

# On the Eve of Web-Harvesting and Web-Archiving for Libraries in Greece

Maria Bottis, Marinos Papadopoulos, Christos Zampakolas & Paraskevi Ganatsiou\*

## Abstract

This conference paper submitted on the occasion of the 8th International Conference on Information Law and Ethics (University of Antwerp, December 13-14, 2018) that focused on modern intellectual property governance and openness in Europe elaborates upon the Text and Data Mining (TDM) issue in the field of scientific research, which is still-by the time of composition of this paper-in the process of discussion and forthcoming voting before the European Parliament in the form of provision(s) included in a new Directive on Copyright in the Digital Single Market. TDM is included in the proposal for a Directive of the European parliament and of the Council on copyright in the Digital Single Market-Proposal COM(2016)593 final 2016/0280(COD) that was submitted to the European Parliament.

**Keywords:** web harvesting, data analysis, text & data mining, TDM: Proposal EU Copyright Directive

*The paradox of intellectual property lies in a “system that promotes, or at least, aspires to promote knowledge [...] by restricting it.”*  
P. Bernt Hugenholtz

## 1 What TDM Is

This conference paper submitted on the occasion of the Eighth International Conference on Information Law and Ethics (University of Antwerp, 13 and 14 December 2018) focused on modern intellectual property governance and openness in Europe elaborates upon the text and data mining (TDM) issue in the field of scientific research, which is still – by the time of composition of this article – in the process of discussion and forthcoming voting before the European Parliament in the form of provision(s) included in a new Directive on Copyright in the Digital Single Market. TDM is

\* Work co-funded by Greece and the European Social Fund (ESF) through the Operational Program ‘Human Resources Development, Education and Lifelong Learning’ for the implementation of the ESF & the Youth Employment Initiative in Greece.

Maria Bottis, Associate Professor, Department of Archives, Library Science and Museology, Ionian University, Corfu, Greece. botti@otenet.gr.  
Marinos Papadopoulos, Attorney-at-Law, Independent Researcher, PhD, MSc, JD, Athens, Greece. marinos@marinos.com.gr.  
Christos Zampakolas, Archivist/Librarian, Independent Researcher, PhD, MA, BA, Ioannina, Greece. christoszampakolas@gmail.com.  
Paraskevi Ganatsiou, Educator, MA, BA, Prefecture of Ionian Islands, Corfu, Greece. pganatsiou@gmail.com.

included in the proposal for a Directive of the European parliament and of the Council on copyright in the Digital Single Market – Proposal COM(2016)593 final 2016/0280(COD) that was submitted to the European Parliament.

On 11 September 2018, the aforesaid proposal was furnished to the European Parliament for voting in its plenary sitting. However, controversies concerning Article 11 that caters for protection of press publications regarding digital uses and Article 13 that pertains to use of protected content by information society service providers storing and giving access to large amounts of works and other subject matter uploaded by their users resulted in the proposal being referred back to the Commission pursuant to Rule 59(4) of the Rules of Procedure of the European Parliament; Parliament’s first reading was therefore not closed and negotiations with the Council begun.

This article considers the text of the Proposal COM(2016)593 final 2016/0280(COD) as it has been until the composition of this writing, that is, until January 2019. This COM(2016)593 final 2016/0280(COD) proposal is consistent with the existing EU copyright legal framework. This proposal is based upon and complements the rules laid down in Directive 96/9/EC – the Database Directive<sup>1</sup> –, Directive 2001/29/EC – the Information Society (InfoSoc) Directive<sup>2</sup> –, Directive 2006/115/EC,<sup>3</sup> Directive 2009/24/EC,<sup>4</sup> Directive 2012/28/EU<sup>5</sup> and Directive 2014/26/EU.<sup>6</sup> Those Directives, as well as the Proposal COM(2016)593 final 2016/0280(COD) contribute to the functioning of the internal market, aim at the smooth development of the Digital Single Market within the EU, ensure a high level of protection for right holders and facilitate the clearance of rights.

1. Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.
2. Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society.
3. Directive 2006/115/EC of the European Parliament and of the Council of 12 December 2006 on rental right and lending right and on certain rights related to copyright in the field of intellectual property.
4. Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs.
5. Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on certain permitted uses of orphan works.
6. Directive 2014/26/EU of the European Parliament and of the Council of 26 February 2014 on collective management of copyright and related rights and multi-territorial licensing of rights in musical works for online use in the internal market.

TDM is understood as the automated computational analysis of information in digital form, such as text, sounds, images or data, that is enabled through the use of new computational technologies.<sup>7</sup> In a broad sense, TDM is called any activity where computer technology is used to index, analyze, evaluate and interpret mass quantities of content and data.<sup>8</sup>

The statutory exception of TDM pertains to activities that are confined to acts of ‘automated processing of large amounts of structured digital textual content, for purposes of information retrieval, extraction, interpretation, and analysis’,<sup>9</sup> which are undertaken for scientific research purposes. In her benchmark 2011 report, Eefke Smit refers to TDM as ‘automated tools, techniques or technology to process large volumes of digital content that is often not well structured – to identify and select relevant information; to extract information from the content, to identify relationships within/between/ across documents and incidents or events for meta-analysis’.<sup>10</sup> Aside from the term ‘text and data mining’, which is usually referred with the TDM initials, the notions of text mining, text data mining, content mining and computational text analysis are often used interchangeably with the ‘text and data analysis’ or the ‘text and data mining’ with the aim to describe a TDM inquiry<sup>11</sup> or an analytical TDM approach.<sup>12</sup>

TDM works in the following manner:<sup>13</sup>

1. It identifies input materials to be analyzed, such as works, or data individually collected or organized in a pre-existing database;
2. It copies substantial quantities of materials – which encompasses
  - a. pre-processing materials by turning them into a machine-readable format compatible with the

technology to be deployed for the TDM so that structured data can be extracted and

- b. possibly, but not necessarily, uploading the pre-processed materials on a platform, depending on the TDM technique to be deployed;
3. It extracts the data; and
4. It recombines data to identify patterns into the final output.

Once access to content is available or granted, TDM generally implies the reproduction of the text or the data, either temporarily, for example, by caching the content or permanently, for example, by creating a database of key elements for facilitating searches (index). There are also TDM technologies that allow for analyzing content without making any copies of the analyzed content, for example, by website crawling or screen-scraping. TDM tools involving minimal copying of few words or crawling through data and processing each item separately could be operated without running into potential liability for copyright infringement. This follows from the fact that copyright law does not protect data but only original expressions within copyright-protected subject matter. In this respect, the proposal for a new Directive on copyright in the Digital Single Market clarifies that ‘text and data mining may also be carried out in relation to mere facts or data which are not protected by copyright and in such instances no authorization would be required’.<sup>14</sup> Obviously, although the proposal, as it has by the time of this writing, fails to specifically mention that, also works and other subject matter not protected by copyright or the sui generis right can be freely mined.<sup>15</sup>

Content that is text and data mined may come in different formats, such as machine-readable formats (*e.g.*, XML) or PDFs, which may be more or less easily mined. The data retrieved often needs to be normalized, annotated and aggregated into a corpus to allow for an efficient use of mining software. The normalization and annotation can be done either by the publishers, including as part of a commercial offer (*e.g.*, data in an XML format, provided in a structured way) or by the researchers themselves, which is more the case for researchers in the public interest research organizations, who tend to prefer using their own tools (relying also more on PDFs than commercial users). The normalization and annotation phase of TDM activity involves the preprocessing to standardize materials into machine-readable formats; activity in this phase might trigger infringement of the right of reproduction of works found online.<sup>16</sup> Likewise, the uploading of the pre-processed material on a platform – which might occur or not depending on whether the TDM technique adopted makes use of a TDM software crawling data to be analyzed directly from the source – might also violate the right of reproduction. The process of analyzing the texts or data is to be distinguished from its result. The output

7. See, Proposal COM(2016)593 final 2016/0280(COD), Recital 8; see also, European Commission, COM(2016)593 final 2016/0280(COD) (2016), according to which *Text and Data Mining (TDM)* is a term commonly used to describe the automated processing (‘machine reading’) of large volumes of text and data to uncover new knowledge or insights.

8. M. Caspers, L. Guibault, K. McNeice, S. Piperidis, K. Pouli, M. Eskevich & M. Gavriilidou, *Reducing Barriers and Increasing Uptake of Text and Data Mining for Research Environments Using a Collaborative Knowledge and Open Information Approach*, Baseline report of policies and barriers of TDM in Europe (extended version of D3.3) (2016) 9, available at: [https://cordis.europa.eu/project/rcn/197301\\_en.html](https://cordis.europa.eu/project/rcn/197301_en.html) (last visited 20 November 2018).

9. B.F. Reilly, *When Machines Do Research, Part 2: Text-Mining and Libraries* (2012) 75-76.

10. E. Smit and M. Van der Graaf, *Journal Article Mining, a research study into practices, policies, plans....and promises*, Commissioned by the Publishing Research Consortium, Amsterdam (2011), available at: <http://publishingresearchconsortium.com/index.php/128-prc-projects/research-reports/journal-article-mining-research-report/160-journal-article-mining> (last visited 20 November 2018).

11. See, C. Bergman, L. Hunter & A. Rzhetsky, *Announcing the PLOS Text Mining Collection* (2013), available at: <https://blogs.plos.org/everyone/2013/04/17/announcing-the-plos-text-mining-collection/> (last visited 20 November 2018).

12. Reilly, above at n. 9, pp.75-76.

13. C. Geiger, G. Frosio & O. Bulayenko, *The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market-Legal Aspects* (2018) 5-6, available at: [www.europarl.europa.eu/RegData/etudes/IDAN/2018/604941/IPOL\\_IDA\(2018\)604941\\_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/IDAN/2018/604941/IPOL_IDA(2018)604941_EN.pdf) (last visited 20 November 2018).

14. Recital 8 of COM(2016) 593 final.

15. Geiger *et al.*, above at n. 13, p. 6.

16. *Ibid.*

of TDM might consist, for example, of a summary of the analyzed text and data, visualizations such as graphics or charts and also of new knowledge, patterns and combinations of data that may lead to new discoveries and research results.<sup>17</sup> However, the analysis and extraction of the TDM process, that is, the phase where data is finally extracted – can also infringe upon the right of reproduction depending on the mining software deployed and the character of the extraction.<sup>18</sup>

Regarding TDM activity on databases, TDM might involve the reproduction, translation, adaptation, arrangement and any other alteration of a database protected by copyright, which means the original selection and arrangement of the database's content.<sup>19</sup> TDM activity might, also, infringe sui generis database right, in particular the extraction – and to a minor extent the re-utilization – of substantial parts of a database or the repeated extraction of insubstantial parts of a database. In this context, even if extraction does occur without reproduction of the original materials, extraction itself would infringe upon the exclusive sui generis right provided to the database owner.<sup>20</sup> According to the CJ,<sup>21</sup> the infringement occurs by unauthorized actions for the purpose of reconstituting, through the cumulative effect of acts of extraction, the whole or a substantial part of the contents of a database protected by the sui generis right and/or of making available to the public, through the cumulative effect of acts of re-utilization, the whole or a substantial part of the contents of such a database, which thus seriously prejudice the investment made by the maker of the database. Article 7(5) of the Database Directive refers to unauthorized acts of extraction or re-utilization the cumulative effect of which is to reconstitute and/or make available to the public, without the authorization of the maker of the database, the whole or a substantial part of the contents of that database and thereby seriously prejudice the investment by the maker.

## 2 Article 3 on TDM of the Proposed Directive on Copyright in the Digital Single Market

The provision of Article 3 of the proposed Directive on copyright in the Digital Single Market has as follows:

### Article 3 Text and data mining

1. Member States shall provide for an exception to the rights provided for in Article 2 of Directive 2001/29/EC, Articles 5(a) and 7(1) of Directive 96/9/EC and Article 11(1) of this Directive for reproductions and extractions made by research organizations in order to carry out text and data mining of works or other subject matter to which they have lawful access for the purposes of scientific research.
2. Any contractual provision contrary to the exception provided for in paragraph 1 shall be unenforceable.
3. Rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject-matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.
4. Member States shall encourage rightholders and research organizations to define commonly-agreed best practices concerning the application of the measures referred to in paragraph 3.

Regarding this provision, the following are striking: The TDM exception's beneficiaries are limited to research organizations. The meaning of 'research organizations' is defined in Article 2(1) of the proposed Directive; according to it:

'research organization' means a university, a research institute or any other organization, the primary goal of which is to conduct scientific research or to conduct scientific research and provide educational services, too:

- a. on a non-for-profit basis or by reinvesting all the profits in its scientific research; or
- b. pursuant to a public interest mission recognized by a Member State;

in such a way that the access to the results generated by the scientific research cannot be enjoyed on a preferential basis by an undertaking exercising a decisive influence upon such organization.

The term 'scientific research' in Article 3 of the proposed Directive on copyright in the Digital Single Market is understood as in the definition of 'research' put forward by the OECD; according to it, research is understood as 'creative work undertaken on a systematic basis in order to increase the stock of knowledge,

17. European Commission, Commission Staff Working Document, 'Impact Assessment on the Modernization of EU Copyright Rules', SWD(2016) 301 final PART 1/3 (2016) 158, available through: <https://ec.europa.eu/digital-single-market/en/news/impact-assessment-modernisation-eu-copyright-rules> (last visited 20 November 2018).

18. Geiger et al., above at n. 13, p. 6.

19. *Ibid.*, p. 7.

20. *Ibid.* See, CJ, *The British Horseracing Board Ltd and Others v. William Hill Organization Ltd*, C-203/02 (9 November 2004), available at: <http://curia.europa.eu/juris/liste.jsf?oqp=&for=&mat=or&lgrec=el&jge=&td=%3BALL&jur=C%2CT%2CF&num=C-203%252F02&page=1&dates=&pcs=Oor&lg=&pro=&nat=or&cit=none%252CC%252CCJ%252CR%252C2008E%252C%252C%252C%252C%252C%252C%252C%252C%252C%252Ctrue%252Cfalse%252Cfalse&language=en&avg=&cid=5068790> (last visited 20 November 2018).

21. *Ibid.*

including knowledge of man, culture and society, and the use of this stock of knowledge to devise new applications'.<sup>22</sup> Scientific research lies in the ambit of that definition. In any case of questionable research activity, the burden would lie on the shoulders of the user to prove that the TDM activity undertaken was carried out for scientific research purposes.

The statutory exception of Article 3 in the text of *COM(2016)593 final 2016/0280(COD)* does not discriminate between types of subject matter covered, between the sources of works or kinds of databases, or between categories of beneficiaries. From the wording of Recital 10 of the aforesaid proposal for a Directive on copyright in the Digital Single Market, it becomes evident that the EU legislator aims at a wide variety of entities throughout Europe – the primary goal of which is to conduct scientific research or to do so together with the provision of educational services. These research organizations or research and educational services providers are the beneficial organizations across the EU, which the EU legislator targets regarding the TDM new mandatory exception. Article 3 of said proposal does not discriminate between types and subject matter covered by the beneficiary organizations since the wording of Article 3 applies to different legal forms and structures of research organizations across Member States, which have in common that they act either on a not-for-profit basis or in the context of a public-interest mission recognized by the State. Such a public-interest mission may, for example, be reflected through public funding or through provisions in national laws or public contracts. For profit organizations, commercial entities are not excluded from the application of Article 3 of the proposal insofar as they operate and/or deploy TDM in the context of a public-interest mission recognized by the State. The fact that for-profit organizations are not excluded from the provision of Article 3 regarding TDM activity gives this provision – the trait of being inclusive.<sup>23</sup> For research organizations in which the commercial undertakings have a decisive influence allowing them to exercise control because of structural situations such as their quality of shareholders or members, which may result in preferential access to the results of the research, Recital 10 of the proposed Directive clearly sets them out of the pool of beneficiary

organizations aimed by the EU legislator regarding TDM mandatory exception.<sup>24</sup>

This approach in the statutory exception of Article 3 of the proposed Directive on copyright in the Digital Single Market, which does not discriminate between types of subject matter covered, between the sources of works or kinds of databases or between categories of beneficiaries coincides with the research exception recognized in Article 5(3)(a) of the Information Society Directive and in Article 6(2) of the Database Directive; said provisions for which there's further analysis below in this text do not discriminate between categories of works, sources or users. The introduction of the statutory exception of Article 3 regarding TDM aims at altering Directive 2001/29/EC and Directive 96/6/EC regarding right-holder's power on copyrighted works and databases, which could hamper the deployment of text and data mining activity. This goal of adaptation of the aforesaid Directives is clear through the text of Recital 5 of *COM(2016)593 final 2016/0280(COD)*.<sup>25</sup>

The proposal for a new Directive on copyright in the Digital Single Market indicates that confining the TDM exception to non-commercial research activities only was not the choice of EU legislator. Such a restriction for the application of the TDM exception could slow down the pace of innovation, for it is not only non-commercial research that generates socially and economically valuable outcomes. Moreover, making the distinction between what is commercial and what is non-commercial may be very difficult in practice, especially in the case of public/private partnerships (PPP), the commercial character of which is often very difficult to ascertain.<sup>26</sup> Thus, in said cases wherein there exists a commercial aspect of the undertaken research, the key-point to consider is the existence of a public-interest mission recognized by the State within which TDM may be deployed leveraging on the exemption of Article 3 of the proposed new Directive on copyright in the Digital Single Market. Thus, in addition to the requirement of the non-commercial nature of the scientific research activity, there is also the requirement of a public-interest mission of the organization which undertakes the research activity. The former notion is narrower than the latter, in the sense that public-interest mission could include commercial research activities.

The requirement of non-commercial research activities follows the lines already set by the Database Directive

22. I. Hargreaves, L. Guibault, C. Handke, B. Martens, R. Lynch & S. Filippov, *Standardisation in the Area of Innovation and Technological Development, Notably in the Field of Text and Data Mining – Report from the Expert Group* (2014), European Union, *Study on the Legal Framework of Text and Data Mining (TDM)*, De Wolf & Partners (2014) 55; F. Manual, *Proposed Standard Practice for Surveys on Research and Experimental Development*, OECD (2002), available at: <https://www.oecd-ilibrary.org/docserver/9789264199040-en.pdf?expires=1542611355&id=id&accname=guest&checksum=39B756986E0ECF728154E3785B2AA363> (last visited 20 November 2018).

23. See, Geiger *et al.*, above n. 13, p. 19, according to who the TDM exception's scope is very inclusive as it applies both to commercial and non-commercial uses and – very importantly – cannot be overridden by contract.

24. According to Recital 10 of the proposed Directive on copyright in the Digital Single Market: 'organisations upon which commercial undertakings have a decisive influence allowing them to exercise control because of structural situations such as their quality of shareholders or members, which may result in preferential access to the results of the research, should not be considered research organisations for the purposes of this Directive'.

25. According to Recital 5 of the proposed Directive on copyright in the Digital Single Market: 'For uses not covered by the exceptions or the limitation provided for in this Directive, the exceptions and limitations existing in Union law should continue to apply. Directives 96/9/EC and 2001/29/EC should be adapted'.

26. Hargreaves *et al.*, above at n. 22, p. 56.

and the Information Society Directive. Recital 42 of the latter Directive specifies that

when applying the exception or limitation for non-commercial educational and scientific research purposes, including distance learning, the non-commercial nature of the activity in question should be determined by that activity as such. The organizational structure and the means of funding of the establishment concerned are not the decisive factors in this respect.

Outside the context of non-commercial research, though, the creation of corpora can be difficult to reconcile with the strict rules of copyright and database rights. In all circumstances, or commercial scientific research or even with non-commercial research, and provided that there's no element of public-interest mission of the organization doing the research, license agreements and website terms of use can impose TDM deployment under certain restrictions.

The provision of Article 3(3) of the proposed Directive on copyright in the Digital Single Market sets a limitation allowing rightholders to introduce measures to protect the 'security and integrity' of their networks and databases where works are hosted. However, such measures shall not go beyond what is necessary to achieve that objective of the provision of Article 3. Recital 12 of the proposed Directive is very clear per subject matter of said limitation: In view of a potentially high number of access requests to and downloads of their works or other subject matter, rightholders should be allowed to apply measures where there is risk that the security and integrity of the system or databases where the works or other subject matter are hosted would be jeopardised. Those measures 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. Besides, according to Article 3(2) of the proposed Directive, any contractual provision contrary to the exception provided for in Article 3(1) shall be unenforceable; thus, the provision of Article 3(2) sets protection from possible contractual override of the TDM mandatory exception.

In accordance with the international obligations of the European Union under Article 10 of the WIPO Copyright Treaty, the new exception described in Article 3 of the proposed Directive for copyright in the Digital Single Market needs to comply with the requirements of the so-called 'three-step-test', for example, that the exception is applicable only in certain special cases that do not conflict with a normal exploitation of the work and do not unreasonably prejudice the legitimate interests of the author.<sup>27</sup> For this reason, Recital 6 of the proposed Directive clearly states that the exceptions and the limitation set out in the proposed Directive seek to achieve a fair balance between the rights and interests of authors and other rightholders, on the one hand, and of

users, on the other hand. They can be applied only in certain special cases which do not conflict with the normal exploitation of the works or other subject matter and do not unreasonably prejudice the legitimate interests of the rightholders. Also, Article 6 of the proposed Directive rules that Article 5(5) and the first, third and fifth subparagraphs of Article 6(4) of the InfoSoc Directive 2001/29/EC shall apply to the exceptions and the limitation provided for under the proposed Directive. Article 5(5) of the InfoSoc Directive rules, the so-called 'three-step-test', while Article 6(4) first subparagraph refers to Technological Protection Means (TPM) and specifically to the obligation of Member States to prevent the derogation from the provision of TDM activity as an exception to copyright in the sense that they are obliged to take appropriate measures to ensure that rightholders make available to the beneficiary of the TDM exception the means of benefiting from that exception, to the extent necessary to benefit from that exception and where that beneficiary has legal access to the protected work or subject matter concerned. Also, Article 6(4) third and fifth subparagraphs refer to the protection afforded to TPMs.

TDM technologies allow researchers to process large amounts of information to gain new knowledge and discover new trends. While TDM technologies are prevalent across the digital economy, there is widespread acknowledgment that TDM can, in particular, benefit the research community, and in so doing encourage innovation. Jonathan Clark<sup>28</sup> notes four main reasons to engage in TDM:

- a. To enrich content: Mining can improve indexing, be deployed to create relevant links, and improve the reading experience.
- b. To engage in systematic review of literature: Mining can help researchers systematically review larger bodies of content, faster than they could do it themselves and to keep up with their field, without missing relevant information.
- c. To discover new knowledge: Mining can be used to create databases that can themselves be mined.
- d. To engage in computational linguistics research: Mining itself is the subject of research, for example to improve the extraction of meaning from texts.

### 3 TDM in a New Copyright Directive in the Digital Single Market

Exceptions and limitations to copyright and neighboring rights have not yet being harmonized at the EU level, and this fragmentation in the implementation of exceptions and limitations from the Member States causes

27. *Ibid.*, p. 54.

28. J. Clark, *Text Mining and Scholarly Publishing, a report for the Publishing Research Consortium*, Loosdrecht, The Netherlands & London (2013) 7.

legal uncertainty that affects TDM Europe-wide.<sup>29</sup> The non-mandatory nature of most of InfoSoc Directive's list of exceptions and limitations to copyright is a cause of failure in the process of harmonization of copyright rules applicable in all Member States of the EU.<sup>30</sup> The non-harmonized EU legal framework for exceptions and limitations, especially those pertaining to scientific research and teaching, which have not implemented nationally by EU Member States in the same way due to their non-mandatory nature, cause significant difficulties in leveraging on the existing legal framework for Copyright for covering the TDM activity. Though these exceptions per research or teaching, a.k.a. education, aim at achieving public policy objectives, there is no sameness in understanding and fulfilling said objectives among Member States. In addition, as new types of uses have recently emerged, it remains uncertain whether these exceptions are still adapted to achieve a fair balance between the rights and interests of authors and other rightholders, on the one hand, and of users, on the other hand. Besides, these exceptions remain national and legal certainty around cross-border uses is not guaranteed. As a consequence, cross-border collaborations of researchers are hindered by the lack of sameness in understanding and applying the research exception or limitation to copyright; this affects TDM activities directly since researchers are unaware – or face high transaction costs for clearance – of whether TDM would be lawful across all EU jurisdictions involved in the research collaboration.<sup>31</sup> The situation of legal uncertainty is further affected by combinations of contractual and technical measures, which are frequently used to create insurmountable hurdles for researchers engaging in TDM projects. Actually, contractual and technological barriers are also frequently used to prevent TDM activities on materials not protected by copyright or on public domain subject matter,<sup>32</sup> and the CJ has ruled that said use of contractual and technological means on non-protected by copyright or the sui generis right databases is not illegal.<sup>33</sup>

In order to overcome the problems caused due to the lack of harmonization of Copyright law, and especially in the field of exceptions and limitations of copyright,

which impedes the Digital Single Market goal, the European Commission has identified three areas of intervention; these three areas of intervention are the following:

- a. digital and cross-border uses in the field of education,
- b. text and data mining in the field of scientific research, and
- c. preservation of cultural heritage.

The objective of the European Commission is to guarantee the legality of certain types of uses in these fields, including across borders. As a result of a modernized framework of exceptions and limitations, researchers will benefit from a clearer legal space to use innovative text and data mining research tools, teachers and students will be able to take full advantage of digital technologies at all levels of education and cultural heritage institutions (*i.e.*, publicly accessible libraries or museums, archives or film or audio heritage institutions) will be supported in their efforts to preserve the cultural heritage, to the ultimate advantage of EU citizens.

Regarding TDM, four options were considered for its proposed regulation:

- a. Option 1 consisted in self-regulation initiatives from the industry. This option pertained to contractual agreements, including clauses allowing for TDM. This option was deemed to be inappropriate for harmonization.
- b. Option 2 consisted in the introduction of mandatory exception for TDM covering uses pursuing a non-commercial scientific research purpose.
- c. Option 3 allowed uses for commercial scientific research purpose but limited the benefit of the exception to some beneficiaries.
- d. Option 4 went further as it did not restrict beneficiaries.

Of these options, the introduction of mandatory exception for TDM covering uses pursuing non-commercial scientific research purposes, but also allowing uses for commercial scientific research purposes limited to some beneficiaries seems to have prevailed – at least, so far – being deemed to be the most proportionate one.<sup>34</sup> This option was deemed to be the best in terms of maximization of legal certainty and minimization of copyright clearance costs for research organizations in the EU, including research projects which are carried out with a possible commercial outcome.<sup>35</sup>

29. Geiger *et al.*, above at n. 13, p. 12.

30. Geiger *et al.*, above at n. 13, pp. 14-15 and references in footnotes 65, 68, 70. A unified and mandatory approach is especially crucial in the digital environment as the Internet involves uses that, most of the time, affect several copyright legislations, leading to a major insecurity regarding what is allowed.

31. Geiger *et al.*, above at n. 13, pp. 12-13.

32. Geiger *et al.*, above at n. 13, p. 13.

33. In case, *Case C-30/14 (2015), Ryanair Ltd v. PR Aviation BV*, the CJ ruled that Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases must be interpreted as meaning that it is not applicable to a database which is not protected either by copyright or by the sui generis right under that directive, so that Arts. 6(1), 8 and 15 of that directive do not preclude the author of such a database from laying down contractual limitations on its use by third parties, without prejudice to the applicable national law. See, CJ's ruling on *Case C-30/14*, available at: <http://curia.europa.eu/juris/document/document.jsf?docid=161388&doclang=EN> (last visited 20 November 2018).

34. See, European Commission, Commission Staff Working Document, 'Executive Summary of the Impact Assessment on the Modernization of EU Copyright Rules', *SWD(2016) 302 final* (2016), according to which 'For TDM, the preferred option is a mandatory exception applicable to research organizations acting in the public interest such as universities or research institutes. The exception would allow them to carry out TDM on content they have lawful access to, for the purposes of scientific research'. The Executive Summary of the Impact Assessment is available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52016SC0302&from=EN> (last visited 20 November 2018).

35. See, European Commission, above at n. 34, according to which 'The new TDM exception would increase legal certainty and reduce rights clearance costs for research organizations, including when research pro-

Option 1, and specifically the view that TDM could be self-regulated in the market through contractual agreements, was supported in France.<sup>36</sup> In July 2014, the High Council on Artistic and Literary Property (*Conseil Supérieur de la Propriété Littéraire et Artistique* [CSPLA]), the advisory body in charge of advising the Ministry of Culture on copyright issues, submitted its report on TDM.<sup>37</sup> The approach, analysis and the recommendations of the French report were almost the opposite of those of the Hargreaves Review in the United Kingdom.<sup>38</sup> While the British report asked how to adapt copyright to the needs of the economy, the French report was more concerned with affording as much protection as possible to copyright against TDM, which it compares to a parasite.<sup>39</sup> The CSPLA report on the legal aspects of TDM concluded that none of the exceptions in French copyright law offered enough guarantees to allow TDM; especially not the teaching exception, the French implementation of Article 5(3)(a) of the InfoSoc Directive, given its very limited scope in French law.<sup>40</sup> According to the CSPLA report, it is not possible to modify national law without a change in the EU framework. Implicitly, the CSPLA report rejected the British analysis on the ability to create a new exception within the existing framework. Crucially, for the French report the creation of such a new exception was not even necessary as contractual solutions should be promoted. It proposed to ‘favor self-regulation over statutory changes’ and set ‘a two-year period after which a sectorial overview will be conducted and the need for legislative change assessed’.<sup>41</sup> The CSPLA report also recommended that the French government should share this wait-and-see approach and oppose any initiative to reform copyright at European or international level.<sup>42</sup> Once more, this was in stark contrast with the Hargreaves review, which had urged the UK government to press the EU to change its copyright law.

Evidence gathered through the review process that preceded to the proposal for a new Directive on Copyright in the Digital Single Market has highlighted that the

jects are carried with a possible commercial outcome, e.g. in the context of PPPs’.

36. N. Jondet, ‘L’ Exception Pour Le Data Mining Dans Le Projet De Directive Sur Le Droit D’ Auteur: Pourquoi L’ Union Européenne Doit Aller Plus Loin Que Les Législations Des États Membres (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)’, 67 *Propriétés Intellectuelles* (2018) 25-35, available at: <https://ssrn.com/abstract=3239374> (last visited 20 November 2018).
37. J. Martin and L.D. Carvalho, *Mission sur l’ exploration de données* (« Text and Data mining ») (2014), available at: <https://docplayer.fr/1430465-Mission-sur-l-exploration-de-donnees.html> (last visited 20 November 2018).
38. I. Hargreaves, *Digital Opportunity – A Review of Intellectual Property and Growth* (2011), available at: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/32563/ipreview-finalreport.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/32563/ipreview-finalreport.pdf) (last visited 20 November 2018).
39. Jondet, above at n. 36; Martin and Carvalho, above at n. 37, p. 2.
40. Jondet, above at n. 36; Martin and Carvalho, above at n. 37, p. 30.
41. Jondet, above at n. 36; Martin and Carvalho, above at n. 37, p. 4, recommendations 5, 6 and 7, respectively.
42. Jondet, above at n. 36; Martin and Carvalho, above at n. 37, p. 5, recommendations 11 and 12, respectively.

research exception has not been implemented in all Member States and that in any event it has generally been implemented without explicitly taking into account of TDM (which can be explained by the relatively novelty of these techniques). So far, a specific TDM exception in the context of the research exception has been adopted by the United Kingdom, which is going to exclude itself from being a Member State of the EU, very soon.<sup>43</sup> Estonia has also introduced a TDM exception to its Copyright Law. France and Germany have amended their Copyright laws passing a TDM provision, too. Greece has yet to amend its Copyright law regarding TDM. However, recently a law regulating subject matter on the National Library of Greece introduced TDM – actually, the Web Archiving – as one of the many responsibilities and statutory goals of the National Library of Greece (NLG).

As a consequence of the fact that very few Member States, such as the United Kingdom,<sup>44</sup> France,<sup>45</sup> Estonia<sup>46</sup> and Germany,<sup>47</sup> have amended their laws allowing for TDM, considerable legal uncertainty exists as to the EU framework applicable to TDM in scientific research and different conditions apply depending on the Member States and rightholders’ licensing practices.<sup>48</sup>

43. European Commission, Commission Staff Working Document, ‘Impact Assessment on the Modernization of EU Copyright Rules’, *SWD(2016) 301 final PART 2/3* (2016) 51, available through: <https://ec.europa.eu/digital-single-market/en/news/impact-assessment-modernisation-eu-copyright-rules> (last visited 20 November 2018).
44. The UK legislator amended its Copyright law by S.I. 1992/3233, regulation 7, S.I. 1997/3032, regulation 8 and S.I. 2003/2498, regulation 9. Section 29A that was added to the Copyright and Rights in Performances (Research, Education, Libraries and Archives) Regulations 2014 came into force on 1 June 2014. The amended Copyright law in the United Kingdom provides for TDM to the lawful user for the sole purpose of computational analysis for non-commercial research, but does not cover the reproduction of databases. See, [www.legislation.gov.uk/uksi/2014/1372/regulation/3/made](http://www.legislation.gov.uk/uksi/2014/1372/regulation/3/made) (last visited 20 November 2018).
45. In France, the legislator of Law No. 2016-1231 for a Digital Republic (Loi pour une République numérique) introduced TDM exceptions both applying to works (Art. L.122-5, 10 of the CPI) and databases (Art. L.342-3, 5 of the CPI). French exceptions cover acts of reproduction from ‘lawful sources’ (materials lawfully made available with the consent of the rightholders) for TDM as well as storage and communication of files created in the course of TDM research activities. The introduction of TDM in the French Intellectual Property Code was implemented in Art. 38 of the Law No. 2016-1231 for a Digital Republic which added paragraph 10 to Art. L.122-5 and paragraph 5 to Art. L.342-3 of the French Intellectual Property Code (Code de la Propriété Intellectuelle, CPI).
46. The Estonian legislator amended the country’s Copyright Act of 1992 and as of 1 January 2017 introduced TDM in paragraph 3 of Art. 19 titled ‘Free Use of Works for Scientific, Educational, Informational and Judicial Purposes’.
47. In 1 September 2017 Germany amended its Copyright law and the amendment has come into force as of 1 March 2018 introducing TDM in Section 60d titled ‘Text and Data Mining’. The TDM exception in German law covers the acts of reproduction necessary for undertaking TDM and the acts of making available of the corpus of materials produced by TDM activity (e.g., source materials that were normalized, structured and categorized) to a specifically limited circle of persons for their joint scientific research, as well as to individual third persons for the purpose of monitoring the quality of scientific research.
48. European Commission, above at n. 43, p. 52.

## 4 The EU Legislator's Approach on TDM

For the EU legislator, TDM is just a means to achieve the goal of Digital Single Market. The goal for an EU Digital Single Marketing is a goal for the free movement of goods, persons, services and capital where individuals and businesses can seamlessly access and exercise online activities under conditions of fair competition, and a high level of consumer and personal data protection, irrespective of their nationality or place of residence. The Digital Single Market strategy<sup>49</sup> considers three pillars in its foundation:

1. Better access for consumers and businesses to online goods and services across Europe. This requires the rapid removal of key differences between the online and offline worlds to break down barriers to cross-border online activity.
2. Creating the right conditions for digital networks and services to flourish. This requires high-speed, secure and trustworthy infrastructures and content services, supported by the right regulatory conditions for innovation, investment, fair competition and a level playing field.
3. Maximizing the growth potential of the European Digital Economy. This requires investment in ICT infrastructures and technologies such as Cloud computing and Big Data, and research and innovation to boost industrial competitiveness as well as better public services, inclusiveness and skills.

Regarding the achievement of the first pillar, that is, better access for consumers and businesses to online goods and services across Europe, there's a requirement for a more harmonized copyright regime which provides incentives to create and invest while allowing transmission and consumption of content across borders, building on Europe's rich cultural diversity. To this end, the Commission has been working on proposed solutions that include:

- a. portability of legally acquired content,
- b. cross-border access to legally purchased online services while respecting the value of rights in the audiovisual sector,
- c. greater legal certainty for the cross-border use of content for specific purposes (*e.g.*, research, education, text and data mining) through harmonized exceptions,
- d. clarification of the rules on the activities of intermediaries in relation to copyright-protected content and
- e. modernization of enforcement of intellectual property rights, focusing on commercial-scale infringements

49. See, *COM(2015) 192 final*, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee, and the Committee of the Regions, A Digital Single Market Strategy for Europe, available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2015%3A192%3AFIN> (last visited 20 November 2018).

(the 'follow the money' approach) as well as its cross-border applicability.

The TDM issue pertains to the harmonization of exceptions and limitations in copyright law of Member States, the creation of legal certainty for cross-border use of content for the purpose of scientific research.

The EU legislator has considered – at least for the time being – recommendations made by various scholars upon the TDM and how it should be regulated in the proposed Directive on copyright in the Digital Single Market. The suggestion that it is best to have a mandatory exception for TDM which would be inspired from and contain partly the same conditions as the scientific research exception, but which would have its own characteristics has prevailed, so far.

The mandatory character of the provision of Article 3 on text and data mining in the text of the proposed Directive on copyright in the Digital Single Market can normally be decomposed into three elements, that is:<sup>50</sup>

- a. be implemented across all Member States in order to ensure effective harmonization of the law;
- b. do not be subject to contractual overrides; and
- c. do not be subject to lock-up behind technological protection measures.

Even when the owner (or holder) of the data cannot exercise copyright or database rights, contractual restrictions or technical protection measures may render TDM more burdensome or even impossible.<sup>51</sup> For this reason, the wording in the proposed Article 3 rules that:

- a. Member States 'shall provide' for an exception ...  
The wording is not 'may provide' but 'shall provide' which indicates the mandatory character of the proposed provision.
- b. Art.3(2) of the provision rules that any contractual provision contrary to the exception provided for in paragraph 1 shall be unenforceable, thus the owner (or holder) of the data cannot exercise copyright or database rights through contractual restrictions that could hamper the TDM activity.
- c. Art.3(3) of the provision rules that 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. These measures include technological protection measures such as DRM. Thus, technical protection measures may not render TDM burdensome or even impossible.

There were many suggestions on how to encourage TDM for research purposes without fear of infringing IP rights. The goal for such an encouragement through legislative action could be achieved in a number of ways:<sup>52</sup> through an adjustment of licensing practices; through a revised, normative interpretation of the

50. Hargreaves, above at n. 22, p. 57.

51. Hargreaves, above at n. 22, p. 59.

52. Hargreaves, above at n. 22, p. 52.

reproduction right in copyright; through the introduction of a new mandatory exception in copyright and database laws, or through the adoption of an ‘open norm’ designed to guide the courts to take a more flexible view of what users are permitted to do.

In consideration of *COM(2016)593 final 2016/0280(COD)*, there’s no doubt that the EU legislator is inclined towards the choice of introducing a mandatory exception for TDM covering uses pursuing non-commercial scientific research purposes, but also allowing uses for commercial scientific research purposes limited to some beneficiaries, and also of ensuring that TDM regulation cannot be over-ridden through the enforcement of restrictive contractual clauses or technological protection measures. The point of contention between the introduction of a new mandatory exception and the facilitation of TDM in consideration of the existing exception for scientific research has found its solution in the introduction of a new mandatory exception. The license option, a.k.a. the encouragement of TDM through licensing was deemed to be inefficient and not adequate for creating legal certainty among Member States regarding TDM for scientific research.<sup>53</sup> The extent to which TDM in Europe is facilitated by any existing exceptions to either EU copyright or database law appeared unclear. The application of a copyright and database exception relating to teaching or scientific research is optional and has not been implemented at all in some Member States. This has contributed to uncertainty in the European scientific research community.<sup>54</sup> Moreover, it was considered that unless a TDM mandatory exception applicable horizontally for all Member States were passed, the possibility of enacting different TDM legislations in Member States is possible, and as a consequence, the fragmentation of the Single Market is more than likely to increase over time as a result of Member States adopting TDM exceptions at national level which could be based on dif-

ferent conditions, which is likely to happen in the absence of intervention at EU level.<sup>55</sup>

The introduction of a new mandatory exception in copyright and the database law may take one of two forms:

- a. an exception specifically permitting TDM for the purpose of scientific research or
- b. an open norm.

The first form provides more immediate clarity and the second form offers more flexibility in a fast-changing technological environment.<sup>56</sup> With an exception on copyright and database right specifically permitting TDM for the purpose of scientific research the assessment of whether an act of TDM is lawful is made *ex ante* by the legislator, while with an open norm the assessment of the lawfulness of an act of TDM would be made *ex post* by the judge.<sup>57</sup> Article 3 of the Proposal of the European Parliament and of the Council on copyright in the Digital Single Market, a.k.a. *COM(2016)593 final 2016/0280(COD)*, describes clearly the form of an exception specifically permitting TDM for the purpose of scientific research. The EU legislator has opted not to frame TDM through an open-norm description in the proposed Directive for copyright in the Digital Single Market because of considerations for possible legal uncertainty; it was deemed best to address the issue of TDM by providing for a mandatory exception to the right of reproduction and also to the right to prevent extraction from a database. The new mandatory exception should be understood as being without prejudice to the existing mandatory exception on temporary acts of reproduction laid down in Article 5(1) of Directive 2001/29/EC, which should continue to apply to text and data mining techniques which do not involve the making of copies going beyond the scope of that exception.<sup>58</sup>

The open norm as a form to regulate TDM is presented as an option by Ian Hargreaves et al. (2014) expert group report on TDM.<sup>59</sup> The idea for an open norm in European Copyright law is not new. The introduction of an open norm – or general exception – similar to U.S. fair use has long been considered in the EU legal scholarship and policy debate.<sup>60</sup> Supporters of the open-

53. Researchers have generally considered that licenses-based solutions would not be able to fully solve the problems of legal uncertainty they face as regards the use of TDM techniques. This was also confirmed in these stakeholders’ replies to a 2013-2014 public consultation (institutional users such as libraries and universities generally considered licenses an inadequate source of transaction costs for TDM and indicated that a legislative change is needed to introduce a mandatory exception for text and data mining in EU copyright law). See, European Commission, above at n. 43, pp. 51-52.

54. Researchers are generally convinced of the potential of TDM but they put forward legal uncertainty, caused by the current copyright rules, as one of the reasons for the slow development of TDM in the EU (in addition to aspects unrelated to copyright, such as lack of awareness and skills and infrastructural challenges). A considerable level of legal uncertainty exists among researchers regarding TDM and copyright law. Research organizations and researchers do not always know whether TDM is copyright-relevant at all, whether it may be covered by an exception or whether a specific rightholders’ authorization is required. See, more at European Commission, Commission Staff Working Document, ‘Impact Assessment on the Modernization of EU Copyright Rules’, *SWD(2016) 301 final PART 1/3* (2016) 104-5, available at: <https://ec.europa.eu/digital-single-market/en/news/impact-assessment-modernisation-eu-copyright-rules> (last visited 20 November 2018).

55. European Commission, above at n. 54, p.106.

56. Hargreaves, above at n. 22, p. 54.

57. *Ibid.*

58. Recital 10, the proposed Directive for copyright in the Digital Single Market, clarifies that this exception still applies but its application would be limited to TDM techniques which involve only the making of temporary reproductions transient or incidental to an integral and essential part of a technological process which enables a lawful use with no independent economic significance. According to Recital 10, ‘The new exception should be without prejudice to the existing mandatory exception on temporary acts of reproduction laid down in Article 5(1) of Directive 2001/29, which should continue to apply to text and data mining techniques which do not involve the making of copies going beyond the scope of that exception. Research organisations should also benefit from the exception when they engage into public-private partnerships’.

59. Hargreaves, above at n. 22, pp. 6, 54, 57.

60. See, M. Senftleben, ‘The Perfect Match – Civil Law Judges and Open-Ended Fair Use Provisions’, 33 *American University International Law*

norm option in TDM claim that the open norm could introduce flexibility so as to allow TDM activities to take place, along with other types of activities that would pass the test. The introduction of an open norm in Copyright and Database law, though, would have required an interpretation of the ‘three-step test’ in copyright law in a balanced way<sup>61</sup> along the lines of the ‘Declaration on a Balanced Interpretation of the “Three-Step Test” in Copyright Law’.<sup>62</sup> Instead of a restrictive reading of the test that would require exceptions and limitations to be interpreted narrowly, the aforesaid Declaration suggests ‘an appropriately balanced interpretation of the three-step test under which existing exceptions and limitations within domestic law are not unduly restricted and the introduction of appropriately balanced exceptions and limitations is not precluded.’<sup>63</sup> The non-restrictive reading of the ‘Three-Step Test’ in European Copyright law could be seen as an attempt to instil in the European Copyright law the flexibility and adaptiveness to new circumstances in the market imposed by technological evolution that characterizes the provisions of Common Law – especially the flexibility and adaptiveness of American Law on applied Copyright through the ‘fair use’ doctrine. The decision of the EU to propose a new Copyright Directive with the aim to boost the Digital Single Market comes at a time when considerations upon the appropriateness of the existing legal framework for Copyright acknowledge that there’s hardly a solid legal

foundation for TDM in the ‘acquis communautaire’. Provisions such as Article 5(1) or Article 5(3)(a) of the InfoSoc Directive or Article 6(1), Article 6(2)(b) and Article 9(b) of the Database Directive do not suffice for covering TDM.

## 5 Article 4(4)(b) of Greek Law 4452/2017 for TDM of NLG

A recent development in Greece’s legal framework on the National Library of Greece (NLG) stipulates for activities that are within the TDM operation. Specifically, law 4452/2017 which is titled ‘Regulation on State Language Certificate subject matter, on the National Library of Greece and on other provisions’ includes in its text the provision of Article 4(4)(b), according to which the National Library of Greece operates as the official National Depository and Archive of digital publications, data and metadata produced in the country or related to Greek culture. This operation includes the monitoring and archiving of the Internet (web archiving) or other technology environment. To this end, the National Library of Greece shall undertake, allocate and coordinate the actions concerned at national level.

This provision of Article 4(4)(b) of law 4452/2017 is the first provision in the Greek legal system that caters for TDM activities. Said provision is too general, probably vague, and not proper in its wording. However, the analysis in this text does not aim at elaborating upon the bad phrasing or vagueness in the provision of Article 4(4)(b) of law 4452/2017.

Article 4(4)(b) of law 4452/2017 sets the TDM activity under the responsibility of the National Library of Greece which is named as the organization to undertake, allocate and coordinate action of text and data analysis at national level. The ‘monitoring’ of the web is meant to be the web harvesting activity; the archiving of the Internet is meant to be the archiving of works harvested from the Internet. Thus, the National Library of Greece is ruled to be the proper organization for TDM activity in Greece. Other organizations may deploy TDM activities under the coordination of the National Library of Greece, which is the national depository and archive of works on the Internet, including data and metadata, produced in Greece or are related to the Greek culture.

Article 4(4)(b) of law 4452/2017 precedes any EU regulation upon TDM. The proposal for a Directive on copyright in the Digital Single Market has yet to pass the European Parliament’s vote. Article 3 of said proposal has yet to become part of the ‘*acquis communautaire*’.

## 6 Conclusion

Regarding the proposal of the European parliament and of the Council on copyright in the Digital Single Market

Review 231 (2017) 286, available at: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3002275](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3002275) (last visited 20 November 2018); B.P. Hugenholtz, ‘Flexible Copyright: Can EU Author’s Rights Accommodate Fair Use?’, in Irini Stamatoudi (ed.), *New Developments in EU and International Copyright Law*, Kluwer Law International, Leiden, The Netherlands (2016) 417-33; B.P. Hugenholtz and M. Senftleben, *Fair Use in Europe: in Search of Flexibilities*, Amsterdam Law School Research Paper No. 2012-39 – Institute for Information Law Research Paper No. 2012-39 (2012), available at: <https://ssrn.com/abstract=2013239> (last visited 20 November 2018); M. Senftleben, ‘Comparative Approaches to Fair Use: An Important Impulse for Reforms in EU Copyright Law’, in Graeme Dinwoodie (ed.), *Methods and Perspectives in Intellectual Property*, Edward Elgar, Cheltenham (2014); A. Dnes, ‘Should the UK Move to a Fair-Use Copyright Exception’, 44(4) *International Review of Intellectual Property and Competition Law* 418 (2013) 444; R. Van der Noll, S. Gompel, L. Guibault, J. Weda, J. Poort, I. Akker & K. Breemen, *Flexible Copyright: The Law and Economics of Introducing an Open Norm in the Netherlands*, SEO Economic Research Report N. 2012-60 (2012), available at: [https://www.ivir.nl/publicaties/download/Flexible\\_Copyright.pdf](https://www.ivir.nl/publicaties/download/Flexible_Copyright.pdf) (last visited 20 November 2018); C. Geiger, ‘Flexibilising Copyright – Remedies to the Privatisation of Information by Copyright Law’, 39(2) *International Review of Intellectual Property and Competition Law* 178 (2008) 197, available at: [https://www.researchgate.net/publication/43233985\\_Flexibilising\\_Copyright\\_-\\_Remedies\\_to\\_the\\_Privatisation\\_of\\_Information\\_by\\_Copyright\\_Law](https://www.researchgate.net/publication/43233985_Flexibilising_Copyright_-_Remedies_to_the_Privatisation_of_Information_by_Copyright_Law) (last visited 20 November 2018).

61. Hargreaves, above n. 22, p. 6; Geiger, above n. 13, p. 16.

62. See, *Declaration – A Balanced Interpretation of the “Three-Step Test” in Copyright Law*, available at: <https://www.jipitec.eu/issues/jipitec-1-2-2010/2621/Declaration-Balanced-Interpretation-Of-The-Three-Step-Test.pdf> (last visited 20 November 2018).

63. See, C. Geiger, D.J. Gervais & M. Senftleben, *The Three-Step-Test Revisited: How to Use the Test’s Flexibility in National Copyright Law* (2013), available at: <https://digitalcommons.wcl.american.edu/cgi/viewcontent.cgi?referer=https://www.google.gr/&httpsredir=1&article=1041&context=research> (last visited 20 November 2018).

– Proposal COM(2016)593 final 2016/0280(COD) as its text has till January 2019 – for a new Copyright Directive, the major positive impact of it lies in its focus on harmonization of Member States' Copyright laws, through a mandatory solution for TDM. Directive 2001/29/EC has failed to address adequately the intersection between Copyright, technological measures and contracts. Rather, it has provisioned in such a way that Copyright exceptions may easily prove to be ineffective because of the contractual override enabled by the interplay of electronic contracts setting out conditions of legal access to the copyrighted work and access- and copy-control technologies such as DRM systems. Also, the InfoSoc Directive has failed to put in place regulatory mandatory ceilings which could have an adverse effect for the possibility to derogate from existing limitations on a contractual basis.<sup>64</sup> Substantive ceilings should have been provisioned in the InfoSoc Directive stipulating that exceptions such as non-transformative private use of works or use of them for the purpose of scientific research are mandatory in the sense that they can neither be contracted away nor being denied through the use of DRM technology which expands technological exclusivity beyond Copyright laws or has the capacity to bypass statutory limitations to Copyright.<sup>65</sup> Said regulatory mandatory ceilings could positively oblige rightholders to ensure that beneficiaries can exercise the exceptions and limitations in Copyright law in spite of contractual agreements or DRM technology that leave room to the contrary.<sup>66</sup>

The proposal for a new Copyright Directive which understands – at last – that harmonization in Member States' Copyright laws may come through the mandatory nature of exceptions or limitations to Copyright is a welcome arrangement that promotes harmonization and, therefore, the Digital Single Market. As such, a harmonized framework for TDM research will be driving innovation in the expectation for a Digital Single Market in the EU, promoting EU-wide, integrated, larger research projects. Said harmonization as well as EU's competitiveness will also be supported by an expansive scope of the limitation, covering both commercial and non-commercial uses of the TDM output, and the unenforceability of contrary contractual provisions or a deviation from DRM technology that could nullify TDM attempts.

Further than that, there's serious consideration whether the TDM exception's beneficiaries should not be limited to 'research organizations'. We believe that they should not. Leveraging on text and data analysis will gradually become a key-point for development for all legal entities; consideration for individual researchers or physical persons should be described in the TDM mandatory exception, too. Actually, the existing UK exception for text and data analysis does not discriminate between legal and physical persons, but rather allows TDM activity to any person with lawful access to a work.

Also, consideration for cultural heritage institutions should be taken, too. Limiting beneficiaries only to 'research organizations' would undermine a widespread assumption that the 'right to read should be the right to mine'.<sup>67</sup> In addition, limiting the TDM mandatory exception only to non-commercial research does not seem reasonable. Both, commercial and non-commercial research could fit in the TDM mandatory exception from the reproduction right of the copyright holder.

The notion of 'lawful access' to a work could hamper TDM in the sense of de facto subject TDM research to private ordering. According to the European Copyright Society, 'the exception can effectively be denied to certain users by a right holder who refuses to grant "lawful access" to works or who grants such access on a conditional basis only'.<sup>68</sup> In addition, subjecting TDM to

64. There are examples in international treaties' law of such mandatory exceptions which can override contractual limitations, such as Arts. 5(2), and (3) and Art. 6 in connection with Art. 9(1) of EC Directive 91/250/EEC on the legal protection of computer programs. Also, Art. 5(1) of the InfoSoc Directive 2001/29/EC, as well as Art. 6(1), and Art. 8, and Art. 15 of the EC Directive 96/9/EC on the legal protection of databases. In addition, EU Commission's, 'Green Paper on Copyright in the Knowledge Society', (COM(2008) 466 final, available at: <https://eur-lex.europa.eu/legal-content/GA/TXT/?uri=CELEX:52009AE0613> (last visited 28 December 2018) raises the issue of making certain categories of exceptions from the InfoSoc Directive mandatory for all EU Member States.

65. According to EU Commission's, above n. 64, at present, the Community list of copyright exceptions comprises one mandatory exception and twenty optional exceptions; Member States being therefore free to decide whether or not they wish to implement the optional exceptions. The EESC believes that this represents a key obstacle to the genuine harmonisation of those exceptions which may be justified in a knowledge economy, via technological methods which are constantly changing in the digital age. However, since this list is exhaustive, it prevents the introduction of other exceptions by various Member States. Furthermore, through the application of the 'three-step test' drawn up by the WTO and the WIPO, such limitations are subject to three conditions: they may apply only to certain special cases (e.g., visually impaired users), they may not be in conflict with the normal exploitation of the work and they may not unreasonably prejudice the legitimate interests of the right holder.

66. Prescribing a minimum framework of mandatory public policy exceptions which must be available in all national laws of WIPO members, i.e., prescribing a regulatory-ceiling framework for exceptions and limitations of Copyright is a task that has already been undertaken by the WIPO Standing Committee on Copyright and Related Rights through a proposal made by Brazil, Chile, Nicaragua and Uruguay. See, World Intellectual Property Organization, *Proposal by Brazil et al. Relating to Limitations and Exceptions* (SCCR/16/2), 17 July 2008, available at: [www.wipo.int/meetings/en/doc\\_details.jsp?doc\\_id=107712](http://www.wipo.int/meetings/en/doc_details.jsp?doc_id=107712) (last visited

28 December 2018). See, also, The A2K (Access to Knowledge), *Treaty on Access to Knowledge*, draft text (2005), available at: [www.cptech.org/a2k/a2k\\_treaty\\_may9.pdf](http://www.cptech.org/a2k/a2k_treaty_may9.pdf) (last visited 28 December 2018) which contains a catalogue of mandatory limitations and exceptions to Copyright including provisions regarding distance education, provisions for persons with disabilities, the first sale doctrine for library use, provisions for ISPs, for DRMs, for orphan works, for the term protection for Copyright, provisions expanding the knowledge commons, provisions promoting the Open Standards, etc. Though the Treaty on Access to Knowledge has been drafted with the aim to become part of WIPO agenda, it has yet to become part of it officially.

67. See, P. Murray-Rust, 'Open Knowledge Foundation', *The Right to Read Is the Right to Mine* (2012), available at: <https://blog.okfn.org/2012/06/01/the-right-to-read-is-the-right-to-mine> (last visited 28 December 2018).

68. European Copyright Society, *General Opinion on the EU Copyright Reform Package* (2017) 4, available at: [https://www.ivir.nl/publicaties/download/ECS\\_opinion\\_on\\_EU\\_copyright\\_reform.pdf](https://www.ivir.nl/publicaties/download/ECS_opinion_on_EU_copyright_reform.pdf) (last visited



Ευρωπαϊκή Ένωση  
Ευρωπαϊκό Κοινωνικό Ταμείο

Επιχειρησιακό Πρόγραμμα  
Ανάπτυξη Ανθρώπινου Δυναμικού,  
Εκπαίδευση και Διά Βίου Μάθηση  
Ειδική Υπηρεσία Διαχείρισης  
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης



ανάπτυξη - εργασία - αλληλεγγύη

lawful access will make TDM research projects harder to run by raising related costs.<sup>69</sup>

The notion of 'normal use' of a database might receive multiple interpretations according to the Member States in which it is applied. Therefore, harmonization of TDM exception could be accompanied with a delineation of the notion of 'normal use' of a database Europe-wide.<sup>70</sup>

Regarding the provision of Article 4(4)(b) of law 4452/2017 in Greece, this is far from being the introduction of the mandatory TDM exception in the Greek legal framework. As is described above hereto, it is merely a provision assigning NLG with the responsibility of undertaking TDM activity in Greece as well as of coordinating TDM activities enacted by other organizations. Said provision in the Greek law sets the notion of 'web archiving' for the first time in the Greek legal system. Despite the fact that the Greek legislator's aim was to describe the notion of TDM which includes the activity of 'web archiving' through this provision – it was hardly achieved through a clear and articulate provision in law – the important thing is that it describes in the statutory goals of NLG the activity of researching the Web on the purpose of archiving works or data which refer to the Greek culture. By no means this should drive to the conclusion that NLG is or will become the only entity with the capability for TDM activity in Greece. However, by Article 4(4)(b) of law 4452/2017 NLG is assigned with – and will probably manage to keep – prime role in the TDM activity deployed on the purpose of scientific research upon Greek culture. It remains to be seen.

The fact that the Greek legislator passed a law catering for the NLG's statutory responsibility for TDM on the purpose of scientific research for Greek culture subject matter, but has yet to pass a law on the exception of TDM to copyright is not uncommon. There are other Member States which have assigned the responsibility for TDM to their National Library, but in which TDM has yet to become an exception to copyright. Ireland is one such Member State. The National Library of Ireland (NLI) has a long-standing tradition of collecting, preserving and making accessible the published and printed output of Ireland. The NLI has been archiving the Irish web on a selective basis since 2011, and it has

over 17 TB of data in the selective web archive,<sup>71</sup> openly available for research through the NLI website. In 2007 and 2017, the NLI undertook domain crawling projects and there is now over 43 TB of data archived from these crawls. The National Library of Ireland is a legal deposit library, entitling it to a copy of everything published in Ireland. However, unlike many countries in Europe, legal deposit legislation in Ireland does not currently extend to online material so the NLI cannot make these crawls available to the public. Despite these barriers, the NLI remains committed to preserving the online story of Ireland in whatever way it can.<sup>72</sup>

An amendment to the Greek Copyright Law 2121/1993 is expected, regarding a provision for TDM, at least, and as a consequence of the pass of the proposed Directive on copyright in the Digital Single Market – when it will be set for voting before the European Parliament, again. In addition to the text of the forthcoming Directive provisioning the TDM mandatory exception, there are expectations from the Greek legislator regarding the amendment of the Greek Copyright Law per TDM exception. The Greek legislator should firmly resist over-regulation of TDM activity which does not prejudice the central objective of copyright, namely the provision of incentives to authors. Thus, aside from defining the notions of 'lawful access' or 'normal use' we would welcome a provision setting the TDM exception in the Greek law which allows TDM to persons, both legal entities and physical persons without discriminating against individual researchers for scientific research purpose; introduce the TDM activity as an exception to the right of reproduction as well as to the right of communication to the public of the author's work including the author's right to database and the database right-holder's sui generis right; specifically mention that works and other subject matter not protected by copyright or neighboring rights can be freely mined; and enable freely the storing and communication to the public of research files created for TDM, a.k.a. the TDM output.

28 December 2018); see also, Max Planck Institute for innovation and Competition, *Position Statement on the Proposed Modernisation of European Copyright Rules*, available at: [www.ip.mpg.de/en/research/intellectual-property-and-competitionlaw/position-statement-modernization-of-european-copyright-rules.html](http://www.ip.mpg.de/en/research/intellectual-property-and-competitionlaw/position-statement-modernization-of-european-copyright-rules.html) (last visited 28 December 2018); Geiger, above at n. 13, p. 22.

69. Geiger, above at n. 13, p. 22.

70. Geiger, above at n. 13, p.25.

71. See, NLI's selective web archive collections, available at: <https://www.nli.ie/en/udlist/web-archive-collections.aspx> (last visited 28 December 2018).

72. See, NLI's blog post on IIPC's site, available at: <https://netpreserveblog.wordpress.com> (last visited 28 December 2018).