawful access”. Thus, the TDM exception applies to the right of reproduction of authors and holders of related rights, as well as to the *sui generis* database extraction right. In addition, it applies to the new right of publishers of press publications concerning the online uses of their press publications as it had been established under Article 15 of the DSM Directive.

As analysed before, among the different stages of TDM techniques, the right of reproduction and of extraction may be triggered. Specifically for the reproduction right, it is to be noted that under Directive 2001/29 should be understood in a very broad sense, thus covering “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”<sup>28</sup>.

TDM inherently requires at its different stages, apart from the reproduction and the extraction, adaptation or transformation/modification and digitization of the original works and other subject-matters of protection. As a result, apart from the said rights, it is also the adaptation right that could be involved during the mining procedure<sup>29</sup>. Regarding the adaptation right there is no harmonization in the Union level.

Moreover, the wording of the exception of Article 3 does not include the right of communication to the public or the making available right. Following the letter of the Union law, this means that any act that involves the communication to the public does not fall within the scope of the exception and consequently a prior authorization must be required.

Regarding the results of the scientific research, in the majority of cases, the results and outputs in case of TDM projects do not include any parts of the pre-existing works, but TDM techniques by recombining the data offer new in-

<sup>27</sup> *Ibid*, p. 26-27.

<sup>28</sup> Article 2 Directive 2001/29/EC.

<sup>29</sup> See also Geiger, Frosio, & Bulayenko (2019).

sights and patterns<sup>30</sup>. Although consisting of a third (to EU) state and had expressed its intent of not implementing the DSM Directive<sup>31</sup>, it is noteworthy to cite the example of UK law under which it is provided that such results can be freely communicated insofar as they do not entail copyright-protected material beyond the extent of the content acceptable under the pre-existing quotation exception.

An interesting issue is whether the datasets produced in order to be mined are covered by the TDM exception or not. In this regard, Geiger et al. proposed that “in light of the increasing research focus on the quality and verifiability, a TDM exception should enable (not only) storing” but also “communication of research files created for TDM”<sup>32</sup>. Examining the implementation choices made in other Member States, it is noteworthy that the TDM exception provided under the German law allows for such communication to a limited though cycle of recipients and explicitly for two purposes that had been clearly specified. More precisely, it is stated that “the exception covers the acts of reproduction necessary for undertaking TDM and the making available of a “corpus” (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 quality assurance”<sup>33</sup>. As Geiger et al. mentioned regarding this German exception “once the TDM project is completed, the “corpus” can be sent to institutions designated by law for long term storage. Any other copy made should be deleted”<sup>34</sup>. Moreover, in the process of verifiability of the results, the researchers may be required to communicate the datasets to their peers and thus, the right of communication to the public may be implicated.

As many librarians had strongly supported, the re-use and dissemination among researchers and even users in general of the datasets produced by the means of TDM techniques would be of crucial importance. However, there is not a specific provision regarding the re-use of these data by other parties, i.e., researchers or librarians, in the DSM Directive.

## 2.4. The Notion of “Storage”

The production of sufficient datasets requires both significant terms of time and strong efforts by the researchers and the persons in general involved in these acts. Accordingly, it was recognized that the preservation of these datasets does not only consist of an important step for the verification of the quality of the findings produced but also of an inherent prerequisite of the relevant procedures<sup>35</sup>. Finally, even if it was not provided in the initial text of the Proposal of

<sup>30</sup>Triaille, de Meeùs d’Argenteuil, & de Francquen (2014).

<sup>31</sup>Rosati (2020).

<sup>32</sup>Geiger, Frosio, & Bulayenko (2018).

<sup>33</sup>*Ibid*, p. 18.

<sup>34</sup>*Ibid*.

<sup>35</sup>The European Parliament Must Improve the Text and Data Mining (TDM) (2018) exception to benefit European research and innovation, Hugenholtz (2019). “*This is important because empirical scientific research generally requires research data to remain available for corroboration purposes*”.

Directive<sup>36</sup>, in the final text of the DSM Directive, there is a relevant specific provision. According to Article 3(2) of the DSM Directive, “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”.

It could be said that this storage and retention capability had been dictated by the inherent need of the scientific and research community to rely upon the ongoing availability of the results produced from preceding analyses that should be accordingly maintained in safe networks for the essential purpose of verifying, supporting or even questioning prior outcomes and reach such new conclusions.

However, such a storage requires an “appropriate level of security”, while this scope of this retention is also limited to the purposes of scientific research including for the verification of the relevant given results. However, the person liable, otherwise the person bearing the responsibility for such a storage and retention is not clarified in the text of the Directive. It could be said that notwithstanding the wording of Article 3(2) under which such a storage and retention possibility is explicitly offered to the beneficiaries of this exception, at the same time it seems that this right is rather limited “in certain cases”. More precisely, Recital 15 provides that “research organisations and cultural heritage institutions could ‘in certain cases’, for example for subsequent verification of scientific research results, need to retain copies made under the exception for the purposes of carrying out TDM. In such cases, the copies should be stored in a secure environment. Member States should be free to decide, at national level and after discussions with relevant stakeholders, on further specific arrangements for retaining the copies, including the ability to appoint trusted bodies for the purpose of storing such copies”. Consequently, it is up to the national legislator-provided though that relevant discussions had previously taken place with the stakeholders involved to determine the specific arrangements or the rules in general applicable to the copies’ storage and retention that should profoundly include the relevant timeframe, the medium and even the potential entrustment of such activities to a third body. In any case, the text of the Directive does not provide for a relevant provision or a complementary guideline meaning that the question concerning the time-period of the storage and retention of the derived data, the specific way and conditions of such uses remains unanswered<sup>37</sup> and may probably lead to derogations among the national laws of the Member States.

In respect of this entrustment to other bodies for the storage and retention of copies and more precisely against those “trusted intermediaries<sup>38</sup>”, UK LACA had stressed that “universities and libraries are trusted to spend billions on subscriptions and acquisitions each year and trusted to preserve in-copyright material”. Consequently, it had been outlined that they should also be trusted to

<sup>36</sup>Geiger, Frosio, & Bulayenko (2018); Hilty & Richter (2017).

<sup>37</sup>See in relation to issues of storage, Bensamoun & Bouquerel (2020).

<sup>38</sup>LACA (2019). The Right to Read is the Right To Mine: But Not When Blocked by Technical Protection Measures.

maintain the data deriving from TDM activities urging as such researchers and librarians to resist to any potential legislative initiatives under which their new right as provided by the exception would be restricted by requiring or entrusting third parties to hold the said data. At this point, it should also be reminded that according to the German TDM exception the “corpus” of such data can be sent to the institutions designated by law for long term storage<sup>39</sup>. As a result, an additional issue that remains unclear is the identification of the person who is going to bear the cost for the storage and retention of the relevant datasets.

Furthermore, this goal implies also extra costs as indicatively referring to the operation, development and maintenance of the relevant and rather necessary equipment. As aligned to the technical issues mentioned above, there are further matters that need to be clarified as related to the determination of the person(s) responsible for the backups of the system that will be undoubtedly required, as well as and/or for the determination of the multitude of backups and/or other relevant acts. In addition, it should be noted that the Directive does not provide for a timeframe in maximum terms within which the storage of the derived data will be allowed.

Thus, the question that additionally arises is who will define how long this data will be stored and retained and in what specific way and under which conditions; all these issues though as related to the fulfillment of the condition of the “appropriate level of security”, remain unanswered.

### **3. Challenges Regarding the Application of the TDM Exception**

Beyond the number of notions included in the letter and the spirit of the Union legislature that need to be clarified as stated above, the application in practice of the TDM exception may be hindered by several additional challenges. These challenges are related to the provision of access and other features of data, to the co-operation between the stakeholders involved and to the best practices that have to be determined at national level.

#### **3.1. Measures to Ensure Security and Integrity of Networks and Databases**

According to the provision of Article 3(3) of the DSM Directive, “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”. A potentially high number of access requests to, and downloads of works or other subject-matters of protection could imply a risk that the security and integrity of the systems or databases of rightholders could be jeopardized<sup>40</sup>. For this reason authors and holders of related-rights are enabled to apply protection measures. Those measures had been further indicatively stated under Recital 16

<sup>39</sup>Geiger, Frosio, & Bulayenko (2018); Jondet (2018).

<sup>40</sup>Recital 16 of the DSM Directive.

of the DSM Directive as those ones that are being used to ensure that only persons having lawful access to their data can have access to it; such an identification could take place through IP-address validation or user authentication.

This rule had been criticized as an “unclear provision whose interpretation is likely to be difficult”, meaning that it could impinge on the application of the TDM exception<sup>41</sup>. This provision is under the test of the principle of proportionality and that of “necessary extent”; more precisely, as the text of Directive clarifies, “those measures should remain proportionate to the risks involved and should not exceed what is necessary to pursue the objective of ensuring the security and integrity of the system and should not undermine the effective application of the exception”<sup>42</sup>.

### 3.2. Technological Protection Measures (TPMs)

TPMs are often used by rightholders as a means to protect their works or other subject-matters of protection against unlawful digital uses. The Directive 2001/29/EC had explicitly provided under Article 6 for a protection scheme against the circumvention of any effective technological measures. This scheme is also applicable to Article 3 of the DSM Directive as it is explicitly provided under Article 7(2) of the latter.

More precisely, the first, third and fifth subparagraphs of Article 6(4) of the Directive 2001/29 shall apply to the TDM exception. The relevant subparagraphs of the Article 6(4) provide as follows:

“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.

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.

When this Article is applied in the context of Directives 92/100/EEC and 96/9/EC, this paragraph shall apply *mutatis mutandis*”.

According to Article 6(1) of the Directive 2001/29, “Member States shall provide adequate legal protection against the circumvention of any effective technological measures, which the person concerned carries out in the knowledge, or with reasonable grounds to know, that he or she is pursuing that objective”.

As stated under Recital 7 of the DSM Directive, the protection of technologi-

<sup>41</sup>Míšek (2019).

<sup>42</sup>Recital 16 of the DSM Directive.

cal measures established in Directive 2001/29/EC remains essential to ensure the protection and the effective exercise of the rights granted to authors and to other rightholders under Union law. In this regard, rightholders should have the opportunity to ensure that they will remain free through voluntary measures to choose the appropriate means of enabling the beneficiaries of the exceptions and limitations provided for in this Directive to benefit from them. In the absence of voluntary measures, Member States should take appropriate measures in accordance with the first subparagraph of Article 6(4) of Directive 2001/29/EC, including where works and other subject matter are made available to the public through on-demand services. However, it should be simultaneously ensured that the use of technological measures will not prevent the enjoyment of the exceptions and limitations provided for in this Directive.

As it is clearly shown from the above, the issues arising and the concerns that had been respectively raised have to be resolved or there is at least an opportunity to be resolved by national legislators. For this purpose, a number of proposals had been made; for instance, the Association of European Research Libraries (Ligue des Bibliothèques Européennes de Recherche—LIBER) had suggested that the issues related to the potential restriction of researchers to access the content needed for TDM have be resolved within a maximum period of 72 hours. An additional relevant suggestion was to provide for the (same) deadline of maximum 72 hours including though financial penalties in case of non-compliance and to promote simultaneously actions (including legal actions) if access is blocked and not quickly resolved by publishers<sup>43</sup>.

### 3.3. Establishment and Enhancement of Co-Operation between the Stakeholders Involved

Apart from the issue of security, an additional feature traversing the DSM Directive in its entirety is the highlighting and encouragement of the principle of co-operation between the various stakeholders involved in each case at issue. Except for the profound co-operation needed between beneficiaries and copyright and related-rights holders and the building of trust and strong bonds between the so-called “providers” and “users” of creative content, further co-operation schemes must be implemented. Focusing on the TDM exception, it is clear that besides the issue of data *per se* as mostly aligned to its provision, extent and overall availability, the effectiveness of the relevant procedures demand in parallel a high-level of digital literacy from the beneficiaries’ behalf. As a result, it could be said—as a general comment aligned to the results reached from the interviews concluded with Greek stakeholders as they will be analyzed—that the success of the new regulatory framework cannot exclusively rely upon one category of stakeholders (i.e., libraries and their personnel) but instead it shall be administered through an extensive and in-depth collaboration with experts from

<sup>43</sup>As suggested in the Guidelines issued by 4 associations representing libraries, i.e., EBLIDA, IFLA, LIBER and SPARC Europe under the title “Transposing the Directive on Copyright in the Digital Single Market: A Guide for Libraries and Library Associations” (Proudman et al., 2019).

IT sector in order to have the desirable successful results of TDM techniques. In addition, librarians had stressed that in order to be able to perform successfully their role and to support effectively researchers and even negotiate with publishers, they should not be merely informed about the new regulatory framework but instead they should be similarly trained in this regard in order to achieve an in-depth understanding of the rules that they will be called to implement in practice. Consequently, the goal is to equip librarians and in general the staff of libraries as the beneficiaries of the exception with both the legal and technical expertise by the establishment and maintenance of strong bonds between the IT and the legal sector.

### **3.4. Best Practices Concerning the Application of the TDM Exception**

Lastly, an additional obligation to Member States is provided by the Union legislator under Article 3(4) thus dictating that they should encourage rightholders, research organisations and cultural heritage institutions to co-operate with each other and define “commonly agreed best practices” with regard to the application of the storage obligation under Article 3(2) and of the measures to safeguard the integrity and the security of such storage medium under Article 3(3). As stated above the Union law had repeatedly given great attention and emphasis to this principle of co-operation (as related to the implementation of certain provisions of the DSM Directive) with a view to leave some room to the stakeholders directly involved in this new legal framework to discuss, exchange their views and end up together to the self-considered as appropriate specific tools and measures for the effective application of the re-invented copyright regime.

These practices are particularly aligned to data management as being further analysed to data storage and documentation, to the sharing of the data used and produced, as well as to ethical issues such as the management of the data including sensitive information and the principle of “responsible conduct of research”<sup>44</sup>. It seems that notwithstanding the fact that the significance of such tools is widely recognized in particular from the librarians’ point of view, there are not yet respective initiatives at least in Greece. In any case, though this issue is strictly aligned with the desired and required legal certainty on the extent of the uses permitted under the TDM exception; consequently, the majority of librarians had underlined the need and the importance for the establishment of clear and robust national legal rules.

## **4. The Impact of the TDM Technique on Libraries and the Benefit Pursued**

The objective pursued under the establishment of the mandatory TDM exception on the promotion and enhancement of scientific research as aligned to the generation of information that only indicatively involves new patterns, norms, trends, correlations and networks, is more than profound.

---

<sup>44</sup>Boston College Libraries (2021).



At the same time though the new provisions are expected to play a pivotal role—among others—in the overall operation of libraries which are explicitly included in both categories of the beneficiaries of this exception as they are further specified into academic libraries in the case of research organizations and within the second category of cultural heritage institutions into public, national and publicly accessible libraries<sup>45</sup>.

Indeed, TDM techniques had been recognized as undoubtedly influencing and contributing not only to the building of a solid library system and of a strong collection in different reference fields but also to their overall growth as related to any aspect of their operation. These fields are indicatively including the content selection and acquisition policies, the identification of circulation thrust areas, the amendment of the statutory rules governing the financial and the human resources sector as being further related to the determination and improvement of the staff's "behavioral pattern" towards readers and users in general, the re-invention of the libraries' marketing policies and provision of information and other services, and in general the introduction of new technologies and techniques on the grounds of the usability and demand surveys' results<sup>46</sup>.

Consequently, the impact of the new possibility afforded to libraries is profound since they will be able to leverage and build upon the limitless applications of TDM in the field also of the external relationships that they develop at national, Union and even at international level; libraries should adopt a new holistic model for their development and redesign their way forward on the grounds of their strong and multidimensional collections. In this regard, they will be also able to analyze the habits and preferences of readers and to re-address accordingly their policy for the purpose of attracting new audience. Indeed, it had been found that the whole search system of a library—especially of a digital library—can be redesigned on the grounds of the "valuable information extracted" with regard to the manner through which such searches are performed by users as accompanied with the difficulties that may potentially arise. TDM contributes to the formulation of digital libraries thus optimizing automatic information processing, improving the quality of the information disseminated, strengthening the relevant collections, while simultaneously impacting on purely managing issues since reducing the costs implied. As a result, the use of such an experience provides libraries with the ability to predict future obstacles and to inverse their negative impact by addressing anew users' information needs, while providing them with better user-oriented applications<sup>47</sup>.

Moreover, the TDM exception is expected to enhance the role of libraries as being further inexorably intertwined with the Open-Science movement<sup>48</sup> that runs in connection with their repositories. It should be noted that this policy entails in addition the open access, open data and fair data policies which had

<sup>45</sup>Preamble of the Directive, Recitals 12 and 13.

<sup>46</sup>*Ibid*, p. 39-45.

<sup>47</sup>Kovacevic, Devedzic, & Pocaajt (2010).

<sup>48</sup><https://ec.europa.eu/digital-single-market/en/open-science>.

been defined as a “loyalty friend” of mining beyond its own contribution to the maximization of the researchers’ ability in proceeding to automated text and data analysis<sup>49</sup>.

Focusing on academic libraries. Their role on the basis of Research Data Management (RDM) is rather crucial and needs to be in-depth comprehended under the new regime as related in particular to the multiple tasks that their staff could successfully undertake if leveraging upon the new possibilities. Accordingly, one of their role is their support (as also implied by the wording of the Directive) in respect of data collection, access to data and conduct of searches, to data curation (being defined as a form of managing the data deriving from TDM procedures), data carpentry, data integrity, data analysis and visualization, while also entailing their “embedded roles in research project teams”, as well as promotional and raising-awareness activities including the users’ training on the appropriate use of data. In this regard, an additional aspect of the new library role refers also to the calculation of metrics (bibliometrics and altmetrics) for the purpose to assist researchers and users in understanding their own work and implementing in practice the principle of “responsible use”, while this technique could also contribute to the management and decision-making policy of libraries themselves<sup>50</sup>.

Indeed, all these specified aspects of the data’s overall management and control are falling within the re-invented competence of libraries which had been now empowered with the right to undertake mining techniques without the requirement of the author’s/rightsholder’s previous authorization (and payment of a fee) since they are explicitly covered by the new exception. They can now remove or ignore any contractual clauses deriving from agreements or licenses which are now contrary to the new regime, while they are also competent to negotiate with publishers and in general holders of copyright and/or related rights within the context of adopting mutually accepted best practices. In addition, libraries may undertake actions (including legal proceedings) for the purpose of lifting any access-blocking requirements which are not promptly resolved<sup>51</sup>. These actions may further refer to the protection of personal data since TDM activities may involve such a processing “for archiving purposes in the public interest, or processing for scientific or historical research or statistical purposes<sup>52</sup>” and of the privacy of researchers within the context of protection of academic freedom meaning that they are enabled to react against potential requests from publishers and rightholders in general as related to the provision of further in-

<sup>49</sup>White (2020).

<sup>50</sup>Cox (2018).

<sup>51</sup>As it had been respectively stated, the role of libraries in supporting the research community may entail the undertaking of legal actions against infringers, as well as the dealing of unusual behaviours and/or access-blocking requests on the grounds of the “Digital Rights Management”. Stewart, Secker, Morrison, and Horton (2016).

<sup>52</sup>See more at: Vavouis, Papadopoulos, Gerolimos, & Xenakis (2020).

<sup>53</sup>As it had been respectively stated by representatives of the Copyright Working Group of REBIUN (2020) (i.e., the network of university libraries in Spain).

formation in relation to mining<sup>53</sup>.

Within this context, some light should also be shed on the supportive role of libraries to the research activities conducted by their users by the means of TDM techniques. As it had been respectively stated, the research community may address a number of issues that would in practice impinge on their role and competence to provide for new results as being indicatively related to download limits or to other limitations provided by publishers such as “unusual behaviour” reports and to the imposition of Digital Rights Management (DRM) technologies that do not allow the conduct (at least without a prior permission) of TDM procedures. In the light of the above, libraries may now provide and promote TDM services as a service offered by themselves encouraging simultaneously the initiation of relevant projects through the development and enhancement of partnerships. The role of librarians as facilitators and their subsequent contribution to the exploration and review of new research trends and to the comparison of the relationships arising as a result of those searches are expected to highlight the intertemporal and traditional role of academic libraries as a “source for peer reviewed research and other scholarly literature”<sup>54</sup>. In addition, they shall provide for a clear and comprehensive framework under which resources are made available to users further entailing the extent of the permitted intervention, and they should also be firm with data with an aim to safeguard the rights of the research community. Concluding, it had been stressed that libraries shall strictly follow and ensure the implementation of the data protection legislation<sup>55</sup> as it is further aligned to the support of researchers in respect of their privacy rights as mentioned above.

Moreover, the rules governing TDM are widely considered as contributing to the establishment of a strong library system since allowing the development of an analysis mechanism concerning data from differentiated sources and perspectives. Within this context, it is assumed that the forthcoming implementation of mining techniques will influence or even re-invent the policy decisions concerning the libraries’ holistic development approach thus applying to a variety of their operation fields. TDM can be used by libraries in order to find useful but undiscovered or unknown or hidden patterns and previously undetected relationships on the basis of a large collected data that could inter alia contribute to the libraries’ own policy-making. As a result, the new regime is expected to form a solid ground for the implementation of new activities and competences that will contribute to the libraries’ actual rebuilding. These decisions may refer among others to the selection or acquisition of content, to the building of a strong and multidimensional collection, to the realization of the pros and cons deriving from the actual operation of a library in order to draw the future strategic plan, as well as to the means through which networking will be enhanced<sup>56</sup>.

However, the formulation of this exception into an effective legal rule at national level does not suffice *per se* for the achievement of the objectives men-

<sup>54</sup>Rattan (2019).

<sup>55</sup>Stewart, Secker, Morrison, & Horton (2016).

<sup>56</sup>*Supra note* 53, p. 44.

tioned above; it has to be implemented in practice by constructing a new dynamic for libraries as being further analysed into operational terms since they will be called upon to regulate a number of crucial but still complicated technical and legal issues on the basis of the new copyright system. For this purpose, and in order to be able to successfully perform all the aforementioned activities and fulfil the objective pursued, a number of key categories of operation had been proposed. First of all, it had been stated that libraries should record their content, resources and relevant subscriptions and share the knowledge accrued. In addition, the implication of the leadership of libraries had been proposed in order to structure a support-data driven research system. This element is further aligned to the development of new institutional policies, including research data management policies, as well as external-relationship policies, i.e., with publishers and overall networking building. As the stepping-stone of this proposal comes the understanding of the current situation<sup>57</sup>. Accordingly, Section V5 as follows is exactly dedicated to the analysis of the relevant environment in Greece.

### **5. Mapping Greek Reality: The Conclusions Arising from Interviews with Librarians, IT and Legal Experts**

The Greek copyright law (Law No. 2121/1993)<sup>58</sup> does not provide for a TDM exception, as was the case in other Member States (UK, Germany, France, Estonia). The Greek copyright legislation does not include either the optional exception or limitation for scientific research<sup>59</sup> provided in the Directive 2001/29/EC.

At this point, it should be mentioned though that the Greek legislator had already embraced TDM activities by the means of the Law No. 4452/2017. However, national law had empowered (exclusively) the National Library of Greece (NLG) with such a competence without though regulating simultaneously the relevant regime foremost in terms of copyright law. More precisely, it is provided under Article 4(4)(b) of Law No. 4452/2017 that the NLG shall operate as the official National Depository and Archive of digital publications, data and metadata which are produced in Greece or which are related to the Greek culture. As it had been further analyzed, the monitoring and archiving of the Internet (web archiving) or of any other technology environment shall fall within the scope of such a competence. For this purpose, the NLG shall allocate and coordinate the relevant actions at national level. Following the establishment of this provision, TDM techniques had been used by the NLG in order to implement web harvesting and archiving in Greece as it had been further analyzed to various stages. The last one included the developing of a “National Archiving System of Greek Web” in order to operate as the national “user/librarian interface” meaning that users could have access through special tools to the “archive

<sup>57</sup>White (2020).

<sup>58</sup>Official Government Gazette (FEK) A’/25/04.03.1993); <https://www.opi.gr/en/library/law-2121-1993>.

<sup>59</sup>Article 5(3)(a) of the Directive 2001/29/EC.

<sup>60</sup>Kanellopoulou-Botti, Papadopoulos, Zampakolas, & Ganatsiou (2019).

formed from the Greek web harvesting process and TDM procedure”<sup>60</sup>.

In view of the above and taking under consideration the forthcoming revision and reform of the relevant framework in the light of the implementation of the DSM Directive<sup>61</sup>, the mapping of the relevant area, was considered rather necessary.

As already mentioned, this article has been composed in the context of a research project titled “The exception of text and data mining in copyright law regarding Academic Libraries” within the framework of the Operational Program “Human Resources Development, Education and Lifelong Learning” of NSRF-Partnership Agreement 2014-2020. Aim of the project was to explore the interrelation between the TDM exception and academic libraries with special focus on the Greek environment.

In the framework of this project concrete methods have been chosen for production of data; 1) the creation and the circulation of a questionnaire to academic libraries and 2) the conclusion of a number of interviews with experts for the needs of this research project a questionnaire has been circulated to librarians and a number of interviews have been concluded concerning both the issue of TDM in the libraries under the aim of mapping the relevant national status, while also intending to explore the views of national stakeholders. Although this questionnaire consisting of 13 questions, some of more general nature while some other of more technical character—was sent to 155 libraries, only 19 provided for a feedback. Regarding the interviews law professors, researchers, librarians and IT experts had been approached to touch base on TDM issues. Attention was paid in order to contact experts that either hold a position of responsibility or they are aware of TDM. In total 12 interviews had been concluded; five with legal experts, two with IT specialists and five with librarians. The outcome and the results reached under this initiative are very interesting.

From the analysis of the questionnaires the following conclusions are deducted:

1) A percentage of 22% of the representatives of academic libraries are not aware of TDM techniques. A percentage of 33% of them was not aware of the TDM exception provided under the DSM Directive. Further, half of the representatives replied that they are aware of the means through which TDM can contribute to the services provided by academic libraries, while 61% was not able to respond on the specific ways through which academic libraries could use effectively TDM.

With regard to the above-mentioned, the technical nature of TDM techniques should be highlighted. Although TDM techniques are used for many years, they are still considered as a nascent tool<sup>62</sup>. This specific and technical character of

<sup>61</sup>As well as of the Directive 2019/789/EU of the European Parliament and of the Council of 17 April 2019 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, and amending Council Directive 93/83/EC.

<sup>62</sup>European Commission (2016).

TDM could justify the level of knowledge of the personnel of libraries. In addition, Greek law does not provide for a similar provision—such as the legal systems of other Member States—having as result that TDM techniques were not widely used in the framework of academic libraries.

2) 61% of representatives estimate that the personnel of academic libraries are not sufficient for the application of the TDM exception, while the majority of the representatives and specifically 72% of them claim that even the existing personnel does not have the necessary skills to support TDM techniques. Further, the totality of the representatives of academic libraries admitted that they consider the elaboration of best practices in relation with TDM as absolutely necessary and rather imperative.

These conclusions must be read in conjunction with the fact that the representatives of academic libraries highlighted the issue of understaffing in academic libraries which is of critical importance, while they also expressed their fear that it will be even more stressed in light of the new competences to be afforded to the latter.

3) In addition, it has to be mentioned that the vast majority and specifically 85% of the persons interrogated focus on both theoretical and practical training via educative and informative seminars, webinars, workshops, team collaborations, as well as via manuals, brochures, guidelines and policies.

4) According to the feedback received, the TDM exception can contribute to the operation of the libraries and notably to three specific domains: a) to the optimisation of the services provided, b) to the enrichment of the collection of libraries and c) to the anticipation of users' needs.

5) Moreover, it is deducted that the role and overall contribution of academic libraries to the effective application and implementation of the TDM exception consists of the following activities:

- Communication with publishers.
- Continuous training of the personnel of academic libraries.
- Obligation of submission of works which are created within the operation of Universities and in particular within the institutional repositories, and providing access to works.
- Educational and informational work from the side of academic libraries addressed to institutional repositories and members of the academic community.
- Drafting and elaborating on the policy and overall management of institutions.

6) In respect of the licensing agreements concluded between academic libraries and publishers, half of respondents mentioned that they are not aware of the content of such contracts. From the rest of replies, the following results are deducted:

- It is deemed as necessary that the negotiations to be conducted in respect of the conclusion of licensing agreements shall be based on collective represen-

tation schemes.

- There is also a need to review the agreements already in force in the light of the new TDM exception and the national law to be respectively implemented with the aim of building in parallel and establishing a climate of trust between publishers and academic libraries.

Apart from the conclusions reached from the feedback received from the questionnaires, a number of concrete conclusions were also reached following the analysis of the personal interviews undertaken with national stakeholders. The intended purpose was to record their positions and views in both technical, legal and actual parameters, while also to shed light to the current environment and its desired transformation.

The targeted questions developed and addressed to the respective recipients were classified into three categories thus serving accordingly three main purposes: 1) to understand the technical issues, including interoperability, to which TDM is analyzed from IT experts, 2) to aggregate and present the views of librarians in respect of the role that the TDM technology as well as libraries themselves can play in the achievement of the objectives pursued and the maximization of the benefit for the later, and 3) to highlight the legal aspects and the challenges that the legislator—and especially the Greek one—has to deal with the implementation of Union law into the national copyright regime in accordance to the view of law experts.

Prior to the presentation and analysis of the specific results reached, a general but still crucial observation shall be made; the initial targeting of the recipients of the said questions and accordingly the sample of respondents were rather wide. However, what was observed was an inability or a reluctance to provide for specific answers to the questions posed indicating as such a lack of information or at least of sufficient information in order for them to be further able to provide and share a safe position on the issues addressed. Consequently, what should be at the outset stressed is the imminent need for stakeholders both at individual and institutional level to be in-depth informed about the upcoming reality indicatively by the means of specified seminars and other educational and training activities since not merely the information about these issues but the specialization on the TDM technique and the opportunities provided therefrom consist of the *sine qua non* prerequisite for the effective implementation and success of the new regulatory framework.

### 5.1. Technical Issues: Interviews Results

The interviews conducted with experts from IT sector focused firstly on the TDM technology *per se*. As was highlighted, these techniques vary on the grounds of their subject-matter meaning whether they concern data or texts. In the case of Data Mining, statistical techniques and recognition of patterns are applied under the purpose of identifying the “secret casual relationships” between various data consisting as such of a useful tool for the analysis of the phe-



nomenon at issue. On the contrary, in the case of Text Mining, the analysis of a set of texts is conducted by the means of general mining in order to extract the information related to such data. Subsequently, such information is grouped in terms of similarity/relevance and it is further analyzed by the means of the Natural Language Processing (NLP) method<sup>63</sup>.

What is noteworthy is that data mining as a sub-area of information technology had been used and was widely known even from 1970's and/or 1980's, while its rapid spread through the passage of time is rather self-evident since it had been aligned with the overflow of new applications, as well as with the qualitative improvement of algorithms and other relevant methods both in academia and in a wider context.

Emphasis was also paid to the stages of the TDM techniques. As the relevant stages have been analysed before, IT experts as well identified three stages. The first step includes the "pre-processing" of data, which implies that data should be at a "clean version", given the deficiencies that may arise or their undistorted rates by virtue of their substance in predetermined formats. The second stage consists of the data mining in algorithmic terms meaning the algorithm itself and the sequence of the steps that must be taken for the purpose of the data's comparison, its further classification, or the formulation of a subjective outcome. Within the third stage, the visualization of this outcome takes place which may further take various forms such as that of a table which consists of the most common one.

Furthermore, the technical issues arising are primarily devoted to the safeguarding of the mining procedure itself and of the safe storing of data in different but always transparent operating systems.

Within this context, Union legislator underlined—and subsequently the national legislator should underline—the data storage and maintenance made by the user who proceeded to a mining technique into a safe environment. As a result, it had been stressed that it is absolutely necessary to determine the specific procedure that has to be applied in order to ensure compliance with the aforementioned requirement. Experts had respectively raised the need to formulate an accurate procedure and a safe storage of the results achieved which should be further analyzed into specific instructions to be given to the person responsible for the conducting and overall operation of TDM techniques. More precisely, storage should take place into a restrictively accessible area, otherwise into access points that will require a special authorization, while back-ups should also be made into a number of units that should be dedicated to this purpose and kept by the network and security manager of a given beneficiary institution. This sense of security as connected to the protection against accidental destruction forms its first dimension, while the second one relates to the protection of data against malicious attacks by various means such as the so-called "firewalls", the regular updates of the systems' applications and programs and the implication of

---

<sup>63</sup>Chopra, Prashar, & Sain (2013).

all safety protocols indicatively stating the secure navigation protocol <https>. As experts had respectively signified (consisting in parallel of a significant outcome of this activity), these two aspects and notions of security form simultaneously the cumulatively applying preconditions for its own efficiency and success. As a result, administrator(s) of the relevant systems should take into account and apply specific measures for both dimensions of security, while also implementing appropriate tools for its maintenance.

Furthermore, Union law requires the safeguarding of the integrity of networks and databases where copyright works and/or other subject-matters of protection are hosted<sup>64</sup> including the information extracted as the result of mining technologies. This inversely means that all necessary measures shall be taken for the ensuring of this integrity that also refer to the effective treatment and management of the relevant risks. As it had been respectively clarified, the medium through which information is transmitted is the network itself meaning that its appropriate treatment should not solely entail the issue of security but also that of capacity. In respect of the storage of the results extracted and of their further visualization, what is crucial is the identity of the database within which the said storage will take place. For this reason, and in any case, apart from the security measures cited above, the experts suggested the application of the principle of restricted access to this content, as well as that of classification of the access allowed on the basis and in proportion to the range of the recipients, otherwise of the public to which a given information is addressed.

Regarding the legal dimensions of the TDM technology and the implication of copyright law, the question concerning the factors which could impinge on the implementation of this new exception to libraries was answered by experts by stressing instead of the application of technological measures (as expected) the failure to secure the availability of data or to facilitate the disposal of the content that will subsequently form the subject-matter of mining. Furthermore, this important aspect did not merely entail the availability and/or facilitation or not of the data's disposal but it also referred to the volume of the data to be provided meaning that it should not be restricted to a given part of the relevant data's category. As result, it was not only the qualitative but also the quantitative nature of the data's provision that was highlighted that further entailed the principle of the data's timelessness and sustainability in order for researchers and users to be able to draw up satisfactory results.

Lastly, the benefit of TDM technology for libraries had been addressed; at first, experts stressed the need for the establishment of a supportive (to mining) information system which must be enriched with new entries meaning through the updated registration of the writings, textbooks and of the content in general of a given library in order for the relevant scientific research to become more efficient. In respect of its technical parameters, the mining technique involving documents pre-requires the digital format of a given content that is usually

<sup>64</sup>Article 3(3) of the DSM Directive.

made by the means of Optical Character Recognition (OCR) tools. The usage of such tools is also contributing to the further disposal of the information accrued that is complemented by the basic elements of such documents, i.e., its title, author and relevant keywords. It is profound that the new information produced is on the one hand enriching the relevant documentation, while on the other hand is capable of effectively contributing to the linguistic analysis of the texts used for the purpose of improving their classification. The direct effect of this procedure is the efficiency and success in maximum terms of the texts' detection in both simple and complicated searches consisting simultaneously of an additional benefit for libraries themselves.

## 5.2. Librarians: Interviews Results

Librarians are qualified themselves as the beneficiaries of the TDM exception for scientific research since they are explicitly included into the *rationae personae* of this mandatory provision as members of the staff of libraries. Being as such librarians are called upon to comprehend and implement in practice the new regulatory framework. Thus, their personal view regarding firstly the contribution of the TDM technology to the overall operation of libraries and secondly their role vis-à-vis this new possibility were the main issues of the interviews with them.

In the interviews conducted, librarians underlined that these TDM techniques are very useful for researchers and they will contribute to the progress of scientific research. At the same time, librarians will have the possibility to produce certain results in relation to the collection kept by a library by means of targeted searches both to bibliographic data and to databases.

In the light of the above, the views concerning the current modus operandi of libraries signified a diverging conceptualization on whether they remain or not user-oriented. This element is rather significant because it impacts on the services that libraries provide and subsequently on the intended enhancement of these services through the implication of mining techniques. In addition, Greek librarians are (complementary to the relevant literature) sharing the view that mining technologies are significantly contributing to the increase of the content that is made available to researchers and users. They underlined that this increase of the content results to the enhancement of the visibility of the conclusions reached and to the reformulation of the scientific research itself by aligning the relevant (existing) writings and publications with new findings. The impact of TDM on libraries' resource savings was stressed, since it won't be necessary anymore for them to pay extra costs for the licensing of access and use of the content made available through digital channels; in this regard, the recording and updating of the content of the libraries' collections (as to be made by librarians) will lead to the reconsideration and revision of their budget allocation.

Inversely, the answers given in relation to the role that the libraries may play for the purpose of leveraging mining technology under the new regime had demonstrated on the one hand the unanimous acceptance of this technique's un-

limited possibilities but on the other hand an incapability to provide for robust assessments on the extent of the opportunities arising therefrom.

More precisely, structured information is not and cannot be currently utilized since both connection and interconnection to information systems are not updated. On the one hand, the notion of connection is construed as the internal connection with databases. An additional problem is observed since certain qualitative features—such as typographical errors to their content—had not been yet resolved impinging as such on the success of searches through the forthcoming mining technologies. On the other hand, the notion of interconnection is rather critical for the practical implementation of mining technologies since the repositories of libraries—as related among others to Open Access policy and the information systems included therein are not “communicating” with each other having as such significant consequences for users since they are deprived from obtaining the overall image of the subject-matter that they are from time to time investigating. As a result, interoperability is not only related to the data storage as the experts from the IT sector had stressed above but also to the need to link and interconnect different information systems by the means of a multidimensional tool in order to satisfy the relevant needs in their entirety. However, librarians stressed that the undertaking of new and foremost drastic measures of a technical nature does not suffice *per se* for the achievement of the said objective; what is also needed is a change of attitude and the realization of the absolute necessity of the data’s identification and interconnection since these two parameters are actually consisting of the pillars of the further success of the search conducted through mining techniques.

Within the context of analyzing the role of their institutions librarians made some further suggestions—except from the establishment of new databases in order to assist and support scientific research and work—with their eyes cast to the future; these suggestions included the training of both librarians, relevant staff and users in respect of the advantages arising from mining technologies that should further entail the proper and safe use of these new tools. Moreover, the establishment and support of digital content repositories had been suggested such as that of CRIS<sup>65</sup> as an Information to Scientific Activities System, as well as the enhancement of collaboration at national and international level.

In relation to the licensing agreements that they conclude with publishers, librarians stressed the need for libraries to be collectively represented and to undertake horizontal action through their respective federations and associations such as that of the Association of Greek Academic Libraries (Sindesmos Ellinikon Acadimaikon Bibliothikon (SEAB)) with an aim to strengthen their negotiating power and to explore new resources.

Moreover, the impact of technical and legal issues to the operation in practical

<sup>65</sup>Current Research Information Systems (CRIS) had been considered as playing a determinative role in Open-Science Movement with regard to the maximization of publication of research results, as well as of their complementary information such as the specific channels through which such a content is shared and the project under which they had been concluded. See more at: [Evans \(2019\)](#).

terms of mining technologies was also raised by librarians; first they stressed their need to be informed or more precisely be familiarized with the technical aspects of mining asking for a clear guidance and relevant instructions. In addition, they highlighted the need to be trained on the legal dimensions of this technique. Also, they raised the issue of the applicable law, taking into account that libraries have to deal with works whose authors are not only nationals but also foreign authors and related-rights holders.

A view of major importance was that while TDM is widely used at international level, with reference made to the overall “ontological system” of libraries, TDM is not actually widespread in Greece due to the lack of the relevant know-how and expertise. This lack of human resources and foremost the absence of the required specialization and training of the existing personnel of libraries in order to deal with mining techniques had been demonstrated from the vast majority of librarians as an issue of critical importance. Librarians should themselves be familiarized and trained in this field with the assistance of specialized in this area experts recognizing simultaneously though that this requires both time and in-depth engagement<sup>66</sup>.

On the grounds of this last reference that is indisputably essential on the way towards the establishment and further implementation of national rules on TDM exception, librarians were asked about the challenges arising and which according to their own point of view should be reformed into practical and effective solutions. Accordingly, what was stressed was the need to find an appropriate equilibrium between the implementation of this exception and the prerequisites set-out for this application as being further related to secure data storage and maintenance of the integrity of the networks and databases in which copyright works and other subject-matters of protection are hosted.

Moreover, the development of open access databases and their constant support by various means (such as Arxiv and Repec)<sup>67</sup> was also identified as a challenge of a similarly crucial nature; in this regard, librarians also stressed the need for those databases’ interconnection through sophisticated and advanced tools and materials such as ontologies, the so-called “semantic Web” and artificial intelligence technologies. In addition, it had been stated that training activities should also be addressed to readers/users with regard to the proper use of both databases and of the data included therein further entailing also Big Data, Data Science and Data analytics. Lastly, the issue of personal data protection was also raised as being further related to users who have access to libraries’ databases

<sup>66</sup>It should be mentioned that there was also one point of view according to which such responsibilities should be assigned to experts of the information technology sector and not be conducted as such from the libraries’ personnel.

<sup>67</sup>According to the relevant definition “arXiv” consists of “a free distribution service and an open-access archive for 1,865,498 scholarly articles in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics”. See more at: <https://arxiv.org/> (last access April 2021). In addition, “RePEc (Research Papers in Economics) is a collaborative effort of hundreds of volunteers in 102 countries to enhance the dissemination of research in Economics and related sciences”. See more at: <http://repec.org/> (last access April 2021).

and overall information systems thus calling for the strict application of the General Data Protection Regulation (GDPR) and of the relevant national law.

This group of participants was also asked about whether already existing agreements with publishers should be modified or not in the light of the implementation of the DSM Directive and if affirmatively about their respective suggestions. However, the answers given to this question reflected the difficulties or even the dead ends that libraries are currently facing with regard to their efforts in finding new content and were not as such focused on the issues that have to be regulated at contractual level. More precisely, librarians indicated as an example the database created by SEAB under the title “Portico”<sup>68</sup>. Simultaneously though they expressed their reluctance in respect of the content provided therefrom in quantitative terms, thus stating accordingly that it is quite often that even more access restrictions are respectively arising having as a result the incapability of proceeding to mining techniques. The same restrictions are also observed in other sources thus limiting access to their own environment.

It is noteworthy that certain respondents said that existing agreements with publishers do not entail a specific clause referring to TDM. On the other hand, some said that this exception is already included into agreements made with foreign publishers giving further emphasis to the concept of data storage thus consisting of a point of friction between the parties involved. Furthermore, under the aim of safeguarding the work of researchers, librarians noted that the platforms to be used for TDM purposes should be authorized by libraries within the context of their competence and responsibility for the protection of both academic freedom and privacy; for example, it had been said that mining technologies should take place exclusively within the context of the libraries’ official structures and services and not through the researchers’ private blogs or other similar medium.

The issue of data storage and maintenance was also addressed. It should be noted at first that there were cases where either an incapability to respond was assessed or that a misconception was identified. In respect of the answers given, it was underlined that this objective could be achieved by both contractual means and technological measures. Moreover, the results reached from these interviews had highlighted one additional dimension of this issue as related to the distinction made between the storage made by the institution or the user himself. In the first case, it had been said that in order for libraries to be able to comply with their obligation to secure data storage and maintenance, they should be supported at national thus institutional level. This implies through the adoption and implementation of a targeted policy concerning storage space that should further encompass a number of qualitative matters such as the extent of the storage of data, the person liable for its conduct and/or for the case of data loss, the purpose that storage should serve, as well as personal data protection

<sup>68</sup>“Portico” consists of a community-supported preservation archive that safeguards access to e-journals, e-books, and digital collections. See more at the following link: <https://www.portico.org/> (last access April 2021).

issues. An important finding in this regard was also the correlation of data storage—as aligned with the searches made by users—to the potential implementation of filtering systems that would lead to the inspection or control of research activities that would conflict with the inherent character of libraries as areas of freedom of expression.

Moreover, the issue of best practices was also addressed as being aligned to the secure storage of the copies made during the operation of a given mining procedure, as well as to the integrity of the networks and databases in which the works made are hosted. It is also in this area observed that recipients were either not capable of providing for certain solutions since making a reference to information technology sector as a general comment or that they made a similarly general reference to the best practices existing in other jurisdictions (and not in Greece). In any case, it seems that there is not a specialized knowledge on this issue especially if considering that according to some points of view best practices should be construed at individual level meaning that they should be aligned and assessed in accordance with each person's capacity.

Lastly, the position of librarians was asked with regard to the feasibility or not of developing case studies and best-practices guide for the purpose of presenting the possibilities of TDM technologies. Here the answers given were universally positive thus mentioning both the feasibility and the usefulness of such studies and guides; however, it was simultaneously stressed that those studies do not suffice *per se* since what is crucial is the achievement of legal certainty and the subsequent accurate information both on the legitimacy and the extent to which specific acts will be allowed under the new rules to be established.

### 5.3. Legal Issues: Interviews Results

Both the exploring of the technical dimensions of TDM technologies and of the point of view of the librarians as the beneficiaries of this new exception, intended to record and highlight the different features and parameters that the Greek legislator has to take into account in order to be able to formulate an effective regulatory framework. The balance between the need on the one hand to remain as closer as possible to the letter and the spirit of Union law for the purpose of achieving a real harmonization of the Members-States' legal systems and of the Digital Single Market objective and to correspond on the other hand to the specific characteristics of the sectors affected at national level, forms undoubtedly a difficult equation and legal experts are called upon to contribute themselves to the reform of copyright law at national basis.

Within this context, certain key questions were addressed to experts in copyright law which had been considered as traversing the new exception as its basic pillars. At first, a general comment should be made; contrary to a part of the relevant literature that faces the exception established under Article 3 of the Directive as of a rather limitative character and scope, Greek experts run in favor of its restrictive scope (thus concerning exclusively research organizations and cultur-



al heritage institutions) stressing accordingly its positive sign towards the promotion of scientific research. This was not the case in relation to the exception provided under Article 4 of the Directive which seems to be dealt rather reluctantly on the grounds of whether it will be able or not to provide for an efficient regulatory framework for the actual facilitation of TDM techniques and of the massive use of the information produced within the context of machine learning of artificial intelligence systems.

Starting from the challenges arising from the establishment of both TDM exceptions<sup>69</sup>, what was first addressed was the issue of their implementation in practice—especially that of Article 4 of the Directive—as well as the clarification of the crucial concept of “lawful access” as being further aligned to the identification of the beneficiary institutions and of the relevant sources. In respect of the exceptions’ conceptual features, a concern was raised with regard to the exception dedicated to scientific research concerning the means and the way through which the non-profitable character of the respective beneficiaries will be assessed including in addition the implementation in practice of the significant opportunity provided to them concerning the Public-Private Partnerships to be sought. It is noteworthy that in relation to Article 4, the major challenge stated was the possibility provided under the Union legislator to overrule or at least not implement the exception as aligned with the prior consent and willingness of the holder of copyright and related rights over the works and other subject-matters of protection to be used for mining purposes.

As it had been respectively highlighted, the primary concern and essential condition for both the establishment and the implementation of the new rules of law, is the identification and clarification of the concepts requiring disambiguation. Greek legal experts had also themselves adopted and expressed their own concerns in respect of the immediate need to determine and deepen in the concepts of “lawful access”, of “scientific research”, of the right of copyright holders to opt-out in respect of the application of the exception provided under Article 4 and of the notion of “best practices” under Article 3 of the Directive directing in parallel their attention to the manner through which they will be at least initially determined.

Moreover, the issues arising from TDM technology *per se* were explored encompassing in addition its operation in practice for both libraries and users. At this point, it could be said that three different perspectives were respectively developed; the first one concerned the impact of reproduction either in whole or in part of the libraries’ databases. The second one focused mostly on the issue of users’ determination thus wondering whether all users of a library’s content should fall within the scope of application of this exception and if affirmatively under which circumstances and prerequisites. The third approach was based on how the libraries as beneficiaries of the exception will deal with the technical protection measures block access to works and other subject matter that are be-

<sup>69</sup>Meaning not only the exception for scientific research provided under Article 3 but also the general—under specific requirements—exception established under Article 4 of the DSM Directive.

ing lawfully accessed. On these grounds, an additional concern was raised on the exact means and tools through which libraries will be able to guarantee the implementation of this exception.

It is undoubted that the conflict between copyright and the right to access information (as made available from beneficiary institutions) as further triggering other fundamental rights such as those over personal data and/or the right to privacy, constitutes the quintessence of formulating fair at first and subsequently effective regulatory rules. Consisting simultaneously of the continuous exercise of the relevant case-law both at national and foremost at EU level, legal experts developed a general consideration on the necessary balancing of these diverging and conflicting rights and interests that needs to be achieved also in the integration of TDM exception(s) at national level. In respect of personal data, there is an imperative need to take into account and fully comply with the relevant regime given the fact that TDM exceptions are exclusively concerning copyright (and not other rights).

According to legal experts, apart from the said conflict of rights and interests and the fair balance that needs to be reached, other factors that may have an impact on the implementation of this exception for libraries, are TPMs as stated above and wider security issues. An additional issue raised was that of the need or at least of the feasibility to modify the terms under which information and content in general are provided from libraries to users especially in the case where the number of those willing to benefit from this new exception will be multiplied.

Consequently, legal experts highlighted three fields which according to their point of view need special consideration during the process of introducing these Union rules into the national legal order; the first one relates to the determination of their relationship and interference with TPMs in order to avoid the elimination of the application of this exception in practice. The second one is aligned to the non-establishment of a payment obligation on behalf of the beneficiaries of Article 3, while the last one concerns the need to provide for special mechanisms for the benefit of users. However, it should be noted that except for these general remarks the said suggestions were not further analyzed. It should also be noted that legal experts emphasized the need for the Greek legislator to follow at least at this stage the letter of the provisions of the DSM Directive further stating that any subsequent specification of the new regime should take place on the grounds of the relevant case-law and of the interpretation to be provided (if any) by the CJEU. In addition, it was underlined that these new provisions should be combined with the existing exceptions and limitations under national law<sup>70</sup> and foremost with those concerning libraries for the purpose of formulating a comprehensive and uniform legal framework.

Lastly, the position of legal experts was also sought in terms of the conditions under which the prerequisite of secure data storage and maintenance on behalf

<sup>70</sup>Provided under Chapter 4 of the Law No. 2121/1993 (<https://www.opi.gr/en/library/law-2121-1993#ch4>).

of the users proceeding to mining techniques would be fulfilled. On the one hand, some stressed the need to avoid the establishment of specified mechanisms and/or relevant solutions but instead to provide for general rules as being further aligned to the principle of proportionality. On the other hand, some special measures were accordingly suggested such as the implication of encryption methods and the avoidance of medium such as clouds for data storage; in this regard, TPMs were once again brought forward in an emphatical way thus being considered as effectively contributing to the achievement of the intended objectives under this new regime.

#### 5.4. Recommendations

After having examined both the conclusions of the questionnaire and the interviews with the IT and the legal experts as well as with the librarians a number of recommendations eventuate as inevitable in order TDM to bring out its full potential for academic libraries. As a first recommendation, awareness needs to be raised among the librarians. The better understanding of the TDM techniques and the concrete provision of the TDM exception in Union and soon national-level, its benefits for the research community and for academic libraries should be shared among them. Seminars, conferences and workshops could form a solution with regard to the need of these beneficiaries to be actually engaged in the digital environment. By attending such training programmes, the personnel of academic libraries will be informed with regard to the new legal regime and TDM techniques. In addition, guidelines could be drafted and distributed among academic libraries contributing thus to the targeted information and education of the personnel of academic libraries. Besides, librarians already have demanded through the questionnaires for the dissemination of information regarding the TDM techniques and the mandatory exception.

An additional recommendation is the training of the academic libraries' personnel in technical issues and specifically in the functioning and overall operation of TDM techniques aiming among others to the strengthening of their skills. This suggestion is further directly related to the capability of academic libraries to perform their activities and to provide their services in the most effective manner.

In this regard, the collaboration between academic libraries and the exchange of knowledge and know-how is deemed as significantly contributing to the fulfillment of the objectives pursued under this provision.

Further, as another recommendation, synergies between academic libraries and IT sector should be strengthened.

In view of the above-mentioned conclusions, another suggested recommendation is the elaboration of an institutional policy of the academic libraries vis-à-vis the performing of TDM, that could contribute to the effective use of TDM techniques while conducting scientific research and to the legal certainty.

Additionally, the enhancement of the cooperation between academic libraries

and publishers in order to conclude licensing agreements according to the new legislative framework and to resolve any issues arising while performing TDM is highly recommended. Regarding the data *per se*, the point of view of librarians and IT experts concentrated on the importance of homogenisation of text and data, that implies the update of libraries' databases and the interconnection between different databases and information systems. Reference was also made by the librarians to the development of open access databases and to the creation of digital content repositories. The storage of the datasets including copies of works and other subject-matter with the appropriate level of security to a safe environment with limited access should be ensured. Last, guidelines including best practices with respect to TDM is expected to support effectively the libraries

## 6. Conclusion

The adoption of mandatory exceptions at Union level is undoubtedly designating a new era for both *acquis communautaire* and the copyright regimes of Member States. The exceptions covering TDM, and foremost that of Article 3 that is devoted to scientific research, are expected to transform the scope of action and overall operation of libraries. On the other hand, it could be mentioned that the scope of application of this provision had been considered by part of the literature as extremely narrow. This limited scope of Article 3 under this perception extends also to the objective to be pursued (thus limited to scientific research)<sup>71</sup>. Moreover, an additional issue had been addressed on the grounds of the nature of the beneficiaries of the TDM exception as public entities in most of the cases; as it had been respectively stated, in the event where an external entity wishes to have access to the content produced and/or retained by such a public organisation that benefits from the exception, the legislation concerning the provision and re-use of public sector information is called into application. Under the relevant regime, the re-use of such information shall be free and, in any case, facilitated to the greatest extent possible and the usage of such data through TDM techniques falls profoundly within the scope of such a permitted re-use<sup>72</sup>. Except though from the criticism addressed what is crucial in practice is the non-bungling of this mandatory exception by the national legislator. Even if the success—in terms of legal certainty—of the TDM provisions could be questioned, national law shall provide for “one single provision on which TDM researchers with existing access to material can rely on makes for a clearer and more predictable system of norms compared to the current patchwork of exceptions, limitations and licences, none of which had quite been designed for this application”. In this regard, special emphasis had been given to the regulation of the right of copyright and related-right holders to apply for security measures

<sup>71</sup>This argument mainly lies on the alleged narrow definition of research organizations as provided under Article 2(1) of the Directive, and to the condition provided under Article 4 according to which text and data mining falls within the scope of application of the exception provided that the rightsholders had not restricted the use of their works and/or other subject-matters of protection.

<sup>72</sup>*Supra note 46.*

<sup>73</sup>Lindholm (2020).

and overall technological measures through which though the implementation in practice of the new rules could be hindered<sup>73</sup>.

Indeed, as the majority of the national copyright legal systems (including the Greek) did not provide for such a rule, the current or at least upcoming establishment of the exceptions concerning TDM indisputably consists of a significant step onwards. The traditional rules which did not by nature cover online and cross-border uses<sup>74</sup>—profoundly resulting to a rather limited (or even absent in practice) protection of copyright and related-rights holders in respect of the pervasive digital environment—will now be replaced, otherwise complemented in order to guarantee on the one hand the high level of protection and on the other hand the necessary legal certainty from users' behalf.

As stated above, TDM consists of the summation of statistical techniques, computer science and machine learning allowing as such the analysis of data from different disciplines that further results to the development of primary knowledge. Consequently, this tool—as bearing now the necessary legislative recognition—will profoundly contribute not only to the enhancement of scientific research but also to the enhancement of the role of libraries, of their openness and overall dynamics. Currently the role of national legislators is more than crucial thus having to transform these new provisions to effective national legal rules with an eye beyond the borders. Within this context, there is a need to clarify the notions that remain ambiguous or that in any case may impinge on the effectiveness of the rules to be established, while also identifying the beneficiaries of the exception. Besides, as it was highlighted from the above-mentioned conclusions with legal experts, the main challenges focus on the clarification of notions, such as “scientific research”, “lawful access”, on the application of “public-private partnership” or the definition of commonly agreed best practices according to Article 3(4) of the DSM Directive and last, on the co-existence of the TDM exception with the already provided exceptions and limitations in national legislations and in *acquis communautaire* and on the co-existence with other legal provisions relevant to TDM, such as personal data protection.

Moreover, it had been proposed that the medium to which data will be safely stored should be identified following a relevant dialogue between the interested parties, and that there should be clear provisions with regard to the application of TPMs.

From the above-mentioned conclusions of the surveys conducted through questionnaires and interviews, concrete recommendations arise in order for TDM to deploy its full potential for the academic libraries. The most important is raising awareness on both legal (Union legislation, national legislation and applicable law including) and technical issues in order for the librarians to understand the potential of TDM techniques and fully benefit therefrom; strengthening collaboration and synergies among libraries and between libraries and IT sector; cooperation with publishers; establishment of institutional policy of li-

<sup>74</sup>At least concerning the exceptions and limitations provided under Chapter IV of the Law No. 2121/1993.

libraries; the homogenisation of text and data, emphasizing on the interconnection between different databases and information systems; the storage of the datasets including copies of works and other subject-matter with the appropriate level of security to a safe environment with limited access. Last and most importantly guidelines including best practices with respect to TDM are expected to support effectively the libraries.

Additionally, it has to be mentioned that the tools established under the regime of Law No. 4451/2017 regulating TDM activities by NLG as they had been already implemented in practice could perform as a guideline at least in technical means for libraries which will undertake TDM techniques under the forthcoming implementation of the TDM exception at national level.

In conclusion, the era of Big Data and Artificial Intelligence brings new opportunities and challenges. TDM as a powerful tool has the potential to strengthen and enhance the conduction of scientific research for the benefit of the whole society. Copyright law had always been related to the evolution of technology. The DSM Directive provides for rules to adapt certain exceptions and limitations to copyright and related-rights to digital and cross-border environments, while keeping a high level of protection of copyright and related rights. At the same time, as the DSM Directive characteristically provides: “The objectives and the principles laid down by the Union copyright framework remain sound”<sup>75</sup>.

## Acknowledgements

This research is co-financed by Greece and the European Union (European Social Fund—ESF) through the Operational Programme “Human Resources Development, Education and Lifelong Learning 2014-2020” in the context of the project “Copyright and the exception of text and data mining in academic libraries” (MIS 5050521).

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- Bensamoun, A., & Bouquerel, Y. (2020). *Transposition des exceptions de fouilles de textes et de données: Enjeux et propositions* (pp. 1-105). Rapport de Mission CSPLA.
- Binctin, N. (2019). *TDM: Un enjeu de l'intelligence artificielle* (pp. 5-35). RIDA 262.
- Boston College Libraries (2021). *Data Management: Best Practices in Data Management*. [https://libguides.bc.edu/dataplan/best\\_practices](https://libguides.bc.edu/dataplan/best_practices)
- Bottis, M., Papadopoulos, M., Zampakolas, Ch., & Ganatsiou, P. (2019). On the Eve of Web-Harvesting and Web-Archiving for Libraries in Greece. *Erasmus Law Review*, No. 2, 178-189. <https://doi.org/10.5553/ELR.000135>
- Canellopoulou-Bottis, M., Papadopoulos, M., Zampakolas, C., & Ganatsiou, P. (2019).

<sup>75</sup>Recital 3 of the DSM directive.

- Text and Data Mining in the EU “Acquis Communautaire” Tinkering with TDM & Digital Legal Deposit. *Erasmus Law Review*, 12. <https://ssrn.com/abstract=3501670>
- Chopra, A., Prashar, A., & Sain, Ch. (2013). Natural Language Processing. *International Journal of Technology Enhancements and Emerging Engineering Research*, 1, 131-134. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.407.6907&rep=rep1&type=pdf>
- Copyright Working Group of REBIUN (2020). *Text and Data Mining (Articles 3 and 4 of the EU-DSM) by REBIUN’s Copyright Working Group*. <https://blogs.ifla.org/lpa/2020/06/10/text-and-data-mining-articles-3-and-4-of-the-eu-dsm/>
- Cox, A. (2018). Academic Librarianship as a Data Profession: The Familiar and Unfamiliar in the Data Role Spectrum. *eLucidate*, 15, 7-10.
- European Commission (2016). *Impact Assessment on the modernization of EU Copyright Rules, Part 1*. SWD, 301 Final, pp. 1-200.
- Evans, I. (2019). *How Using a Current Research Information System (CRIS) Can Support Open Science Implementation*. <https://www.elsevier.com/connect/how-using-a-current-research-information-system-cris-can-support-open-science-implementation>
- Ferri, F. (2020). The Dark Side(s) of the EU Directive on Copyright and Related Rights in the Digital Single Market. *China-EU Law Journal*. <https://link.springer.com/content/pdf/10.1007/s12689-020-00089-5.pdf>  
<https://doi.org/10.1007/s12689-020-00089-5>
- FutureTDM (2017). *Legal Guidelines for TDM Practitioners*. <https://www.futuretdm.eu/wp-content/uploads/5.3.1B-Legal-Guidelines.pdf>
- Geiger, C., Frosio, G., & Bulayenko, O. (2018). *The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market-Legal Aspects*. European Parliament. [https://www.europarl.europa.eu/RegData/etudes/IDAN/2018/604941/IPOL\\_IDA\(2018\)604941\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/IDAN/2018/604941/IPOL_IDA(2018)604941_EN.pdf)  
<https://doi.org/10.2139/ssrn.3160586>
- Geiger, C., Frosio, G., & Bulayenko, O. (2019). Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU. In C. S. García, & R. E. Llorca (Eds.), *Propiedad intelectual y mercado único digital europeo* (pp. 27-71). Valencia, Tirant lo blanch, Centre for International Intellectual Property Studies (CEIPI) Research Paper No. 2019-08. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3470653](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3470653)  
<https://doi.org/10.2139/ssrn.3470653>
- Gerrish, Ch., & Skavlan, A. M. (2019). European Copyright Law and the Text and Data Mining Exceptions and Limitations: In Light of the Recent DSM Directive, Is the EU Approach a Hindrance or Facilitator to Innovation in the Region? *Stockholm Intellectual Property Law Review*, 2, 1-80.
- Hilty, R., & Richter, H. (2017). *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)*. <https://www.ip.mpg.de/en/publications/details/position-statement-of-the-max-planck-institute-for-innovation-and-competition-on-the-proposed-modernisation-of-european-copyright-rules-part-b-excepti.html>  
<https://doi.org/10.2139/ssrn.2900110>
- Hugenholtz, B. (2019). *The New Copyright Directive: Text and Data Mining (Articles 3*



- and 4). Kluwer Copyright Blog.  
<http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/>
- Jondet, N. (2018). 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. *Propriétés Intellectuelles*, No. 67, 25-35.
- Kanellopoulou-Botti, M., Papadopoulos, M., Zampakolas, C., & Ganatsiou, P. (2019). Legal and Technical Issues for Text and Data Mining in Greece. In D. Wittkower (Ed.), *2019 Computer Ethics-Philosophical Enquiry (CEPE) Proceedings* (19 p).  
[https://digitalcommons.odu.edu/cepe\\_proceedings/vol2019/iss1/11](https://digitalcommons.odu.edu/cepe_proceedings/vol2019/iss1/11)
- Kovacevic, A., Devedzic, V., & Pocajt, V. (2010). Using Data Mining to Improve Digital Library Services. *The Electronic Library*, 28, 829-830.  
<https://doi.org/10.1108/02640471011093525>
- LACA (2019). *The Right to Read Is the Right to Mine: But Not When Blocked by Technical Protection Measures*. <https://uklaca.org/629/>
- Latreille, A. (2020). *Les aménagements favorisant l'accès à la connaissance* (pp. 221-246) (Articles 3-11 et 14 de la directive "DSM"). RIDA 264, 04/2020.
- Lindholm, M. (2020). *Text and Data Mining under Finnish Copyright Law before and after the DSM Directive*. Helsinki: Department of Accounting and Commercial Law, Hanken School of Economics.
- Míšek, J. (2019). Exception for Text and Data Mining for the Purposes of Scientific Research in the Context of Libraries and Repositories. In *12th Conference on Grey Literature and Repositories* (pp. 1-10).  
[https://repozitar.techlib.cz/record/1424/files/Misek\\_fulltext.pdf](https://repozitar.techlib.cz/record/1424/files/Misek_fulltext.pdf)
- Paramithiotis, Y. (2020). *Recent Developments in Copyright Law (Prosfates Exelixeis sto Dikaio Pneumatikis Idioktias)*. Research Documents (Ereunitika Keimena) IME GSEBEE. (In Greek)  
<https://imegsevee.gr/wp-content/uploads/2020/12/paramithiotis.pdf>
- Proudman, V., Stratton, B., Vézina, B., White, B., & Wyber, S. (2019). *Transposing the Directive on Copyright in the Digital Single Market: A Guide for Libraries and Library Associations*. [https://zenodo.org/record/3552203#.X7RFJxZS\\_IU](https://zenodo.org/record/3552203#.X7RFJxZS_IU)
- Quintais, J. P. (2020). The New Copyright in the Digital Single Market Directive: A Critical Look. *European Intellectual Property Review*, 42, 28-41.  
<https://doi.org/10.2139/ssrn.3424770>
- Rattan, P. (2019). Data Mining: A Library Utility Model. *European Journal of Research*, No. 1, 39-45.
- Regulating Text and Data Mining in the European Union: Issues and Challenges* (2020).  
<https://www.lexology.com/library/detail.aspx?g=3ac6190c-f9b9-479c-bafd-290897127b2f>
- Rosati, E. (2018). An EU Text and Data Mining Exception for the Few: Would It Make Sense? *Journal of Intellectual Property Law & Practice*, 13, 429-430.  
<https://doi.org/10.1093/jiplp/jpy063>
- Rosati, E. (2019). Copyright as an Obstacle or an Enabler? A European Perspective on Text and Data Mining and its Role in the Development of AI Creativity. *Asia Pacific Law Review*, 27, 198-217. <https://ssrn.com/abstract=3452376>  
<https://doi.org/10.1080/10192557.2019.1705525>
- Rosati, E. (2020). *The United Kingdom Will Not Transpose the DSM Directive*.  
<https://ipkitten.blogspot.com/2020/01/the-united-kingdom-will-not-transpose.html>

- Sobrino-García, I. (2020). Copyright in the Scientific Community. The Limitations and Exceptions in the European Union and Spanish Legal Frameworks. *Publications*, 8, 27. <https://www.mdpi.com/2304-6775/8/2/27/html>  
<https://doi.org/10.3390/publications8020027>
- Stamatoudi, I. (2016). Text and Data Mining. In I. Stamatoudi (Ed.), *New Developments in EU and International Copyright Law* (pp. 251-282). Leiden, Netherlands: Kluwer Law International.
- Stewart, N., Secker, J., Morrison, Ch., & Horton, L. (2016). *Liberating Data: How Libraries and Librarians Can Help Researchers with Text and Data Mining*. <https://blogs.lse.ac.uk/impactofsocialsciences/2016/07/12/how-libraries-and-librarians-can-help-with-text-and-data-mining/>
- The European Parliament Must Improve the Text and Data Mining (TDM) (2018) *Exception to Benefit European Research and Innovation*. [https://eare.eu/assets/uploads/2018/03/OpenLetter-to-JURI-Committee-on-TDM\\_26March2018.pdf](https://eare.eu/assets/uploads/2018/03/OpenLetter-to-JURI-Committee-on-TDM_26March2018.pdf)
- Triaille, J.-P., de Meeûs d'Argenteuil, J., & de Francquen, A. (2014). *Study on the Legal Framework of Text and Data Mining* (pp. 1-122). De Wolf & Partners for the European Commission.
- Vavousis, K., Papadopoulos, M., Gerolimos, M., & Xenakis, Ch. (2020). Text and Data Mining for the National Library of Greece in Consideration of Internet Security and GDPR. *Qualitative and Quantitative Methods in Libraries*, 9, 441-460. <http://www.qqml-journal.net/index.php/qqml/article/view/637>
- White, B. (2020). *Research Libraries: How You Can Support Text and Data Mining*. [https://zenodo.org/record/3801114#.X7LySxZS\\_IV](https://zenodo.org/record/3801114#.X7LySxZS_IV)