

A European solution for Text and Data Mining in the development of creative Artificial Intelligence

With a specific focus on articles 3 and 4 of the Digital Single Market Directive

By Kristina Christensen

ABSTRACT

In today's data-driven society, Text and Data Mining (TDM) has become an essential tool in managing Big Data in its different sizes and forms, which is also an inherent part of Artificial Intelligence (AI) research. TDM techniques highly depend on datasets derived from TDM to self-learn and to make autonomous decisions. Through the lens of copyright and related rights, TDM may be used to train AI for the purpose of AI-driven creativity, where AI has already helped in completing paintings, composing music and producing movie trailers. However, since TDM typically involves the acts of copying and/or extracting of works and other subject-matter protectable by copyright and related rights, legal restrictions under the EU acquis might be in place.

In this regard, the importance of TDM has been acknowledged by the EU Legislators, which introduced two mandatory exceptions for TDM contained in articles 3 and 4 of the Directive 2019/790 on copyright and related rights in the Digital Single Market (DSM Directive). The present article analyses the relationship between TDM and AI-driven creativity by, firstly, explaining the notion of TDM and common technical steps within its process. Secondly, it examines the copyright and related rights issues regarding TDM and possible pre-existing exceptions and limitations under the EU acquis that might be applicable. Lastly, this article critically analyses the mandatory TDM provisions under the DSM Directive and concludes that these still contain shortcomings that may significantly restrict the possibility to undertake unlicensed TDM for AI creative purposes within the EU.

(EU) – often referred to as the Fourth Industrial Revolution – highly depends on the value created by data. Every day new data is produced by the quintillions of bytes, and it is estimated that by 2023 the world will be populated by 29 billion smart connected devices that are capable of collecting and sharing data in real time and making autonomous decisions.¹ Nevertheless, the increasing availability of data is the key driver to the existing growth of AI.² Consequently, adapting within this magma of online information has become a challenging but vital task – leading to difficulties in regulating this new environment.³

As a matter of fact, the value of data does not lie in the data or text taken separately, but rather in the extraction of value.⁴ This requires an analysis of the large volumes of digital text and/or data to enable the discovery of new patterns and relations. While such analysis is nearly impossible to perform manually, TDM techniques allow this to be performed easily.⁵ Generally, TDM can be described as automated computational analysis of large amounts of information in digital form, including data, images, text and sound contained in Big Data, to gain new knowledge and uncover patterns, tendencies, and correlations.⁶

Through the lens of copyright and related rights, TDM may be used to train AI for the purpose of AI-driven creativity.⁷ Accordingly, AI has already helped to actualize paintings such as ‘The Next Rembrandt’; to compose music in the style of ‘The Beatles’; and to produce a movie trailer for the film ‘Morgan’ – just to mention a few.⁸ In this regard, through (i) access, (ii) copying and/or extraction, and (iii) mining of the material, TDM generates robust and varied data sets that are further used to feed and train AI for creative purposes. As a result, there is a tension between TDM techniques and IP protection, since works or subject-matter used during TDM process may be protected under Directive 2001/29/EC (InfoSoc Directive)⁹, Directive 2009/24/EC (Software Directive)¹⁰ or Directive 96/9/EC (Database Directive)¹¹, where such use requires authorization from the relevant rightsholder.¹²

As a matter of fact, the unlicensed uses of TDM may be covered by the exceptions and limitations under the EU *acquis*, where the authorization from the rightsholders is generally not required. In this regard, several Member States within the EU (post-UK, France, Estonia and Germany) have already implemented specific TDM excep-

1. INTRODUCTION

The current and future sustainable economic development and its innovative environment within the European Union

tions into their national laws; each of those having different characteristics. However, as digital technologies permit new types of uses, supplemented by divergent national implementations of the exceptions and limitations, it remains unclear whether these provide sufficient space for enabling TDM.¹³ To solve legal uncertainties and to compete with the legal systems that offer a more friendly environment for TDM (such as Japan, UK and US), the EU adopted the Directive 2019/790 (DSM Directive) comprising two mandatory TDM exceptions; article 3 as a specific scientific research exception for non-commercial purposes and article 4 as a general exception or limitation, that may cover commercial data analytics and AI.

The aim of this article is to analyse to what extent a use of protected work or subject-matter for TDM purposes in the field of AI-driven creativity is controlled by the exclusive rights of the relevant rightsholder. The analysis will be done on the basis of EU copyright and related rights, by examining the exclusive right of reproduction and sui generis database right as well as possibly applicable pre-existing exceptions and limitations under the Infosoc, Database and Software Directives. In addition to that, an assessment of whether the newly introduced mandatory TDM exceptions contained in articles 3 and 4 of the DSM Directive are sufficient to risk-freeing unlicensed uses of TDM and thus unlocking new business opportunities for AI innovators within the EU.

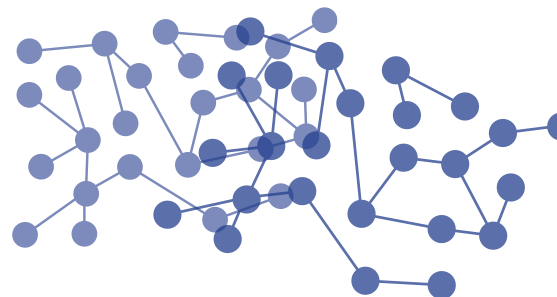
This article is structured as follows: section (ii) examines the technicalities of TDM and the copyright and related rights issues that become relevant when performing this activities; section (iii) examines whether TDM may be covered by pre-existing exceptions and limitations that

are available under the EU acquis and presents an overview of the national TDM exceptions to demonstrate the divergent implementation of these, due to the non-mandatory character of the research exception that constitutes a legal basis for the transposition; section (iv) examines the mandatory TDM exceptions introduced in the DSM Directive, by considering the rationales for such exceptions and the positive and negative impact of these on TDM when it is used for the purpose of AI-driven creativity.

2. LEGAL BARRIERS TO TEXT AND DATA MINING IN THE EU

2.1 Definition of Text and Data Mining

Article 2 of the DSM Directive defines TDM as ‘any automated analytical technique aimed at analysing text and data in digital form in order to generate information, which includes but is not limited to pattern, trends and correlations’, whereas the digital information can constitute ‘text, sounds, images or data’.¹⁴ In other words, TDM involves the deployment of automated software tools, that enable everyone with the right level of knowledge to accumulate massive quantities of text and data, such as Big Data, to further uncover new insights and patterns.¹⁵



¹ Yann Meniere et. al., *Patents and the Fourth Industrial Revolution – The global technology trends enabling the data-driven economy* (2020), European Patent Office, page 7 and 14 [http://documents.epo.org/projects/babylon/eponet.nsf/0/06E4D8F7A2D6C2E-1C125863900517B888/\\$File/patents_and_the_fourth_industrial_revolution_study_2020_en.pdf](http://documents.epo.org/projects/babylon/eponet.nsf/0/06E4D8F7A2D6C2E-1C125863900517B888/$File/patents_and_the_fourth_industrial_revolution_study_2020_en.pdf) accessed 10 October 2021.

² European Commission, *White Paper on Artificial Intelligence: a European approach to excellence and trust*, Brussels, 19.2.2020 COM(2020) 65 final, page 1.

³ Christophe Geiger et. al., *Text and Data Mining in the Proposed Copyright Reform: Making the EU ready for an age of Big Data?* (2019), *International Review of Intellectual Property and Competition Law*, Volume 49, Issue 7, page 815.

⁴ Eleonora Rosati, *An EU text and data mining exception for the few: would it make sense?*, *Journal of Intellectual Property Law & Practice*, 2019, Volume 13, Issue 6, page 429.

⁵ Eleonora Rosati, *The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market: Technical Aspects, Briefing requested by the JURI committee, Policy Department for Citizens' Rights and Constitutional Affairs*, European Parliament, page 2.

⁶ Recital 8 and Article 2(2) Directive 2019/790 of

the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, OJ L 130, 17.5.2019, p. 92-125.

⁷ Eleonora Rosati, *Copyright as an obstacle or an enabler? A European perspective on text and data mining and its role in the development of AI creativity* (2019), *Asia Pacific Law Review*, Volume 27, Issue 2, page 198-199.

⁸ Steve Schlackman, *Who holds the Copyright in AI created art?* (2020), *Artpreneur* <https://alj.artpreneur.com/the-next-rembrandt-who-holds-the-copyright-in-computer-generated-art/> accessed 10 October 2021; *Flow Records, About Hello World* (2017) <https://www.helloworldalbum.net/#> accessed 10 October 2021; John R. Smith, *IBM Research takes Watson to Hollywood with the first "Cognitive Movie Trailer"* (2016), *IBM* <https://www.ibm.com/blogs/think/2016/08/cognitive-movie-trailer/> accessed 10 October 2021.

⁹ 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, OJ L 167, 22.6.2001, p 10-19 (InfoSoc Directive).

¹⁰ Directive 2009/24/EC of the European

Parliament and of the Council of 23 April 2009 on the legal protection of computer programs (Codified version) OJ L 111, 5.5.2009, p. 16-22 (Software Directive).

¹¹ Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, OJ L 77, 27.3.1996, p. 20-28 (Database Directive).

¹² Christophe Geiger et. al. (n 3), page 5; WIPO Conversation on Intellectual Property (IP) and Artificial Intelligence (AI), Draft issues paper on intellectual property policy and artificial intelligence – prepared by the WIPO Secretariat, Second Session, WIPO/IP/AI/GE/20/1, 1 May 2021, page 5

¹³ Recital 5 and recital 19 DSM Directive; See further European Commission, *Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market*, Brussels 14.9.2016, COM(2016) 0593 final, 2016/0280 (COD), page 2.

¹⁴ See further recital 8 DSM Directive.

¹⁵ Ian Hargreaves, *Digital Opportunity – A Review of Intellectual Property and Growth* (2011), Department for Business, Innovation and Skills, page 10 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/32563/ipreview-finalreport.pdf accessed 11 October 2021.

To examine what legal issues may arise in TDM process, it is important to grasp how it operates. TDM involves a chain of activities that needs to be performed to cost-effectively mine a large amount of text and/or data. In this regard, there are three common – but not necessarily required – steps, which include (i) accessing the input material to be analysed, such as work or data collected individually or organized in a database; (ii) copying substantial quantities of the material and/or extracting the data, which may also include (a) pre-processing of the material by turning it into a machine-readable format, and (b) uploading of the pre-processed content on a platform; and (iii) mining the data and recombining it to discover new knowledge and patterns into the final output.¹⁶

2.2 The reproduction right: authorial works and expressive subject-matter

One of the central principles of copyright is that it only protects authorial works that are sufficiently original, in the sense of being ‘author’s own intellectual creation’. The EU *acquis* does not provide an exhaustive list of original works, but it can be anything from books, music, paintings and photographs to databases and computer programs.¹⁷ In addition to copyright subsisting in original works, the EU legislation expressly requires Member States to protect a closed list of unoriginal expressive subject-matter by related rights, aiming to protect the economic and legal interests of certain persons or entities that have contributed with financial, organizational or creative resources to the production of that subject-matter. For instance, rights that might be at stake when TDM is used for the purpose of AI-driven creativity is article 2(c) of the InfoSoc Directive that protects certain rights of phonogram producers in respect of their published and unpublished phonograms and article 15 of DSM Directive covering the rights of publishers in relation to their press publications.¹⁸

Once a created work can be considered as original or a subject-matter that expresses enough creativity, the rightsholders are granted the exclusive right of reproduction.¹⁹ This right is defined in article 2 of the InfoSoc Directive – supplemented for original computer programs in article 4(1)(a) of the Software Directive, original databases in article 3(1) of the Database Directive and non-original press publications in article 15 of DSM Directive – and provide rightsholders with the exclusive right to authorize or prohibit ‘direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part’ of their works or subject-matter.²⁰ Indeed, article 2 of the InfoSoc Directive introduces a broad definition of acts that may be covered by the reproduction right, alongside with the CJEU’s broad interpretation of this concept aiming to ensure legal certainty within the internal market.²¹

2.3 The sui generis database right: databases

In addition to copyright protection of databases as original works, the non-original part of a database can also be protected under the *sui generis* database right. Accordingly, the maker of the database that has made qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents of the database is entitled to object to the extraction and/or re-utilization of all or substantial part of the contents of that database, evaluated qualitatively (the scale of investment) and/or quantitatively (the volume of data).²² The CJEU interpreted the rights of extraction and re-utilization broadly, by understanding the former as temporary or permanent transfer (copying) from one medium to another by any means or in any form, and the latter as any form of making available to the public.²³ In addition, article 7(5) of the Database Directive entitles the maker of the database to prevent repeated and systematic extractions and/or re-utilization of ‘insubstantial’ parts of the

¹⁶ Eleonora Rosati (n 7), page 203–204; See also Jean-Paul Triaille et. al., Study on the legal framework of text and data mining (TDM) (2014), De Wolf & Partners, Funded by European Commission, European Union, page 28.

¹⁷ See article 2 Berne Convention for the Protection of Literary and Artistic Works of September 9, 1886 for the example of works protectable by copyright.

¹⁸ See Article 3 International Convention for the Protection of Performers, Producers of phonograms and Broadcasting Organizations, Done at Rome on October 26, 1966, where phonograms are defined as ‘any exclusively aural fixation of sounds of a performance or of other sounds’; See also article 2 InfoSoc Directive and article 7 Database Directive for the protection of databases with *sui generis* database right; See also article 15 DSM Directive.

¹⁹ European Commission Green Paper of 27 July 1995 on Copyright and Related Rights in the

Information Society COM(95) 382 final – Not published in the Official Journal.

²⁰ This is also consistent with the wording of the reproduction right itself, following article 9 Berne Convention – ‘in any manner or form’.

²¹ Recital 21 InfoSoc Directive; Judgement of 16 July 2009, Infopaq, C-5/08, EU:C:2009:465, para 43; Judgement of 1 December 2011, Painer, C-145/10, EU:C:2011:798, para 96.

²² Article 7 Database Directive; See also the British decision in the British Horseracing Board Limited & Ors v William Hill Organisation Ltd [2005] RPC 35, [2005] ECDR 28, confirms that without further verification of the content, the existing material in the database is not protected by *sui generis* right; See also Judgement of 9 November 2004, The British Horseracing Board Ltd and Others v. William Hill Organization Ltd, C-203/02, EU:C:2004:695, para 70–71.

²³ Ibid, para 51; Judgement of 18 October 2012, Football Dataco Ltd and Others v. Sportradar

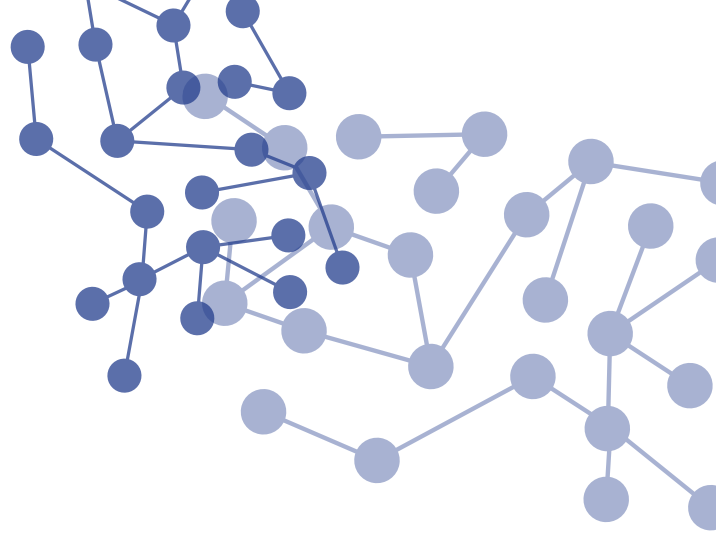
GmbH and Others, C-173/11, EU:C:2012:642, para 20–21; See also Judgement of 5 March 2009, Apis Hristovich EOOD v. Lakorda AD, C-545/07, EU:C:2009:132; Also confirmed by the Advocate General Stix-Hackl in Opinion of Advocate General delivered on 8 June 2004, BHB v. WH, C-203/02, EU:C:2004:695.

²⁴ BHB v. WH, C-203/02, para 86.

²⁵ Federico Ferri, The dark side(s) of the EU Directive on copyright and related rights in the Digital Single Market (2020), China EU Law Journal, Department of Legal Studies, University of Bologna, page 11 <https://doi.org/10.1007/s12689-020-00089-5> accessed 13 October 2021.

²⁶ Eleonora Rosati (n 7), page 200; See also Eleonora Rosati, (n 5), page 5.

²⁷ Recital 33 InfoSoc Directive; See further CJEU, Judgement of 26 April 2017, Stitching Brein v Jack Frederik Wullems, C-527/15, EU:C:2017:300, para 65 and 69, where the CJEU examined the ‘lawful use’ in article 5 (1)



contents of the database, where the unauthorized acts would seriously prejudice the investment of the database maker.²⁴

In the context like the one presented above, the EU *acquis* establishes a two-tier system for the protection of expressive subject-matter that involves copyright protection for original works and related rights protection for their non-authorial counterparts.²⁵ This means that whenever TDM is used to mine Big Data containing protectable works or other subject-matter – also included in a database – for the purpose of AI-driven creativity, both copyright and related rights may become relevant. For instance, if an AI developer wishes to train AI system to create songs, that developer not only have to consider the author's exclusive right of reproduction of the authorial text, but also the reproduction right of a phonogram producer that first fixes the sounds and the database maker's exclusive right of extraction that stores these songs in its database.

2.4 Text and Data Mining: An apt technique encroaching the exclusive rights?

2.4.1 Text and Data Mining as an act of reproduction?

Given the broad scope of the right of reproduction as well as the extraction and re-utilization, the following questions arise: can protected works and subject-matter be used for AI creative purposes within the TDM context without falling within the scope of the exclusive rights? First of all, it must be noted that not all TDM activities involve copying and/or extraction of the material at the outset, which mostly depend on the use of the material, technical tools and the extent of the mining procedure.²⁶ Nor are all acts of copying subject to prior authorization, e.g., when such acts fall within the scope of the exceptions and limitations under the EU *acquis*.²⁷

Additionally, TDM carried out on mere information, facts or data does not amount to copyright or related rights infringement.²⁸ However, even if the source of data

used for TDM is protected, the threshold for infringement may not be met if TDM reproduces only parts of the work or subject-matter so minimal that it falls below the threshold for protection.²⁹ This was indeed confirmed by the CJEU in *Infopaq I*, C-5/08 where it held that words considered in isolation are not *per se* an intellectual creation and cannot be protected as such.³⁰ Also, as has been underlined by numerous scholars, the act of reading a work by computers is random access memory does not result in copyright infringement.³¹ Thus, putting it in the context of TDM, the 'right to read is the right to mine'.³²

Conversely, whenever TDM techniques involve copying and/or extraction of the material relevant for AI project, then legal restrictions may be in place.³³ In this regard, the CJEU has in *Infopaq I*, C-5/08 confirmed that at least '11 consecutive words' contained in a newspaper constitute an approximate threshold for originality that may be applied *mutatis mutandis* in respect of all authorial works, including computer programs and databases.³⁴ Within this context, since AI relies on processing masses of data sets stemmed from TDM, especially in cases when TDM is carried out on Big Data containing protectable works, the likelihood for copyright infringement exists.³⁵

InfoSoc Directive and confirmed that this exception cannot be relied upon by users where the pre-installed add-ons allow access to private servers on which copyright-protected works have been made available to the public without the consent of the rightholder. See further Judgement of 4 October 2011, *Football Association Premier League Ltd and others v. QC Leisure and others v. Media Protection Services Ltd*, C-403/08, EU:C:2011:631, para 168; Judgement of 17 January 2012, *Infopaq International A/S v. Danske Dagblades Forening*, C-302/10, EU:C:2012:16 (Infopaq II), para 42; Judgement of 5 March 2015, *Copydan Båndkopi v. Nokia Danmark*, C-463/12, EU:C:2015:144, para 79.

²⁸ This is also confirmed by the recital 9 DSM Directive.

²⁹ However, the CJEU has ruled in *Ryanair Ltd v. PR Aviation BV* that the absence of copyright, related rights or sui generis database right protection does not exclude the possibility of

the rightholder to impose restrictions through contractual provisions (also covering TDM activities), See judgement of 15 January 2015, *Ryanair Ltd v. PR Aviation BV*, C-30/14, ECLI:EU:C:2015:10.

³⁰ Judgement of 16 July 2009, *Infopaq International A/S v. Danske Dagblades Forening*, C-5/08, ECLI:EU:C:2009:465 (Infopaq I), para 45-46; See also reasoning by Christophe Geiger et. al. (n 3), page 818.

³¹ See also Peter Murray-Rust, *The right to read is the right to mine* [2012] <https://blog.okfn.org/2012/06/01/the-right-to-read-is-the-right-to-mine/> accessed 14 October 2021.

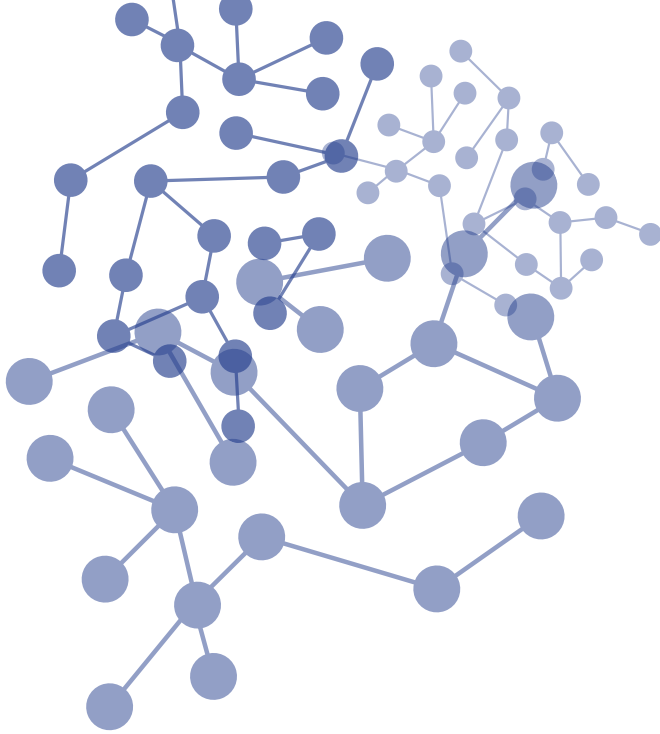
³² Sean Flynn, *Implementing user rights in the field of Artificial Intelligence: A call for international action* (2020), *European Intellectual Property Review*, Issue 7, WCL Research Paper No. 2020-12 page 4.

³³ Eleonora Rosati (n 7), page 206-209.

³⁴ A similar reasoning was also conducted by the CJEU in *FAPL*, C-403/08, para 159, where it

held that reproducing approximately four audio or video fragments that constitute authorial works embedded within the films constitute and falls within the exclusive rights of reproduction of the relevant rightholders; See further Kim Martineau et. al., *Towards artificial intelligence that learns to write code* (2014), *Massachusetts Institute of Technology* <https://news.mit.edu/2019/toward-artificial-intelligence-that-learns-to-write-code-0614> accessed 14 October 2021.

³⁵ See also Judgement of 2 May 2012, *SAS Institute Inc. v. World Programming Ltd*, C-406/10, ECLI:EU:C:2012:259, para 66-67, where the CJEU confirm the test in *Infopaq* in relation to reproduction of computer programs.



In case where the content copied and/or extracted is contained in a database, then both copyright and *sui generis* database right might come into consideration. With regard to copyright, TDM may infringe the reproduction right of the author of the database, e.g. when pre-processing for the extraction, meaning that the original selection and/or the arrangement of the database is copied in its entirety.³⁶ However, when a huge amount of data is being analyzed, indexed, aggregated, and merged during TDM process, it can be difficult to prove that the data comes from a specific database and, in particular, that TDM infringes upon the selection or arrangement of that database protected by copyright. Consequently, the possibility of the author of a database to claim and prove infringement can be highly unlikely – yet not impossible.³⁷

In so far as related rights are concerned, the CJEU has in *Pelham*, C-476/17 clarified that phonogram producer can prevent a reproduction of even a very short sound sample, for instance approximately 2-seconds rhythm sequence, when it is taken for the purpose of including that sample in another phonogram, unless that sample is modified to the extent that the final result is unrecognizable to the ear.³⁸ Thus, the key criterion for related rights is not ‘originality’ but rather ‘recognizability’, which means that even small pieces of a larger work are able to attract their own related rights protection.³⁹ Consequently, since TDM may involve reproduction that results in the creation of a copy of the protected material, without possible selection of certain pieces from that material during TDM process that may fall below the threshold for recognizability or additional modification of the material *per se*, it is certain that TDM will infringe related rights of the relevant rightsholder.

In sum, since TDM is a copy-reliant technology it becomes qualitatively and quantitatively relevant to the realm of copyright and related rights. This means that any digital copies made out of works protectable by copyright or expressive subject-matter protectable by related rights during TDM processes for the purpose of AI-driven creativity – irrespective of how transient or short from an eco-

nomical perspective these may be – have the potential to meet the threshold infringement for copyright (originality as being ‘author’s own intellectual creation’) and related rights (‘recognizability’), since the main activity of TDM is to copy the content in its entirety without adding or altering it *per se*, and thus will always require prior authorization of the rightsholders.

2.4.2 Text and Data Mining as an act of extraction and/or re-utilization?

Turning to the *sui generis* database right, TDM might infringe the extraction – and to some degree the re-utilization – of a substantial part of the contents of a database, when processing Big Data for AI-driven creativity.⁴⁰ In this regard, it is no coincidence that the notion of TDM often assimilates with the ‘extracting data and/or information’.⁴¹ In fact, TDM techniques identify and collect pre-existing works or subject-matter from different databases in accordance with the relevance to each AI project, where they are eventually stored in one or more servers or other tangible mediums accessible to the programmers.⁴² Even if an extraction occurred without reproduction of the original materials, the extraction *per se* would infringe the exclusive right of the database owner. In this regard, the CJEU has in *BHB v. WH*, C-203/02 confirmed that temporary or permanent transfer of data from one medium to a new one and storage thereof is sufficient to be considered as an extraction, meaning that TDM will fall within the scope of this right as this operation constitutes a necessary step in its process.⁴³

3. PRE-EXISTING LEGAL REGIME APPLICABLE TO TEXT AND DATA MINING

3.1 Exceptions and limitations – A European perspective

Given the broad scope of the exclusive rights, reproductions and extractions made during TDM process for AI creativity would always need to be authorized by the relevant rightsholder. However, such authorization is not required where TDM may be eligible for protection under the mandatory and non-mandatory pre-existing exceptions and limitations contained in the EU *acquis*.⁴⁴ Although the nature and scope of the exceptions and limitations are governed by the domestic laws of individual Member States, these are subject to harmonised EU regulations and should be interpreted strictly.⁴⁵ Thus, several pre-existing exceptions and limitations have been selected as possible candidates to screen unauthorized use of TDM from copyright and related rights infringement.

3.1.1 Temporary act of reproduction

According to recital 9 of the DSM Directive, the mandatory exception in article 5(1) of the InfoSoc Directive still applies to TDM techniques, insofar these involve the making of temporary reproductions that are ‘transient or incidental’ to an integral part of a technological process that enables a ‘lawful’ use of a work with ‘no independent economic significance’ – these must be cumulatively met in accordance with the restrictive interpretation.⁴⁶ As a guideline, the CJEU has in its case law confirmed that ‘transient or

incidental' requires an act to be limited in its duration to what is necessary for the proper completion of the technological process, meaning that after the completion of the process such copies must be deleted.⁴⁷ Additionally, independent economic significance occurs when reproduction generates an additional economic advantage beyond the advantage derived from the lawful use of the protected material or if the reproduction leads to a modification *per se*.⁴⁸

Applying the cumulative conditions to TDM, it may seem that these would not be easily met during its process.⁴⁹ Firstly, copies made during TDM are in most cases not 'transient or incidental' but rather permanent, due to the fact that initial intention of the reproduction is to keep such copies for a longer period of time in order for these to be pre-processed, uploaded into a medium and mined for the development and the training of creative AI. In fact, TDM techniques, if any, do not involve manually activated or automatically performed deletion process of created copies, which highly depend on the will of the ones performing TDM and the subsequent users of the material in accordance with their needs, such as AI developers.⁵⁰

Secondly, according to recital 33 of the InfoSoc Directive, a use is considered lawful when it is authorized by the rightsholder (explicitly or implicitly) or it is not restricted by law.⁵¹ However, since TDM as such encroaches on the exclusive right of reproduction, which is an act prohibited by law, every copy made during TDM process for the pur-

pose of AI-driven creativity needs to be authorized by the relevant rightsholder. This in fact means that it would be impossible to conduct TDM if authorization or license work by work is required, which would also defeat the purpose of TDM to avoid a time-consuming process.

Lastly, the independent economic significance cannot be overlooked since it is highly relevant to the economic value of the TDM as such. Potentially, the results gained from TDM and the sharing of the final outputs, especially in cases of Big Data and AI creativity, are all steps leading to the financial reward.⁵² However, even if an independent economic significance would not be demonstrated, the content could still be modified during the TDM process, for instance when the material is aggregated or altered or also in cases when it is normalized from unstructured data into structured data in order for it to be compatible with the technology to be deployed for TDM.

As a matter of fact, even though being mandatory, exception in article 5(1) of the InfoSoc Directive may be overridden by contracts, meaning that rightsholders can rely on contractual provisions to restrict reliance on it by AI developers.⁵³ Consequently, it is evident that this sole mandatory exception has a limited scope and lacks legal certainty with regards to TDM activities and the lawfulness thereof. Consequently, copies made during TDM that fail to satisfy one of the conditions stipulated in the article voids the application of the exception, which, in fact, indicates the rightsholder's strong monopoly over its original work and expressive subject-matter.

³⁶ Jean-Paul Triaille et. al. (n 16), page 33-34; See also Eleonora Rosati (n 5), page 5-6.

³⁷ Jean-Paul Triaille et. al. (n 16), page 34, referring to the Benoit Michaux, *Droit des bases de données* (2005), No. 116, Bruxelles, Kluwer, page 119.

³⁸ Judgement of 29 July 2019, *Pelham GmbH and others v. Ralf Hütter and Others* (Pelham), C-476/17, EU:C:2019:624, para 31 and 39.

³⁹ Eleonora Rosati (n 7), page 206; Christophe Geiger et. al. (n 3), page 6.

⁴⁰ See further Marco Caspers et. al. *Baseline report of policies and barriers of TDM in Europe* (2016). In *Reducing barriers and increasing uptake of Text and Data Mining for research environments using a collaborative knowledge and open information approach, FutureTDM, Horizon 2020, GARRI-3-2014*, page 22, arguing that since it refers to any act of making available to the public and since the contents of the database are only used to be 'read' by TDM, there may not be any actual disclosure to the researchers themselves, let alone a public.

⁴¹ Jean-Paul Triaille et. al. (n 16), page 38.

⁴² Theodoros Chinou, *Copyright lessons on Machine Learning: what impact on algorithmic art?* 10 (2019), *JIPITEC*, page 402 https://www.jipitec.eu/issues/jipitec-10-3-2019/5025/chinou_pdf.pdf accessed

15 October 2021.

⁴³ *BHB v. WH*, C-203/02, para 65-66; Moreover, TDM may adapt or translate the content, e.g. converse to a different format, which may also fall within the scope of extraction right. Even if not discussed, it is important step during TDM; See also Jean-Paul Triaille et. al. (n 16), page 38-39, stating that it may not be ruled out that TDM can also copy and/or extract elements that are so small that these can be considered as insubstantial.

⁴⁴ Recital 1 and 3 DSM Directive and recital 31 InfoSoc Directive; See also Theodoros Chiou (n 42), page 405; See also Eleonora Rosati (n 7), page 206; Christophe Geiger et. al. (n 3), page 820.

⁴⁵ *Infopaq I*, C-5/08, para 56; This is also supported by the three-step-test under article 9 (2) Berne Convention.

⁴⁶ Recital 9 DSM Directive; See further article 5 (1) InfoSoc Directive; See also *Infopaq I*, C-5/08 para 55-58, where the CJEU held that the provision that derogates from the general principle must be interpreted strictly; See also *Infopaq II*, C-302/10, para 26; *FALP*, C-403/08, para 162; Judgement of 5 June 2014, *Public Relations Consultants Association v Newspaper Licensing Agency Ltd and Others* (PRCA), C-360/13, ECLI:EU:C:2014:1195, para 24; See also Article 5 (5) InfoSoc Directive for

the three-step-test to ensure legal certainty.

⁴⁷ *Infopaq I*, C-5/08 para, para 33, 64, where the CJEU required human intervention when deleting reproduced copies, where this requirement was further reconsidered and removed in *Infopaq II*, C-302/10, para 32, 36, 39 and further confirmed in *PRCA*, C-360/13, para 15.

⁴⁸ *FALP*, C-403/08, para 177; *Infopaq II*, C-302/10, para 51-53.

⁴⁹ Christophe Geiger et. al. (n 3), page 9-10.

⁵⁰ See further Jean-Paul Triaille et. al. (16), page 46; See also Theodoros Chiou (n 42), page 406 and Mark A. Lemley et. al., *Fair Learning* (2020), Vol. 6, No. 11, page 120-121 <https://ssrn.com/abstract=3528447> accessed 16 October 2021.

⁵¹ *Infopaq II*, C-302/10, para 44, where the CJEU stated that where an act of reproduction is not restricted by law (in this case Danish law or EU law) the authorization from the rightsholder is not required; See also *FALP*, C403/08, para 169-171; See *Stichting Brein v Jack Frederik Wullems*, C-527/15, para 65-68.

⁵² Jean-Paul Triaille et. al. (16), page 47; See also Theodoros Chiou (n 42), page 406.

⁵³ Since InfoSoc Directive does not provide any expressed provisions for contractual override.

3.1.2 Scientific research

Another potential candidate to screen TDM from the copyright and related rights infringement is the non-mandatory exception contained in article 5(3)(a) of the InfoSoc Directive, which also constitutes a legal basis used by the Member States for the introduction of specific TDM exceptions into their national laws. Accordingly, this exception applies to reproductions with non-commercial purposes and which have as its sole purpose the illustration for teaching or scientific research, where the source, including the author's name, must be indicated unless it turns out to be impossible. Applying this exception to TDM, several observations can be made.

Firstly, TDM techniques used for the purpose of AI-driven creativity may simultaneously have other purposes than scientific research. Secondly, since TDM is generally associated with quantity, where it may involve copying and/or extraction of thousands, if not millions, of protectable works or subject-matter, the obligation to indicate the source, including author's name, would make it nearly impossible for AI developers that are using TDM techniques to fulfill this requirement and may discourage them from using TDM as a research tool.⁵⁴

Thirdly, the key restriction of the 'non-commercial purpose' seems wholly misaligned to the modern realities of academia, because most of the universities and researchers are striving to obtain funding and budget from private entities in order to carry out most promising projects, not least with regards to AI-driven creativity.⁵⁵ As a matter of fact, this requirement may also bring evidential problems, where the results of research unintendedly turn out to be commercially valuable or where commercial TDM may *per se* qualifies as a 'scientific research'.⁵⁶

As a matter of fact, this research exception can arguably cover TDM activities for the purpose of training AI for creative purposes, based on the framework of human teaching so as to fall within the scope of 'illustration for teaching', but once again this would only be done for the non-commercial purpose and still requires AI developers

to make an effort in tracing all authors and sources of each work or subject-matter to be mined. Thus, this exception is difficult to apply in practice and would permit undertaking of unlicensed TDM only in few cases.

A similar research exception is available under article 6(2)(b) of the Database Directive, which applies to the selection and arrangement of a database. Accordingly, all limitations described in relation to the research exception under the InfoSoc Directive would also apply to databases protectable by copyright, which would equally not be sufficient to cover unlicensed TDM. In fact, article 6(2)(b) imposes the obligation to indicate the source of the database but does not provide for a safeguard clause if 'it turns out to be impossible', which makes it even harder for AI developers to comply with. The difference is more a declamation than a substantial matter because it is the general principle of law that it can never oblige anyone to do the impossible (*impossibilium nulla est obligatio*).⁵⁷

A research exception is also provided for the sui generis database right in article 9(b) of the Database Directive that only covers acts of reproduction made by a 'lawful user' who, according to the CJEU, is a user having lawful access to the contents of a database, e.g. through licensing agreement, or relying upon exceptions by law or contract.⁵⁸ In fact, contrary to the research exception for copyright, article 9(b) of the Database Directive does not include the adjective 'sole' in relation to the purpose; TDM remains within the scope of the exception even if it is conducted partially also for other purposes. Once again, the condition of attribution puts a heavy burden on AI developers as well as requirement of being a lawful user, making this exception difficult to apply in cases of unlicensed uses of TDM for AI creative purposes.

3.1.3 Normal use of the structure of a database

A possible candidate for serving as an exception for TDM is the so called 'normal use of a database' contained in article 6 (1) of the Database Directive, which is the only mandatory exception under that Directive and cannot be

⁵⁴ Maria Bottis et. al., Text and Data Mining in the EU 'Acquis Communautaire' tinkering with TDM & Digital Legality Deposit (2019), No. 2, Erasmus Law Review, page 192 <http://www.erasmuslawreview.nl/tijdschrift/ELR/2019/2/ELR-D-19-00024.pdf> accessed 16 October 2021; See also reasoning conducted by Marco Caspers et. al. (n 40), page 29.

⁵⁵ Rossana Ducato et. al., Limitations to Text and Data Mining and Consumer Empowerment – Making the Case for a Right to "Machine Legibility" (2018), CRIDES Working Paper Series, page 19 https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3278901 accessed 17 October 2021.

⁵⁶ Marco Caspers et. al. (n 40), page 31; See further Jean-Paul Triaille et. al. (n 16), page 63-64.

⁵⁷ Michael Walter et. al., European Copyright Law – A commentary (2010), First Edition, Oxford University Press, page 1042; See also Rossana Ducato et. al. (n 55), page 11

⁵⁸ BHB v. WH, C-203/02, para 58; See Judgement of 3 July 2013, UsedSoft GmbH v Oracle International Corp., (UsedSoft), C-128/11, EU:C:2012:407, para 85.

⁵⁹ Article 15 Database Directive.

⁶⁰ Jean-Paul Triaille et. al. (n 16), page 72-73; In national proceedings of the Ryanair case, the Netherlands court found that the online intermediary comparing prices of flight, including the extraction of information from the Ryanair website, did constitute a normal use of that database. See Ryanair, C-30/14, para 21.

⁶¹ Jean-Paul Triaille et. al. (n 16), page 75-76; See also reasoning from Christophe Geiger et. al. (n 3), page 824

⁶² Jean-Paul Triaille et. al. (n 16), page 77

⁶³ BHB v. WH, C-203/02, para 69 and 73.

⁶⁴ Estelle Derclaye, The legal protection of Databases: A comparative analysis (2008), Edward Elgar, page 111.

⁶⁵ BHB v. WH, C-203/02, para 86 and 89-90.

⁶⁶ See further Rossana Ducato et. al. (n 55), page 14.

⁶⁷ Recital 14 and 15, Article 5(3) Database Directive.

⁶⁸ Lucie Guibault, Blogpoll: towards a Text & Data Mining exception in EU copyright law? (2015), Kluwer Copyright Blog <http://copyrightblog.kluweriplaw.com/2015/09/07/blogpoll-towards-a-text-data-mining-exception-in-eu-copyright-law/> accessed 17 October 2021.

⁶⁹ Ibid; See also Jean-Paul Triaille et. al. (n 16), page 109; See further Marco Caspers et. al., A right to 'read' for machines: Assessing a black-box analysis exception for data mining (2016), Proceedings of the Association for Information Science and Technology, Computer Science, Volume 53, Issue 1, p. 1-15.

waived by contractual provisions.⁵⁹ It permits a lawful user to carry out the act of reproduction of a database without prior authorization of the rightsholder, if the act is ‘necessary’ for accessing the contents of a database and making ‘normal use’ of them. As mentioned previously, a lawful user is the one who can either invoke a contractual authorization, e.g. through licensing agreement, or a legal or contractual exception.

As regards the condition of ‘normal use’, recital 34 of the Database Directive can be used as a guideline: ‘lawful user must be able to access and use the database for the purpose and in the way set out in the agreement with the rightsholder’. This certainly means that a ‘normal use’ considers the ‘purpose’ and the ‘way of access and use’ specifically set out in the agreement, meaning that the database must only be used for the specific purpose provided by the rightsholder. As a matter of fact, the agreement can limit the purpose and modalities of access, also including TDM, or not explicitly address the uses for the benefit of the rightsholder.⁶⁰

All in all, since the main purpose of TDM is to extract new patterns between previously unrelated pieces of information and to get new insights by mining large number of databases, especially when it is used for the training of creative AI, this would neither be considered as ‘necessary’ to access the contents and to use it in a normal manner nor would the aim of TDM normally be the purpose in the context of a ‘normal use’ of a database.⁶¹ Consequently, the exception in article 6(1) of the Database Directive does not give much room for unlicensed TDM and constitutes a remarkable obstacle for AI developers to provide such type of activities.

3.1.4 Extraction and/or re-utilization of insubstantial parts

Another exception that may be relevant for TDM is the mandatory exception in article 8(1) of the Database Directive, which, according to article 15, cannot be overridden by contracts. This exception enables a lawful user to extract and/or re-utilize insubstantial parts of a database protected by *sui generis* right, evaluated qualitatively and/or quantitatively, for any purpose whatsoever and without obtaining prior authorization from the rightsholder. This wording allows a broader interpretation compared to the notion of a ‘normal use’ under article 6(1) of the Database Directive meaning that it can possibly include unintended purposes, such as TDM.⁶²

Furthermore, the terms ‘insubstantial’, ‘qualitatively’ and ‘quantitatively’ are not defined in the Database Directive. In this regard, the CJEU has in *BHB v. WH*, C-203/02 concluded that when any part does not fall within the definition of a ‘substantial part’, evaluated quantitatively and qualitatively, it thus falls within the scope of an ‘insubstantial part’.⁶³ In a nutshell, when TDM extracts any part that does not represent the substantial investment of the database maker, the investment is not harmed and there cannot be an infringement.⁶⁴

Besides, even though article 7(5) of the Database Directive provides that repeated and systematic extraction and/or re-utilization of insubstantial part of the contents of the database are not permitted, these are, however, still

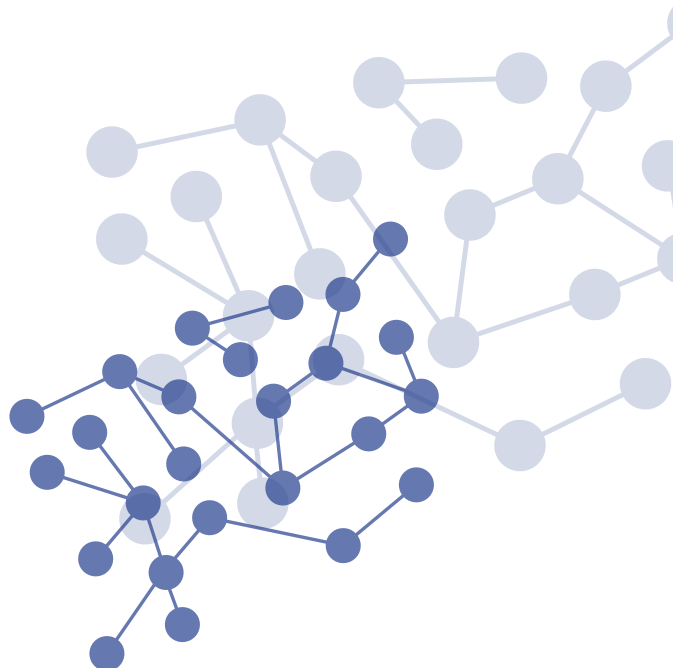
lawful when such acts do not ‘reconstitute’ the whole or substantial parts of the database and, more generally, do not harm the investment of the rightsholder.⁶⁵ In fact, since the aim of TDM is not to reconstitute the database, as required by the CJEU, but rather to identify patterns and extract knowledge, this exception provides enough ‘pass through’ for lawful users of the database with the aim of conducting TDM for the training of creative AI.

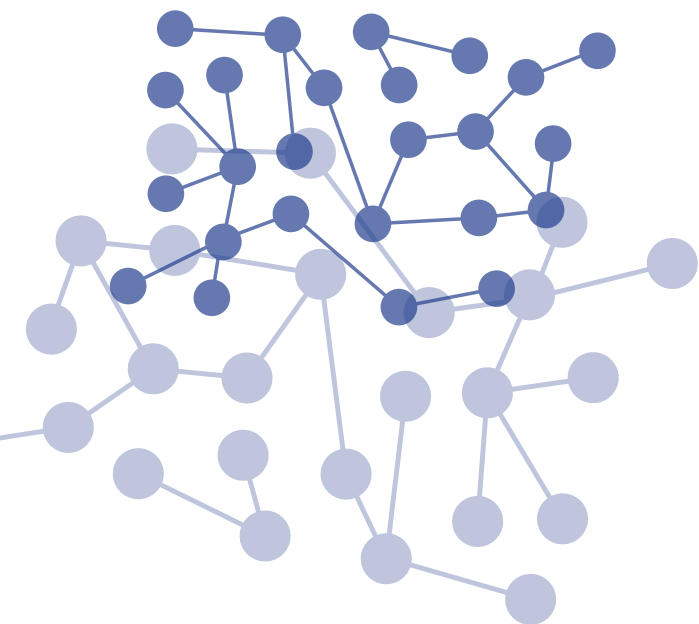
Again, as explained in conjunction with article 6(1) of the Database Directive, this exception will only apply to lawful users using TDM on databases, meaning that contractual provisions made by the rightsholders in a licensing agreement may limit or completely prohibit the uses of TDM on protectable databases.

3.1.5 Mandatory exception to computer programs

Article 5(3) of the Software Directive may also constitute a potential candidate for serving as an exception for TDM, the so-called ‘back box analysis’.⁶⁶ This exception allows the person having a right to use a computer program to observe, study or test the functioning of the program in order to determine the ideas and principles which underlie any element of the program – provided that such act does not infringe copyright in the computer program *per se*.⁶⁷ In fact, this exception does not make a distinction between commercial and non-commercial acts, which is one of a few mandatory exceptions within the EU *acquis*.⁶⁸

Since the fundamental principle of copyright is to protect the expression and not the ideas or the data contained in that expression, a parallel between the permitted acts under this exception can be drawn with the TDM activities; TDM aims at extracting new ideas from the computer program and thus reproduces it to proceed with the analysis and creation of datasets for further training of AI.⁶⁹ Therefore, in these cases the purpose is not to copy the expression of the computer program but rather to extract information from it, meaning that TDM may fall within the scope of this mandatory exception, without infringing upon the copyright in the computer program *per se*.





Remarkably, the CJEU has pointed out that the acts permitted by the exception may only be carried out within the framework of the acts permitted by the licensing agreement.⁷⁰ This means that even if contractual provisions contrary to the exception in article 5(3) of the Software Directive are null and void according to article 8 (2), it can still be possible for the rightsholders, with sufficiently careful drafting, to define the permitted usage narrowly so as to limit opportunities available for the user to engage in TDM for AI creativity while exercising its licensed rights.

In sum, the pre-existing exceptions and limitations under the EU *acquis* do not offer a steady legal framework for conducting unlicensed TDM. The cumulative conditions of the temporary acts of reproduction and the requirement of being a lawful ‘user’/‘acquirer’ make it difficult for AI developers to comply with, especially when TDM is applied on Big Data. In addition, the unharmonized EU legal framework of the research exceptions, which causes legal fragmentation due to their voluntary implementation in Member States, and the requirement of the ‘non-commercial purposes’, further constrains the potential to undertake TDM for the purpose of AI creativity. In fact, it is indeed clear that the licensed-based solutions are inadequate to allow TDM to take place, since ‘take it or leave it’ provisions in the agreement make access conditional upon accepting the rightsholders terms of use, where rightsholders may through sufficiently careful drafting put specific clauses in their licensing that rule out TDM.

3.2 Exceptions and limitations – National perspectives in the EU

Specific TDM exceptions have long been considered within the EU, due to the uncertain application of the pre-existing exceptions and limitations to TDM techniques.⁷¹ In fact, since a system resting solely on licensing agreement was insufficient to cover undertaking of TDM for the research purposes and cross-border uses of protected material, several Member States within the EU (namely UK, France, Estonia and Germany) tried to tackle the situation at the national levels by adopting TDM exceptions within the legal framework of the EU *acquis*, i.e.

article 5(3)(a) InfoSoc Directive. However, as will be seen, different national implementations of this exception to specifically cover TDM do result in a patchwork approach and create legal uncertainties for market players conducting unlicensed TDM within the EU.

Already in 2014, the UK was back then the first Member State within the EU to introduce a mandatory exception that recognizes and permits ‘text and data analysis’ or ‘computational analysis’ (prima facie covering state-of-the-art technologies, including TDM).⁷² Section 29A(1)(a) of the Copyright, Designs and Patents Act (CDPA) explicitly provides that the making of a copy of a protectable work by the person having ‘lawful access’ to that work is permitted when carrying out a computational analysis for the ‘sole purpose of research’ for a ‘non-commercial purpose’.⁷³ Even if this exception does not impose restrictions on the beneficiaries of the exception, there are elements that make it more complicated. Firstly, this exception is only limited to the right of reproduction, where copies made during TDM cannot be shared as it would amount to copyright infringement; this is a wholly undesirable outcome for AI developers. Secondly, this exception does not cover reproduction of databases protected by sui generis database right, meaning that the database maker may prevent through contractual provisions to undertake TDM acts. However, even though creating a semi-certain environment for researchers conducting computational analysis, an explicit exception is indeed a clear step towards a more favourable environment for TDM.

On 7 October 2016, by Law No. 2016-1231 for a Digital Republic (*Loi pour une République numérique*), France introduced two specific exceptions for TDM for both copyright contained in article L122-5, 10 and for sui generis database right contained in article L342-3, 5 of the Intellectual Property Code (Code de la Propriété Intellectuelle) (CPI).⁷⁴ These exceptions cover acts of reproduction made from a ‘lawful source’, e.g. material and databases made available to the public by the rightsholder, which is included in or associated with scientific publications for the purpose of public research, excluding all commercial purposes. Much alike the UK exception, these provide for demarcations that limit its applicability to a fairly large extent.⁷⁵ Firstly, the benefits derived from TDM for research purpose goes beyond the mining of merely scientific publication and writing, especially in case of AI-driven creativity.⁷⁶ Secondly, regarding the copyright exception for TDM, the requirement of ‘public research’ is unsatisfactory solution for the TDM activities carried out by private research institutions. Consequently, even though these exceptions focus on the lawfulness of the source *per se*, without requiring ‘lawful access’, the existence of restrictive aspects makes the French exceptions limited in their utility.

Another TDM exception was introduced in Estonia in 2016, which entered into force on 1 January 2017. According to § 19(3) of the Copyright Act (*Autoriõiguse seadus*), the copyright protectable works may be used without prior authorization from the rightsholder ‘for the purposes of text and data mining, provided that such use does not have a commercial purpose’.⁷⁷ In fact, even though not providing the requirement of ‘lawful access’ or similar, § 19 of the Copyright Act requires the mentioning of the

name of the author and, if possible, the name of the work and the source publication. Furthermore, much alike the UK exception, the Estonian exception covers solely the acts of reproduction of works and thus excludes the communication right, which, as stated above, is a wholly undesirable outcome for AI developers. In addition, this exception does not cover reproduction of databases protectable by *sui generis* database right, where contractual provisions restricting TDM activities may be in place.

A much braver measures were taken by the German legislator, which in 2017 introduced a specific TDM exception in article 60d of the Act on Copyright and Related Rights (*Urheberrechtsgesetz*), which entered into force on 1 March 2018, covering acts of reproduction (copyright) and acts of extraction (*sui generis database right*) for the purpose of scientific research without a commercial purpose. In fact, this exception also covers the making available of a 'corpus' (e.g. a collection of normalized, structured and categorized data) 'to a specifically limited circle of persons' (presumably research team or multi-institutional), as well as to 'individual third persons' for quality assurance.⁷⁸ Nevertheless, after the completion of TDM project, the created 'corpus' may be sent to institutions designated by law for permanent storage; all other copies must be deleted. It is worth noting that, compared to the UK, the German exception does not impose a prerequisite of 'lawful access' or requirement of being a 'lawful user', nor does it limit the source material that can be used for the purpose of TDM, for instance 'included or associated with scientific publications' as required by the French exception. While the German exception can be considered as giving a much clearer guidance for the TDM, one might wonder whether this approach remains within the limits of the EU law; Member States are free to maintain or introduce stricter rules when implementing the directive into national laws,

not vice versa. However, this discussion is outside the scope of this article and requires further considerations.

4. DIGITAL SINGLE MARKET DIRECTIVE – A STEP IN THE RIGHT DIRECTION?

4.1 From the initial proposal to the final output

In 2016, the European Commission issued a proposal for the DSM Directive, which was adopted in April 2019 and came into force on 6 June 2019, with the aim of providing greater legal certainty in the digital and cross-border environment. In fact, during the discussion leading to the proposal of the DSM Directive and its final adaptation, the European Commission had in mind to assess the overall competitiveness of the EU copyright and related rights system with regards to TDM not just internally but also vis-à-vis third countries, such as US (fair use under 17 U.S.C. § 107)⁷⁹ and Japan (reproduction for data analysis under article 47septis Japan Copyright Act)⁸⁰, which constitute EU's main trading partners that have already dealt with TDM issues in their IP regimes.⁸¹ Nearly two years have passed since the adoption of the DSM Directive, which was due to be implemented by June 7, 2021. As of November 21, 2021, only 11 Member States⁸² have implemented the Directive, where the delays may partly be explained by the disruption caused by the COVID-19 pandemic.⁸³

The journey of the DSM Directive has, however, not been easy. The proposal envisaged article 3 of the DSM Directive as the only mandatory TDM exception covering reproductions and extractions made by research institutions for the purpose of scientific research. The aim behind the limited scope of the exception was to ensure EU's competitiveness and scientific leadership, including competitiveness vis-à-vis third countries, which back then in-

⁷⁰ SAS Institute Inc., C-406/10, para 54–55.

⁷¹ European Commission, 'Licenses for Europe' stakeholder dialogue (2021) <https://digital-strategy.ec.europa.eu/en/library/licences-europe-stakeholder-dialogue> accessed 17 October 2021.

⁷² See also HM Government, The Government response to the Hargreaves review of Intellectual Property and growth (2011), The Intellectual Property Office, page 1 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/32448/11-1199-government-response-to-hargreaves-review.pdf accessed 18 October 2021.

⁷³ Regulation 3 of the Copyright and Rights in Performances (Research, Education, Libraries and Archives) Regulations 2014, No. 1372, adding Article 29A to the Copyright, Designs and Patents Act 1988. The Regulations came into force on 1 June 2014.

⁷⁴ Art. 38 of Law No. 2016-1231 for a Digital Republic added paragraph 10 to Art. L122-5 and paragraph 5 to Art. L342-3 of the Intellectual Property Code (Code de la propriété intellectuelle) (CPI).

⁷⁵ Marco Casper et. al. (n 40), page 64.

⁷⁶ Marco Caspers, Some observation of the French TDM exception (2016), Future TDM <https://www.futuretdm.eu/blog/legal-policies/some-observations-of-the-french-tdm-exception/> accessed 19 October 2021.

⁷⁷ WIPO IP Portal, Copyright Act (consolidated text of January 1, 2017), § 19. Free use of works for scientific, educational, informational and judicial purposes. For the English version see <https://wipolex.wipo.int/en/text/429284> (accessed 21 November 2021).

⁷⁸ Article 60d (1) sentence 1 and Copyright Act of 9 September 165 (Federal Law Gazette I, p. 1273), as last amended by Article 1 of the Act of 28 November 218 (Federal Law Gazette I, p. 2014) (UrhG).

⁷⁹ Section 107 in the Copyright Laws of the United States and Related Laws Contained in Title 17 of the United States Code; See further Mark A. Lemley et. al., Fair Learning (2020), Vol. 6, No. 11, page 120–121 <https://ssrn.com/abstract=3528447> accessed 18 October 2021, for further analysis of the application of fair use doctrine on AI and machine learning.

⁸⁰ The Copyright Act 1970 (Japan), Chapter ii,

Sec.5, Subsec.5, Art.47(7)

⁸¹ European Commission, Commission Staff Working Document – Impact Assessment on the modernization of EU copyright rules, Brussels 14.9.2016, SWD(2016) 301 final, part 2/3, page 27; See further European Commission, Standardisation in the area of innovation and technological development, notably in the field of Text and Data Mining (2014), Report from the Expert Group, Luxembourg, European Union, page 43–44.

⁸² National transposition measures communicated by the Member States concerning: 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 (Text with EEA relevance.), PE/51/2019/REV/1, OJ L 130, 17.5.2019, p. 92–125.

⁸³ Eleonora Rosati, Five considerations of the transposition and application of Article 17 of the DSM Directive (2021), IPkitten <https://ipkitten.blogspot.com/2021/02/five-considerations-for-transposition.html> accessed 19 October 2021.

creasingly needed to take place on a larger scale through cross-border and cross-discipline collaboration.⁸⁴ However, many different views have been expressed during the discussions in the European Parliament about the wording of the TDM exception and its narrow scope of application.⁸⁵ In addition, the text was also highly criticized by academics and AI innovators alike, who pointed out the fact that the formulation of the exception excludes startups and innovators to carry out TDM for commercial purposes, since the exception only covers not-for-profit and public research institutions, which leads to difficulties of fulfilling these requirements.⁸⁶

Consequently, considering the criticism and modifying the initial text, the DSM Directive was adopted comprising two mandatory TDM exceptions contained in articles 3 and 4; these are not, however, equally robust. Article 3 of the DSM Directive exempts act of reproduction and extraction made by research organizations and cultural heritage institutions to carry out, for the purposes of scientific research, TDM of lawfully accessed works or other subject-matter – including databases and press publications but excluding computer programs protected under the Software Directive, where a license may be required to undertake the restricted acts. In fact, article 3 of the DSM Directive does not exclude public-private partnership, where research organizations and cultural heritage institutions may rely on their private partners for TDM.⁸⁷ Nevertheless, one important aspect is that con-

tractual provisions overriding this exception are prohibited.⁸⁸

Article 4 of the DSM Directive allows acts of reproduction and extraction for anyone having a lawful access to works and other subject for the purpose of TDM, also including databases, press publications and computer programs. Accordingly, this article encompasses a much broader class of beneficiaries and permits TDM for all kinds of purposes regardless of any underlying commercial motives.⁸⁹ However, article 4 of the DSM Directive contains an opt-out provision, meaning that the rightsholder may expressly reserve in an appropriate manner the right to make reproductions and exactions for TDM purposes, for instance through machine-readable means, by contractual agreement or unilateral declaration.⁹⁰ In addition, article 4(2) of the DSM Directive provides that reproductions and extractions made of works and other subject-matter may be retained for as long as is necessary for the purpose of TDM.

After the preliminary overview of the mandatory exceptions, it is possible to critically review TDM regime introduced in the DSM Directive. In this regard, the upcoming sections aim at providing the positive and negative aspects of TDM exceptions and sets forth possible suggestions and recommendations for improvements that the remaining Member States may consider when implementing the directive into their national laws.

⁸⁴ European Commission, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Towards a modern, more European copyright framework, Brussels, 9 December 2015, COM(2015) 626 final, page 7; See also European Commission (n 79), Part 2/3, page 104.

⁸⁵ Committee on the Internal Market and Consumer Protection, Draft Opinion on the proposal for a directive of the European Parliament and of the Council on copyright in the Digital Single Market (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)), 20.2.20178, page 3; Committee on Industry, Research and Energy, Opinion on the proposal for a directive of the European Parliament and of the Council on copyright in the Digital Single Market (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)), 01.8.2017, page 23; Committee on Legal Affairs, Report on the proposal of a directive of the European Parliament and of the Council on copyright in the Digital Single Market (COM(2016)0593 – C8-0383/2016 – 2016/0280(COD)), A8-0245/2018, 29.6.2018.

⁸⁶ Open letter to European Commission, Maximizing the benefits of Artificial Intelligence through future-proof rules on Text and Data Mining [2018], Agency Submissions, Brussels, European Union, page 1-2.

⁸⁷ See recital 11 DSM Directive; See further Eleonora Rosati (n 7), page 212.

⁸⁸ See further article 7(1) DSM Directive.

⁸⁹ Bernt Hugenholtz, The new copyright Directive: Text and Data Mining (Articles 3 and 4), Institute for Information Law (IvIR), 24 July 2019 <http://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/> accessed 18 October 2021; See also Benjamin White et. al., Articles 3-4: Text and data mining <https://www.notion.so/Articles-3-4-Text-and-data-mining-9be17090ebc545b88ed9ac7d39e4e25a> accessed 18 October 2021.

⁹⁰ Recital 18 DSM Directive; See also Theodoros Chiou (n 42), page 409.

⁹¹ Benjamin Raue, Free Flow of Data? The friction between the Commission's European Data Economy Initiative and the proposed Directive on Copyright in the Digital Single Market [2018], Max Planck Institute for Innovation and Competition, Vol. 49, page 381-382.

⁹² See also recital 2 and 10 DSM Directive.

⁹³ Compared Article 5(1), 5(3)(a) InfoSoc Directive, articles 6 (2) (b) and 9 (b) Database Directive.

⁹⁴ Recital 11 DSM Directive; See further recital 18 DSM Directive, referring to both private and public entities.

⁹⁵ Benoit Van Asbroeck et. al., The EU Copyright

Directive: New exception for text and data mining [2019] <https://mediawrites.law/the-eu-dsm-directive-new-copyright-exception-for-text-and-data-mining/> accessed 19 October 2021.

⁹⁶ Obvious, Obvious is a collective of friends, artists and researchers <https://obvious-art.com/page-about-obvious/> accessed 19 October 2021; See also Ciara Nugent, The painter behind these artworks is an AI program. Do they still count as art? [2018] <https://time.com/5357221/obvious-artificial-intelligence-art/> accessed 19 October 2021.

⁹⁷ Judgement of 27 June 2013, Verwertungsgesellschaft Wort and others v. Kyocera and Others, C-457/11, EU:C:2013:426, para 36-38, which was summarized by Voluntary Memorandum from the UK Department for Business, Innovation and Skills, para 5 <https://publications.parliament.uk/pa/jt201415/jtselect/jtstatin/13/1321.htm> accessed 20 October 2021.

⁹⁸ See Recital 5 DSM Directive: The existing exceptions and limitations in Union law should continue to apply, including to TDM, as long as they do not limit the scope of exceptions and limitations under the DSM Directive.

4.2 Positive impacts on the training and development of creative AI

4.2.1 Harmonization of national laws in the digital age

A positive aspect of the DSM Directive is that it transfers a fundamental principle of copyright and related rights into the digital age, by allowing unauthorised uses of TDM under certain circumstances. It also provides a robust public interest to encourage the creation of new knowledge with the help of TDM, which would not be possible due to the excessive transaction costs. Nevertheless, it enables the rightsholders to participate in the economic value of their works or subject-matter by claiming remuneration for the actual use of the protected material, since both articles 3 and 4 of the DSM Directive require lawful access to the material but does not grant it.⁹¹ In fact, the key benefit of the DSM Directive is that it aims at harmonizing the national laws of Member States through mandatory solutions, meaning that each Member State is obliged to introduce them into their national laws.⁹² This will certainly reduce the national fragmentation and create much more certainty for the relevant market actors using TDM, leading to the promotion of more integrated and larger research projects across the EU and also vis-a-vis third countries.

4.2.2 Covering both commercial purposes and non-commercial purposes

A further justification for the TDM exceptions is that they also cover commercial purposes, as neither article 3 nor article 4 DSM Directive include the 'non-commercial' requirement.⁹³ In fact, article 3 of the DSM Directive includes private-public partnership, meaning that beneficiaries of this exception can rely on their private partners for carrying out TDM, including the use of their own technological tools.⁹⁴ Therefore, this may be an option for start-ups as they are 'time intensive and nearly impossible to handle for small teams'.⁹⁵ This was nevertheless the case in the project led by Obvious Art, consisting of a collective of researchers, artists and AI developers, where AI system, with the help of training data created by TDM, authored a portrait representing a member of the functional Belamy family and which was sold during 2018 for USD 432,500.⁹⁶ In addition, the scope of article 4 is broad in terms of its application, meaning that not only research purposes are covered but also any other TDM activity provided that these fall within the definition of TDM contained in article 2(2) of the DSM Directive.

4.2.3 Unenforceability of contractual provisions contrary to the exceptions

Another important aspect of the DSM Directive with regards to TDM exceptions is article 7(1), which expressly provides that any contractual provisions contrary to the exceptions provided for in inter alia article 3 of the DSM Directive shall be unenforceable. In fact, even though article 4 of the DSM Directive is not explicitly protected against contractual override, the CJEU has in *VG Wort*, C-457/11 stated that the 'default position where contract or license terms are not expressly allowed to limit the scope of an exception is that the exception will prevail over any rights holder authorization'.⁹⁷ Overall, even if the

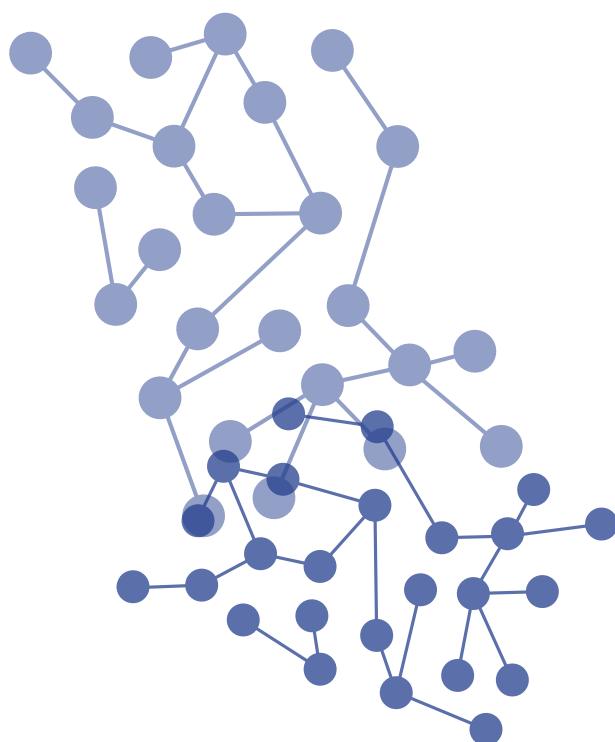
rightsholders may restrict the operation of the contractual override clause by drafting the provisions of it in a way that indirectly may restricts the lawful access which is required under both article 3 and 4 of the DSM Directive, the expressed prohibition to contractual override with regard to TDM is still a welcomed step, especially for AI development.

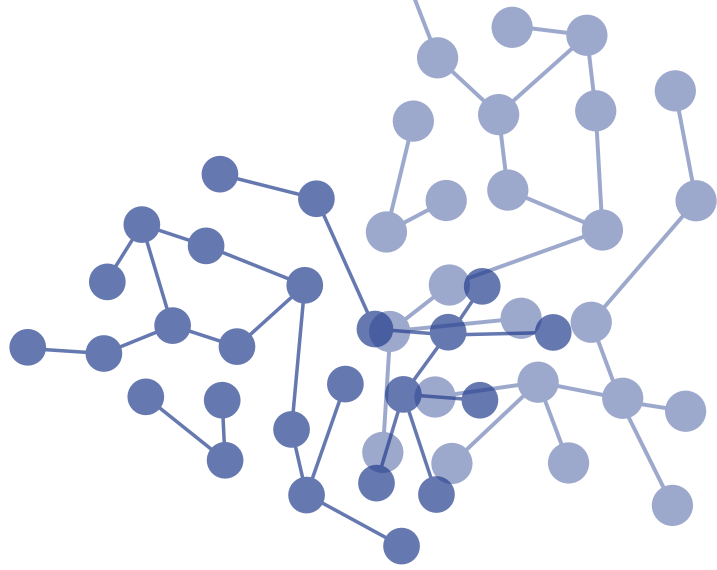
4.3 Negative impacts on the training and development of creative AI

4.3.1 Unresolved legal uncertainty

Despite the presented justification grounds for the TDM exceptions, there remain negative impacts that need to be assessed. As a matter of fact, article 4 DSM Directive obliges to implement either a mandatory exception or a limitation for TDM purposes, which means that Member States still have some discretionary power as to the scope of the provision they choose to implement. Consequently, this can lead to fragmentation and unharmonized treatment of TDM activities, leading to uncertainty as to the financial exposure of AI developers seeking to rely on this exception.

Furthermore, article 25 of the DSM Directive, which was not added in the proposal, clarifies that the Member States may adopt or maintain in force broader provisions with regards to TDM, within the limits of the EU *acquis*, irrespective of the mandatory articles 3 and 4 of the DSM Directive. This means that AI operators using TDM that fall outside the scope of the DSM Directive may still rely on the pre-existing legal framework as a fallback argument.⁹⁸ As a consequence, this discretionary power of Member States is likely to inhibit the harmonization within the internal market that the DSM Directive aims at achieving and provides uncertainty for AI developers using TDM.





4.3.2 Limited scope: Narrow purpose-specific approach

The formulation of the research exception in article 3 of the DSM Directive raises concerns with regards to its scope of application, especially when applying it to TDM for the purpose of AI-driven creativity. In fact, article 3 is limited to *inter alia* research organisations, but the article per se does not give a clear definition of the term. To qualify for the exception, research organizations have to operate either on a not-for-profit basis or by reinvesting all the profits in their scientific research, or pursuant to a public-interest mission.⁹⁹ Even if recital 11 of the DSM Directive provides a possibility for *de facto* public-private partnership, it does not apply to research institutions controlled by a commercial entity, e.g. where research organizations provide preferential access to the results of their research to those entities.¹⁰⁰

Accordingly, this exception fails to recognize the reality of scientific research nowadays, where many research organizations running the most cutting-edge TDM projects are often at least partly supported by private fun-

ding.¹⁰⁰ Also, unaffiliated researchers often conduct TDM projects in the framework of public-private partnership, where these fall within the scope of a research organisation but are decisively influenced by a commercial undertaking.¹⁰² Consequently, the narrow scope of the exception limits the possibility of others conducting 'scientific research' to rely on this exception, which can in turn lead to the following; (i) extreme transaction costs for organizations and commercial private actors conducting AI research as well as unaffiliated researchers that are forced to obtain a license for content they mine during TDM¹⁰³; (ii) discourage the undertaking of TDM on a large amount of data and decrease the quality of the research results, which in turn might impede competitiveness of the EU vis-à-vis third countries¹⁰⁴; and (iii) may also lead to total ignorance of copyright and related rights that will further damage the integrity of these systems.

In addition, even if recital 12 of the DSM Directive describes the term 'scientific research' as covering both natural and human sciences, the lack of reference to a wider purpose in article 3 may restrain the effectiveness of the exception and produce practical difficulties. As an example, there might be different interpretations of the classification of science; if computer science is not classified as natural science *stricto sensu* it does fall outside the scope of article 3 of the DSM Directive.¹⁰⁵ Consequently, this may have a negative impact on AI development and put EU at a competitive disadvantage in the competitive global market for world-class AI, where the most talented researchers will take jobs abroad and commercial private actors will relocate their place of establishment because of the better chances to undertake TDM.

Therefore, recital 12 of the DSM Directive could have been formulated in a wider sense, by describing 'scientific research' as any form of activity carried out in a methodological and systematic way that aims exploring a certain

⁹⁹ See Article 2(1)(a)-(b) DSM Directive.

¹⁰⁰ See further recital 11 and 12 DSM Directive.

¹⁰¹ Rossana Ducato et. al. (n 55), page 19.

¹⁰² Pamela Samuelson, The EU's Controversial Digital Single Market Directive (2018), Communications of the ACM, Volume 61, Issue 11, page 23; See also Geiger et. al., Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU (2019), Center for International Intellectual Property Studies (CEIPI), Research Paper N. 2019-08 <https://ssrn.com/abstract=3470653> accessed 20 October 2021.

¹⁰³ Since TDM is generally associated with quantity, especially when it comes to training of AI system, where TDM may involve copying and/or extraction of thousands, if not millions, of protectable works or other subject-matters, this may be difficult to fulfill.

¹⁰⁴ Also including UK, where the TDM exception in section 29A CDDPA is not limited to certain beneficiaries (in fact, it also includes individual researchers as beneficiaries and anyone having lawful access to the material).

¹⁰⁵ Rossana Ducato et. al., Ensuring Text and Data Mining: Remaining issues with the EU

copyright exceptions and possible ways out (2021), CRIDES Working Paper Series No. 1/2021, page 11.

¹⁰⁶ Recital 18 DSM Directive, stating that the rightsholder may reserve the rights using machine-readable means, including metadata and terms and conditions, and contractual agreements or unilateral declaration. This practical disapplication is also formulated in recital 18 of the DSM Directive, stating that the rightsholders 'remain able to license the uses of their works or other subject-matter' in all cases falling outside the mandatory exception in article 3 of the DSM Directive or article 5(1) InfoSoc Directive, and further confirmed by the article 7(1) by not expressly protecting other profit-making entities or organizations against contractual override.

¹⁰⁷ Geiger et. al. (n 3), page 21.

¹⁰⁸ Rossana Ducato et. al. (n 103), page 15; See further article 8(2) Software Directive for contractual restrictions.

¹⁰⁹ See reasoning conducted by Geiger et. al. (n 3), page 22, which the author of this thesis agrees with.

¹¹⁰ League of European Research Universities,

Europe needs a broad & mandatory TDM exception (2018) <https://www.leru.org/news/europe-needs-a-broad-mandatory-tdm-exception> accessed 22 October 2021.

¹¹¹ For the definition of the information society service provider see further article 1 (b) Directive (EU) 2015/1535 if the European Parliament and of the Council of 9 September 2015 laying down a procedure for the provision of information in the field of technical regulations and of rules on Information Society services (codification), OJ L 241, 17.9.2015, p. 1-15.

¹¹² See further Rossana Ducato et. al. (n 103), page 14.

¹¹³ Tatiana Eleni Synodinou, Lawfulness for users in European Copyright Law: Acquis and Perspectives (2019), JUPIPEC 20, page 26.

¹¹⁴ European Copyright Society, General opinion on the EU copyright reform package (2017), page 4 <https://europeancopyrightsocietydo-torg.files.wordpress.com/2015/12/ecs-opinion-on-eu-copyright-reform-def.pdf> accessed 23 October 2021.

¹¹⁵ See also Pamela Samuelson (n 100), page 23.

subject-matter to discover new data or information or to generate new knowledge to advance the state-of-the-art in a certain field – as it is the case in the general research *per se*. Even though article 3 would still include the purpose of ‘scientific’ research, it would, however, not restrict the scope of application on the scientific area for which the research is undertaken. This may be a solution for the remaining 16 Member States, when transposing the DSM Directive into their national law, to adopt a broad definition of ‘scientific research’ in the context of the mandatory exception covering TDM in article 3 of the DSM Directive.

4.3.3 Limited scope: The ‘opt-out’ mechanism

Another issue is that article 4(3) of the DSM Directive limits the possibility of the concerned beneficiaries to rely on the exception by providing for the ‘opt-out’ mechanism, meaning that it can be easily overridden by any expression of will, whether by contract or unilateral declaration.¹⁰⁶ Consequently, this undermines the general principle that ‘the right to read is the right to mine’, where having lawful access to protected material shall include the right to mine a particular content.¹⁰⁷ In fact, the wording of article 4 may create a Schrödinger’s paradox: for instance, the activity of observing, studying or testing the function of a computer program or the normal use of a database may be restricted by contractual provisions if article 4 DSM Directive is applied, but this is certainly not the case if one considers the voidance of restrictive contractual clauses under the Software and Database Directives.¹⁰⁸ As a result, this may create further uncertainties for AI developers and discourage them from undertaking TDM. Instead, it would certainly be more favourable to consider fair remuneration, when the potential harm could be shown.¹⁰⁹

Another issue that exists under article 4(2) of the DSM Directive is that reproductions and extractions made during TDM may be retained ‘for as long as is necessary’ for the purpose of the analysis. In fact, the wording of the article indirectly allows merely ‘temporary reproductions’, which is highly similar to the mandatory exception in article 5(1) of the InfoSoc Directive.¹¹⁰ As previously been discussed, the exception for temporary acts of reproduction is not enough to cover unlicensed use of TDM for AI creative purposes, leading to a risk where no public or private AI developers will make a large invest that is required to mine data in case where copies may only be retained on a temporary basis. Consequently, article 4 of the DSM Directive efficiently creates and actualizes a derivative market for TDM, which the rightsholders may wish to control, license or also totally restrict. Solution for the Member States, when transposing the directive, would be to provide a clear indication with regards to the ‘appropriate manner’ in which the rightsholder may reserve the use for TDM.

4.3.4 Structured ambiguities in the scope of application

Several inconsistencies also arise with regards to the scope of application of both mandatory TDM exceptions in relation to other provisions of the DSM Directive. Firstly, both articles 3 and 4 of the DSM Directive cover press publishers’ right of reproduction introduced in article 15

of the DSM Directive, but the inclusion of it in article 3 of the DSM Directive may appear to be ambiguous. Since the protection under article 15(1) is granted to publishers in relation to the use of their press publications by information society service providers¹¹¹ in the online environment, such provision certainly does not give any rights against third parties. Consequently, beneficiaries of article 3 of the DSM Directive are not likely to be qualified as information society service provider *per se*, and the reference to that article becomes inadequate.¹¹²

Secondly, only article 4 of the DSM Directive refers to the economic rights granted by the Software Directive, while beneficiaries of article 3 of the DSM Directive are excluded from taking part of this privilege, where a license will always be required. However, this exclusion can partly be explained by the fact that article 5(3) of the Software Directive already provides for the ‘black box analysis’ exception, by allowing the lawful users to study, observe or test the functions of the program without prior authorization. Accordingly, since the mandatory exception in Software Directive only covers non-commercial research purposes, the EU legislators clarified in article 4 of the DSM Directive that reproduction of computer programs outside research sphere can also be done for commercial purpose. This justification is, however, not fully valid, due to the ‘opt-out’ provision and retention period under article 4 of the DSM Directive.

4.3.5 Pre-condition of ‘lawful access’

The requirement of ‘lawful access’ appears to be a pre-requisite for enjoyment of exception in both articles 3 and 4 of the DSM Directive, which closely follows the model of the UK exception. In fact, these articles do not provide *per se* a clear definition of this requirement, where the guidance instead can be found in recital 14 of the DSM Directive explaining that lawful access to protected works and other subject-matter occurs *inter alia* when researchers have access through subscriptions to publications, open-access licensing or through other lawful means, including content freely available on the Internet. However, the recital does not indicate whether lawfulness of access is evaluated only objectively or whether other factors may be taken into account, for instance the presumed ‘state of mind of the user’ in relation to the lawfulness of the source of the work and subject-matter.¹¹³

In fact, compared to the user’s rights, i.e., lawful user, under the Software and Database Directives, the lawful access represents a more strict approach. This makes the exception subject to private ordering, meaning that the enjoyment of both TDM exceptions is dependent on the market decisions of the rightsholders, where these can successfully deny access to works and other subject-matters or only grant access on conditional terms.¹¹⁴ Consequently, it may be difficult especially for start-ups and small and SMEs to negotiate with owners of big data sets about TDM licenses on reasonable terms, which puts them at risk of being excluded from the scope of TDM and AI and seriously jeopardize their innovation opportunities.¹¹⁵

This may, in fact, also lead to a more difficult undertaking of TDM projects, as it will raise related costs and budget considerations will restrict the scope of research.

In addition, this may lead to discrimination that depends on the research organizations market power, meaning that only limited number of organizations will be able to acquire licenses for all the databases that are indeed relevant for the TDM project at stake. This will in turn spread the gap between richer and poorer research institutions and increase the cleave between research in developed and less developed countries.¹¹⁶ As a matter of fact, given the importance of the initial datasets generated by TDM for the training of creative AI, both in terms of quality and quantity, there might be a risk that the outcome of AI will be of a lower quality if it is trained on small datasets that can be easily accessed by AI developers.

A possible solution would be the promotion of data pools, which are centralized repositories of various data/information, where it can be obtained, maintained or exchanged between different market actors.¹¹⁷ Accordingly, the creation of a sole set of works or other subject-matter through aggregation, would ease for the potential AI developers wishing to train AI for creative outputs since licensing work by work would not be required. Consequently, enabling rightsholders to license their content for TDM purposes may motivate them to generate high-quality datasets for commercialization and distribution – beneficial for them and the users.

4.3.6 Issues of coexistence with Technical Protection Measures

Even if contractual limitations are not allowed, at least in relation to article 3 of the DSM Directive, both exceptions are subject to technical protection measures (TPM) established in article 6 of the InfoSoc Directive, as referred to in article 7(2) of the DSM Directive, allowing rightsholders to effectively block access for AI operators seeking to conduct TDM. The reference to TPM can clearly be found in article 3(3) of the DSM Directive referring to ‘measures to ensure security and integrity of networks and databases’ and article 4(3) of the DSM Directive referring to ‘reserved in an appropriate manner’ *inter alia* by machine readable means. Indeed, these should not prevent the enjoyment of the mandatory exceptions and limitations under the DSM Directive and shall not exceed what is necessary to pursue the objectives thereof.¹¹⁸

However, despite the good intentions of the DSM Direc-

tive, the applicability of the anti-circumvention provisions might encroach on users’ privileged uses, meaning that TPM are at risk of limiting or preventing the access to protected material for purposes that may not be restricted by the exclusive rights or for uses that are allowed per se. In fact, the obligation of the rightsholders to make available content, for users to benefit from the exceptions and limitations according to article 7(2) of the DSM Directive, does not limit liability for circumvention.¹¹⁹ Consequently, DSM Directive does not in fact grant any effective protection against TPM since it is not yet clear whether there is a possibility to legally circumvent those technical measures that would unlawfully limit TDM. All in all, considering the fact that this mechanism has not proven to be effective for the past 20 years, since the adoption of the InfoSoc Directive, it will most likely not work for TDM now through the DSM Directive.¹²⁰

5. CONCLUSION

It is certainly true that the DSM Directive meets important policy goals and aims at supporting and promoting the work that is being undertaken in the field of Big Data and AI within the EU, by introducing a mandatory solution for TDM and thus harmonizing national laws between the Member States. However, this article has raised several uncertainties with regards to the possibility of these TDM exceptions to achieve a fair balance between the promotion of technological development, on the one hand, and the interest of the rightsholders, on the other. Accordingly, the wording of the exceptions strongly limits the effectiveness of the reform and its ability to promote competitive advantage within the EU entities engaging in TDM for the purpose of AI-driven creativity.

All this said, even though the DSM Directive follows an approach that better fits the digital environment – compared to the long existing InfoSoc, Database and Software Directives – it does fail to address the new era of the Fourth Industrial Revolution to which AI belongs. This conclusion is justified by the fact that there is *prima facie* clear preference for the protection of copyright and related rights of the relevant rightsholders controlling the content, for instance by the requirement of lawful access and the opt-out mechanism contained in articles 3 and 4 of

¹¹⁶ European Copyright Society (n 112), page 4; See further Geiger et. al. (n 3), page 22.

¹¹⁷ IBM, Data pool (2020) <https://www.ibm.com/docs/en/imdm/11.6?topic=gds-data-pool> accessed 24 October 2021.

¹¹⁸ See further recital 7 and 16 DSM Directive.

¹¹⁹ Geiger et. al. (n 3), page 23.

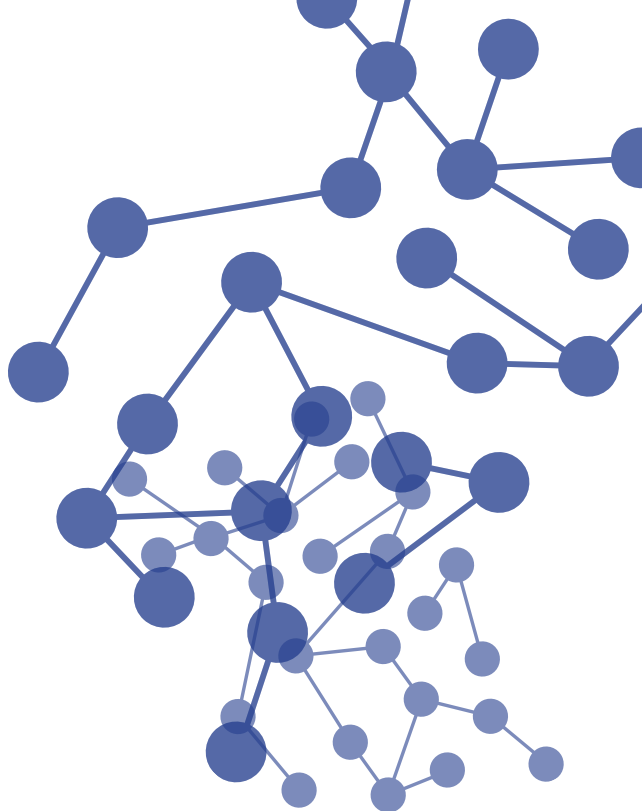
¹²⁰ Rossana Ducato et. al. (n 103), page 16-17; See also Thomas Margoni et. al. (n 358). Thomas Margoni et. al., The Text and Data Mining exception in the Proposal for a Directive on copyright in the Digital Single Market: Why it is

not what EU copyright law needs [2018] <https://www.create.ac.uk/blog/2018/04/25/why-tdm-exception-copyright-directive-digital-single-market-not-what-eu-copyright-needs/> accessed 25 October 2021.

the DSM Directive. Ultimately, the DSM Directive did overlook the opportunity for true modernization of the EU *acquis* on copyright and related rights in the digital single market and it seems that, at some point, it missed to strengthen its competitive position with regards to unlicensed TDM for the purpose of AI-driven creativity, both internally and vis-à-vis third countries such as US, Japan and also including UK.

As a matter of fact, at the time of writing this article the current position of unlicensed TDM for the development of AI, including AI-driven creativity, and the future of these technologies within the EU is undetermined. Therefore, the actual transposition of the DSM Directive into national laws by the rest of the 16 Member States may certainly represent an important opportunity for them to design a more advantageous TDM environment, by improving the mandatory exceptions, particularly through interpretation of ‘scientific research’ in article 3 of the DSM Directive, the ‘opt-out’ mechanism in article 4 of the DSM Directive as well as spell out that contractual and technological measures should not deprive the effective application of the mandatory TDM exceptions.

As a result, this may eventually encourage more researchers and businesses to rely on TDM techniques and thus enhance competition within the EU, including vis-à-vis third countries, which is also per se the ultimate goal of the EU. In addition, the proposed changes that can be made during the national transpositions may also enhance the development of the innovative AI projects in the field of algorithmic creativity. Until then, copyright and related rights still remain an obstacle for AI development, including AI-driven creativity, and it is therefore not accidental that some of the nominal AI generated creations in the EU, for instance ‘The Next Rembrandt’ and ‘SKYGGE’, are based on mining of works from the public domain.



Kristina Christensen

Kristina Christensen is currently working as a risk and compliance consultant at KPMG Sweden. She has an LL.M. in European Business Law from Lund University and an LL.M. in European Intellectual Property Law from Stockholm University. She also has extensive knowledge and interest in IP law and its relation to the most recent technological innovations, such as Artificial Intelligence.