

European copyright law and the text and data mining exceptions and limitations

In light of the recent DSM Directive, is the EU approach a hindrance or facilitator to innovation in the region?

By Charlotte Gerrish and Anders Molander Skavlan

ABSTRACT

The newly adopted European Union (EU) Directive on copyright and related rights in the Digital Single Market¹ (DSM Directive) provides for text and data mining (“TDM”) exceptions for the first time at an EU legislative level. The TDM provisions are a step in the right direction, but the situation remains uncertain for innovators and researchers alike.² The DSM Directive’s TDM provisions demonstrate the EU legislature’s lack of understanding of the technical realities of TDM, and the financial and contractual limitations which are faced by those performing such processes, especially companies etc. in the start-up phase. Through these provisions, the EU has made it difficult for valuable TDM output to originate from European-based operators. To drive innovation in the EU and further the commitment to the digital single market, it would have been more desirable for the EU to implement a broad, all-encompassing TDM exception, not been capable of being overridden³ nor subject to caveats, as is the case in other key innovative jurisdictions. By creating and maintaining a restrictive environment for TDM, which is often a key component of Artificial Intelligence (“AI”) and machine learning, the EU runs the risk of losing a foothold in a growing industry, and causing an exodus of key talent to other TDM-friendly jurisdictions. However, the TDM provisions contained in the DSM Directive are at least one step in the right direction and one can hope that future legislative development in Europe and a sensible approach by Member States will take place in the foreseeable future.

1. INTRODUCTION

We are in a digital age with big data at the heart of our global, digital environment. Exploiting big data by manual means is virtually impossible, meaning that we need to rely on innovative methods such as machine learning and AI to allow us to fully harness the value of big data which is available in our digital society. One of the key processes allowing us to innovate using new technologies such as

machine learning and AI is through the use of TDM carried out on large volumes of big data.

While there is no single definition of TDM, it is universally acknowledged that TDM involves the automated analytical processing of raw and unstructured data sets through sophisticated technological tools in order to obtain valuable insights for society or to enable efficient machine learning and AI development. Such TDM processes include extraction and reproduction of source text, some of which is likely to be protected by copyright. This of course creates tension between the exclusive rights of copyright holders⁴ and the interests of innovators developing TDM technologies and performing TDM. This has caused the relevant players to either avoid TDM activities and the development of and investment in the relevant technology, or has obliged them to resort to licensing solutions. They therefore shoulder the administration and transaction costs of the licenses to avoid widescale copyright infringement claims⁵ when performing TDM on big data.

1.1 Can “big data” be protected by copyright?

Arguably, one of the

“basic and fundamental principles of copyright law is that data is not protected, as copyright only protects the creative form, not the information incorporated in the protected work.”⁶

Indeed, on this basis, perhaps TDM operators should not be concerned by any intellectual property rights, whether copyright or otherwise, as TDM activities potentially fall outside the scope of any intellectual property right monopoly. On this basis, the risk of copyright infringement in respect of TDM processes carried out on data is a non-issue: data in itself is simply not capable of copyright protection.

However, in the context of *big data*, and given the three Vs (volume, velocity and variety)⁷ applicable to it, mere “data” must be distinguished from big data. As such, it is likely that literary copyright subsists in documents, publications, research and analysis, as well as in any technical documents, software and IT architecture which constitute big data, and which is ultimately subject to TDM activities. Copyright within a database (so-called “database copyright”) may apply to big data in some instances, aside



from any independent database right. Indeed, by way of example, the CJEU confirmed in the *Infopaq case*⁸ that the threshold for copyright protection (and the risk of infringement) can occur when there is reproduction of text extracts comprising at least of eleven words of copyrightable material. AI and machine learning, which are developed through TDM, rely on processing masses of data, so it is likely that this de minimis threshold will often be met, meaning that when TDM is carried out on big data, a risk of wide scale copyright infringement exists.

Accordingly, given the rapid pace at which technology is developing, there is no doubt that the legal framework needs to adapt in order to avoid becoming redundant, and more importantly, to prevent outdated legal regimes from hindering innovation. Until the DSM Directive entered into force on 7 June 2019, there was no specific TDM exception at an EU level and innovative firms had to rely on a patchwork of limitations and exceptions to a copyright holder's exclusive rights,⁹ which were not adapted to the realities of TDM activities. Furthermore, the majority of provisions existing prior to the DSM Directive are not

mandatory, unfortunately resulting in a patchwork approach to exceptions and limitations across the EU, including their application to TDM, creating uncertainty for stakeholders.

It was therefore generally hoped that the provisions in the DSM Directive which relate to specifically to TDM¹⁰ would provide the solution which has thus far been absent within the EU. Regrettably, Articles 3 and 4 of the DSM Directive which govern exceptions to copyright infringement when TDM is conducted on copyrighted works are fraught with caveats and limitations. These provisions also create an undesirable distinction between TDM conducted for research on one hand, and for other purposes (i.e. commercial) on the other hand. As drafted, the DSM Directive arguably does little to reduce uncertainty for copyright holders and innovators alike, which risks having an impact on innovative TDM-based developments in Europe. Indeed, firms may simply relocate to jurisdictions offering more legal certainty in the field of innovation and TDM specifically, such as Japan¹¹ or the US.¹²

¹ 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.

² Europe Needs A Broad & Mandatory TDM Exception, Association of European Research Libraries, 13 November 2018, available at: <https://libereurope.eu/blog/2018/11/13/europe-needs-a-broad-mandatory-tdm-exception/> (accessed on 11 May 2019).

³ Ibid.

⁴ Particularly under Article 2(a) of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society.

⁵ Cabrera Blázquez F.J., Cappello M., Fontaine G., Valais S, IRIS Plus, 'Exceptions and Limitations to Copyright', A Publication of the European Audiovisual Observatory, Council

of Europe, 2017, available at: <https://rm.coe.int/168078348b> (accessed on 2 December 2019), p. 67.

⁶ Geiger, C., Frosio, G. & Bulayenko, O., 'Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data? Legal Analysis and Policy Recommendations', IIC (2018) 49: 814. <https://doi.org/10.1007/s40319-018-0722-2> (accessed on 2 December 2019), p. 817.

⁷ Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by "Big Data", OECD Digital Economy Papers, No. 222, OECD Publishing, Paris, available at: <http://dx.doi.org/10.1787/5k47zw3fcp43-en> (accessed on 2 December 2019), p.11.

⁸ Judgment of the Court (Fourth Chamber) of 16 July 2009, Case C-5/08 - Infopaq International, ECLI:EU:C:2009:465.

⁹ For example, Article 5, InfoSoc Directive;

Articles 6 and 5, Software Directive; Articles 6 and 9, Database Directive; Articles 6 and 10, Rental Right and Lending Directive; Article 6, Orphan Works Directive.

¹⁰ See Articles 3 and 4 of the DSM Directive.

¹¹ The 2018 Amendment to the Japanese Copyright Act offers a wide range of unfettered TDM exceptions. For the official provision (in Japanese) visit: http://www.mext.go.jp/b_menu/houan/kakutei/detail/1405213.htm (accessed 5 December 2019).

¹² Where innovators can rely on the doctrine of Fair Use, including for commercial purposes, as recently confirmed in the Google Books saga – see e.g. the US District Court Southern District of New York, Opinion 05 Civ. 8136 (DC) (22 March 2011).

2. DISCUSSION

2.1 The Proposal - Introducing TDM exceptions in EU law

Issues surrounding the scope of the TDM provisions in the DSM Directive were apparent at its inception. Article 3 of the Proposal for the DSM Directive¹³ (“Proposal”) initially envisaged a mandatory exception for TDM activities which extended only to

*“reproductions and extractions made by research organisations in order to carry out text and data mining of works or other subject-matter to which they have lawful access for the purposes of scientific research”*¹⁴

Taking the TDM exception as initially provided for in the Proposal, therefore, was clearly not reflecting the European Commission’s intention to drive innovation within the EU, especially in the area of TDM used on big data for commercial purposes. Indeed, whilst the initial drafting of Article 3 at the Proposal stage imposed a mandatory exception on Member States, which is a positive aspect when compared to the majority of optional exceptions contained in EU legislation to date, and whilst certainty is provided to TDM actors in that the scope of the legal provision is not capable of being overridden by contract, the extent of the Proposal’s TDM exception was highly limited. The scope of the exception not only required operators to have “lawful access” to the copyrighted works, which presumably meant either via open access channels, or pursuant to licence or subscription agreements, but it also limited the scope of the TDM exception to academia, and notably to “scientific research”, which is to the exclusion of start-ups and innovators which carry out TDM for commercial means.

Furthermore, the wording of the Proposal expressly permitted the application of Technical Protection Measures (“TPMs”) to protect copyrighted works, which means that it would have been possible to technically prevent reproduction and extraction of copyrighted works, even when a lawful exception permits such actions. The text of the Proposal was therefore highly criticized by academics and innovators alike.¹⁵

2.2 The DSM Directive – Creation of two regimes

Given the modifications to the finalized text of the DSM Directive, it appears the EU legislators took these criticisms on board when drafting the final version of the DSM Directive, at least to a certain extent. However, turning to the wording of the final TDM provisions of the DSM Directive, Article 3 remains materially unchanged from the wording of the Proposal. There remains a mandatory exception for research organisations for scientific research.

Indeed, the important change contained in the DSM Directive as compared to the Proposal is the inclusion of Article 4 which expressly provides for a mandatory exception or limitation to be implemented by Member States for TDM activities beyond the previous narrow provisions limited solely to scientific research by research organizations as initially laid down in the Proposal. This means the EU has now gone some way to resolve the issues related to commercial TDM activities and thus drive innovation in the region.

We now have two regimes for TDM exceptions and limitations – one for scientific research and another for all other TDM activities. The rights and obligations for each TDM purpose are not aligned, and this is to be criticized – the playing field is not level between different actors and purposes, and the environment for commercial TDM therefore remains unfavourable in Europe and constitutes a potential hindrance to innovation in the region. Despite the improvements made to the final text compared to the initial drafting of the Proposal, it has nonetheless been argued that the “*project to allow Europeans to conduct TDM, which is crucial for modern research and the development of AI, has been obstructed with too many caveats and requirements*”,¹⁶ particularly since we are now faced with two differing regimes for research and “other” TDM operations.

2.3 The limited scope of Article 4

The broader TDM exception is arguably devoid of function due to the possibility for contractual override at Article 4(3) of the DSM Directive. This provision results in a scenario whereby holders of copyrighted works are entitled to expressly disapply Article 4 for all TDM activi-

¹³ Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market COM/2016/0593 final - 2016/0280 (COD).

¹⁴ Cabrera Blázquez F.J., Cappello M., Fontaine G., Valais S, IRIS Plus, ‘Exceptions and Limitations to Copyright’, A Publication of the European Audiovisual Observatory, Council of Europe, 2017, available at: <https://rm.coe.int/168078348b> (accessed on 2 December 2019), p. 67.

¹⁵ Maximising the benefits of Artificial

Intelligence through future-proof rules on Text and Data Mining, Open Letter to the European Commission, Brussels, 9th April 2018 available at: https://eua.eu/downloads/news/openletter-to-european-commission-on-on-ai-and-tdm_9april2018.pdf (accessed on 10 March 2019).

¹⁶ Julia Reda “The Text of Article 13 and the EU Copyright Directive has just been finalized”, available at: <https://juliareda.eu/2019/02/eu-copyright-final-text/> (accessed on 11 May 2019).

¹⁷ Recital 18 of the DSM Directive.

¹⁸ Judgment of the Court (Second Chamber) of 15 January 2015, Case C-30/14 – Ryanair, ECLI:EU:C:2015:10.

¹⁹ See Articles 6(1), 8 and 15 of Directive 96/9/EC.

²⁰ Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (InfoSoc).

ties, save those related to scientific research (which is solely governed by the provisions of Article 3). It is appropriate to criticize the practical application on Article 4 in light of this flexibility, since this wider provision is so easy to disapply, either by technical means or indeed by contract or unilateral declaration.¹⁷ This is confirmed by recital 18 and Article 7(1) of the DSM Directive which do not include Article 4 in the express protection against contractual override; there can be no uncertainty as to the ability of rightsholders to override Article 4 at their discretion. Practically, we can envisage that rightsholders in an online environment will disapply Article 4 as in the PR Aviation/RyanAir case through the application of exclusions to screen-scraping TDM activities in website terms and conditions or website notices.¹⁸

Of course, such practices are inherently difficult to monitor, TDM operated by AI is generally unable to ascertain when such contractual restrictions have been applied to a website, and of course, depending on the status of the TDM operator (i.e., a research organization or otherwise) the legal application of such terms will vary meaning that in some instances the contractual override will be valid, and in other circumstances, not. The difficult application of contractual overrides or unilateral notices is akin to the application of technological restrictions, as are discussed further below, and also potentially creates further confusion for TDM operators. This is depending on the nature of the IP rights protecting the source content – for example, when screen scraping activities include copyrighted works or works covered by the Database Directive, for which such contractual override is not permissible,¹⁹ or when a notice or contractual restriction does not take into account activities which do not require consent of the relevant rightsholder.

In light of the foregoing, we consider that Article 4 essentially acts as an optional exception for broader TDM activities not falling within the scope of Article 3. Is this evidence of the EU simply playing lip-service to the industry criticisms surrounding the narrow scope of the initial TDM exception as contained in the Proposal, or just evidence of a failure to understand the realities of TDM and the likelihood of harm suffered by rightsholders? Either way, it is regrettable that despite much negotiation and effort, the DSM Directive still fails to provide non-research TDM operators with certainty as to their activities and their protection against copyright infringement actions. Such exposure being dependant on an individual rightsholder's reservation – something which is almost impossible to monitor and check as data mining analytics processes huge volumes of information, often coming from thousands of source resources – means that such organisations may find themselves in the precarious position of relying on the provisions of Article 5 of the InfoSoc Directive,²⁰ which are insufficient. As it stands, the ability of rightsholders to exclude Article 4 by various means effectively renders the provision devoid of function, leaving Europe an uncertain environment for TDM actors.

2.3.1 Article 4(2) creates legal uncertainty

Furthermore, the scope of the broader TDM exception remains unclear. On reading Article 4 of the DSM Directive, there is likely to be confusion about the extent of the exception or limitation for non-research purposes. Article 4(2) states that:

“Reproductions and extractions made pursuant to paragraph 1 may be retained for as long as is necessary for the purposes of text and data mining.”

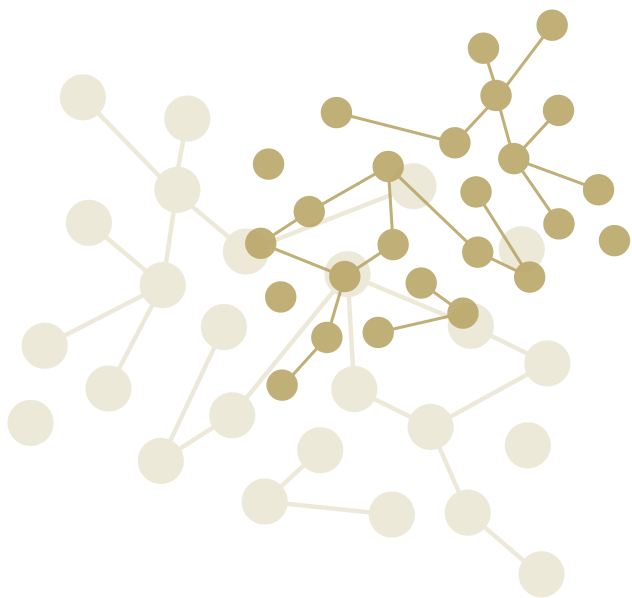
In our view, “necessary for the purposes of TDM” may lead to uncertainty as it could be argued by rightsholders that “necessity” is simply limited to the time required to complete the technical process, whereas operators might wish to extend that time frame to purposes which go beyond the mere technical processes.

This is to be contrasted with the provisions in the DSM Directive regarding scientific research which are not subject to the same restrictions of “necessity”. Indeed Article 3(2) of the DSM Directive states that:

“Copies of works or other subject matter made in compliance with paragraph 1 shall be stored with an appropriate level of security and may be retained for the purposes of scientific research, including for the verification of research results”.

It is therefore undesirable for the DSM Directive to create stark differences regarding the retention of source data depending on whether such data has been mined for research or other purposes. Unfortunately, yet again, the EU has failed to provide certainty for TDM activities across the board and the retention rights of source data is therefore very much subject to the identity of the TDM operator and the purpose that that operator seeks to achieve.





2.4 Article 3 – TDM for scientific research from selected research organisations

Whilst recital 12 of the preamble to the DSM Directive provides for a wide notion of scientific research which extends to both natural and human sciences, and provides certainty for specific categories of beneficiaries of the exception contained at Article 3 by listing potential organisations falling within the provisions,²¹ Articles 2(1)(a) and (b) restrict the scope of qualifying for the exception contained at Article 3 by providing a narrow interpretation of research organisations. Indeed, these provisions provide that research organisations must operate on a not-for-profit basis, or reinvest all its profits into its scientific research or pursue a public interest mission funded by public funds or public contracts, in order to qualify with certainty as to their TDM activities. As such, the scope of the DSM Directive is prohibitively narrow when defining the nature of a research organization.²²

This essentially restricts commercially-backed research organisations from being able to benefit from Article 3, even if they are ultimately carrying out “scientific research” for the purposes of the DSM Directive, and therefore excludes certain research organisations, such as private universities. While on one hand, the EU clearly wished to ensure that scientific research carried out for TDM purposes remains neutral and independent from industry, in the current austerity climate in which public funding and investment is scarce and where regard must be had to the private sector to obtain budget for the most cutting edge research, the restriction to qualification for this exception is perhaps unwittingly a move by the EU, which will result in stunted innovation through research in the region.

2.4.1 Requirement of “lawful access”

The DSM Directive provides that for both research and non-research TDM, the relevant limitation/exception shall only apply to operators which have “lawful access” to the copyrighted content. If access to the volume of source data is limited in any way, or is subject to unnecessary risk (i.e. of copyright infringement), then not only would the TDM output be less thorough and thus less valuable, the

reduced amount of source data processed by machines could also have a negative impact on the development of AI as such technology requires significant amounts of data to machine-learn.

From a research perspective, it has been argued that subjecting TDM to lawful access will make TDM research projects harder to run by raising related costs, meaning that publishers of content might price TDM into their subscription fees, if only those with lawful access can perform TDM research.²³ Most innovative start-ups and research organisations will be effectively prevented from being able to gain lawful access to works due to the cost of such access, which is problematic for TDM operators (specifically research organisations) coming from less economically sound environments where public funding may be scarce.

Ultimately, for both commercial and research TDM, the overall quality and value of the AI development and TDM output is likely to be put at risk whenever there is a requirement for lawful access which results in payments being made or costly subscriptions being taken out, as budget restrictions take over.²⁴ The underlying result is that where cost of conducting TDM is increased, researchers and innovators are less likely to use it, and will potentially favour either other methods or move their activities to territories where such cost-burdens are lower, and which may be located outside of the EU altogether.

As indicated above, lawful access is therefore likely to be subject to licence arrangements, save where such data is freely available in the public domain, or indeed when source data is not protected by copyright and so no consent is required, and therefore no licence is needed to perform TDM. Nevertheless, where consent is required and the mechanism to obtain such consent falls to licensing or contractual arrangements, the situation is problematic.

Arguably, the research sector is more heavily affected by the requirement of lawful access and the potential issues related to the licensing of content to be mined since commercial TDM is often focused on some areas of online analytics (such as retail analytics) which relate to consumer movements and trends gained through the use of cookies, plug-ins or social media.

However, while the content to be mined for commercial purposes is therefore more often freely available, this is not necessarily the case for scientific research where the source data is subject to more onerous access restrictions (online databases or private library content). The DSM Directive nonetheless assumes that profit-making firms can and should get a license to engage in TDM research from the owners of the affected IP rights, which is not necessarily the case, especially for start-ups which have limited access to financial resources.²⁵

2.4.2 Additional hindrance for TDM innovations – technological protection measures

A further issue reducing the effectiveness of the DSM Directive in creating a favourable environment for TDM activities in Europe is the inclusion of provisions which confer on rightsholders the possibility of limiting TDM activities via the application of technological protection

measures. This allows the owners of copyrighted works to block access to operators seeking to carry out TDM²⁶ which may prevent the enjoyment of available exceptions and limitations. The DSM Directive does little to balance the interests of TDM operators and rightsholders in practice as regards the application of anti-circumvention provisions on works which can be mined under a lawful exception, and it will be interesting to note how the Member States approach this dilemma from a national perspective once the provisions have been implemented and are applied to concrete cases.

Ultimately, as with other aspects of the TDM provisions in the DSM Directive, the drafting of the text as regards TPMs means there is a risk of inconsistent implementations across national jurisdictions which might effectively curtail harmonised enjoyment of the new mandatory exceptions, thus limiting the effectiveness of the DSM.²⁷ The result of this may indirectly encourage innovators away from the EU to regions where such restrictions are less likely to impact their TDM activities or indeed which are less likely to hinder reliance on lawful exceptions or limitations, which would ultimately restrict TDM innovation across Europe.

2.5 Not all bad...

Despite the issues surrounding Articles 3 and 4 of the DSM Directive, the legislation does go some way towards creating a stable environment in Europe for TDM operations. For example, a key benefit of the DSM Directive's provisions related to TDM is that for both scientific and non-scientific TDM activities, the provisions impose a mandatory exception or limitation for TDM activities on Member States. It is therefore hoped the drafting of the DSM Directive will reduce fragmentation in the approach and application of national laws for TDM activities from one Member State to another which would not only create much-welcomed certainty for the relevant actors, but would also promote the EU's policy goals for its Digital Agenda, namely to provide a normalised, consistent level playing field across Europe to legally carry out TDM projects.²⁸

A further positive aspect of the DSM Directive's provisions on TDM is that it is the first occasion, at an EU-level, that TDM has been expressly recognized and codified. This evidences a recognition by the EU of innovative and valuable technological tools and mechanisms within the world of data analytics and raises awareness of TDM as a process of harnessing the value of big data. This encourages innovators AI and machine-learning development and encourages TDM in the region, given that such activities have formal recognition, in respect of copyright implications, in official European texts.

Additionally, in respect of TDM for scientific research, Article 7(1) expressly provides for the unenforceability of contrary contractual provisions. In practice, it would not be possible for copyright holders to expressly exclude the application of Article 3 through contract.

In any case, the prohibition of contractual override is crucial so as not to deprive the provision of any practical utility. Previously, copyright holders, such as publishers of scientific research, would have been able to contractually exclude TDM in licence agreements whilst applying high transaction costs on TDM operators to obtain consent to mine content for research.

In light of these considerations, the authorisation for contractual override in respect of Article 4 of the DSM Directive is arguably of limited consequence. Nevertheless, the prohibition to such contractual override is a welcome step towards a favourable environment for EU-based scientific research activities involving TDM, although, as discussed below, this leaves much to be desired for other TDM activities, despite the potential contract law limitations.

Finally, the TDM exceptions and limitations in the DSM Directive create a stronger environment for rightsholders. The drafting of the provisions sets out a strict framework providing for specific instances in which TDM may be lawfully operated on copyrighted source data, without consent of the rightsholder. The specific and stringent requirements which must be met in order for the beneficiaries of these exceptions and limitations to rely on them is wholly justified in order to create a safe environment for rightsholders, and to protect them against unjustified or unfair exploitation of their works.

²¹ Recital 12 of the DSM Directive lists: "universities or other higher education institutions and their libraries, also entities such as research institutes and hospitals that carry out research".

²² Recital 12 of the DSM Directive.

²³ Geiger C., Frosio G., Bulayenko O., "Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data? Legal Analysis and Policy Recommendations, IIC (2018) 49: 814. <https://doi.org/10.1007/s40319-018-0722-2> (5 December 2019), p. 836.

²⁴ Ibid.

²⁵ Legally Speaking: "The EU's Controversial Digital Single Market Directive", Communications of the ACM, Viewpoints, Pamela Samuelson, November 2018, Vol. 61, No. 11, <https://cacm.acm.org/magazines/2018/11/232195-the-eus-controversial-digital-single-market-directive/fulltext>, (5 December 2019), p. 22.

²⁶ Recital 7 of the DSM Directive.

²⁷ Geiger C., Frosio G., Bulayenko O., "Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of

Big Data? Legal Analysis and Policy Recommendations, IIC (2018) 49: 814. <https://doi.org/10.1007/s40319-018-0722-2> (5 December 2019), p. 838.

²⁸ Geiger et al, "The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects, In-depth Analysis", Policy Department for Citizens' Rights and Constitutional Affairs, European Parliament, February 2018, PE 604.941, section 2, p. 20.

2.6 Reflecting on the new dual-regime approach

The TDM provisions are a step in the right direction,²⁹ but the situation remains uncertain for innovators and researchers alike. The DSM Directive's TDM provisions demonstrate the EU's lack of understanding of the technical realities of TDM, and the financial and contractual limitations which are faced by those performing such processes.

By applying the caveats and restrictions to the TDM provisions in the DSM Directive, and establishing a difference in treatment between research and commercial TDM activities, rather than creating an environment of innovation and collaboration favourable to innovation, the EU has

“overlooked the fact that TDM is not about displacing existing content but rather extracting further knowledge from it and, in doing so, rendering it more valuable.”³⁰

Indeed, it is the aggregated result of TDM which is of value and interest and not the single copyright protected work. The EU does not seem to have understood the technical tool TDM actually is, but rather see it as a traditional reproduction/re-use of the rightholders work, as we have seen with online piracy, unlawful downloading and copies of individual works. In overlooking this fact the EU fails to adequately legislate for this technical tool in an appropriate and nuanced way.

Through these provisions, the EU has made it more difficult for such valuable output to originate from European-based operators. To drive innovation in the EU and further the commitment to the DSM, it would have been more desirable for the EU to implement a broad and all-encompassing TDM exception which is not capable of override,³¹ as is the case in other key innovative jurisdictions.

2.7 The way forward: National transposition and alternatives for the future

In the highly competitive global market for world-class AI and data science researchers, the EU may suffer from “brain drain” if its most talented researchers take job opportunities in jurisdictions where TDM is subject to fewer restrictions.³² Perhaps this task falls to the Member States

as they transpose the provisions of the DSM Directive into their respective national laws, something which must be done by 7th June 2021.

2.7.1 Transposing the TDM exceptions into domestic law

In this respect, it is important to note that the DSM Directive is a harmonization directive. A harmonization directive shall be transposed in accordance with Article 288 third subparagraph TFEU, which requires that:

“A directive shall be binding, as to the result to be achieved, upon each Member State to which it is addressed, but shall leave to the national authorities the choice of form and methods.” [emphasis added]

Arguably, if Member States were to widen the scope of Articles 3 and 4 of the DSM Directive in order to remove the undesirable caveats and loopholes, there is a risk that such Member States would be acting contrary to the requirement of “result to be achieved”. Additionally, if Member States were to deviate from the provisions of Articles 3 and 4 of the DSM Directive as drafted, we would then be faced with an issue of balance regarding the stakeholders on the opposite side of the innovators – the copyright holders – meaning that deviance from the initial drafting by Member States in applying a liberal transposition of the DSM Directive may be faced by challenges from rightholders. This is more so the case since the exceptions found in Articles 3 and 4 of the DSM Directive are so limited in scope that it could be seen as contrary to EU law if Member States were to grant a too wide an exception for TDM. Should a Member State attempt to do this, they would naturally run the risk of being targeted by the European Commission for unsatisfactory transposition of EU legal obligations, which could result in an infringement procedure and eventually be found as violating EU law by CJEU.³³

Article 3 of the DSM Directive relates to TDM for purposes of scientific research. This exception can already be found in a number of Member States and Nordic countries' domestic copyright laws.³⁴ It is rather the general exception contained at Article 4 of the DSM Directive which is relevant when looking at the digital development in society in general and specifically in the business

²⁹ “Europe Needs A Broad & Mandatory TDM Exception”, Association of European Research Libraries, 13 November 2018 available at: <https://libereurope.eu/blog/2018/11/13/europe-needs-a-broad-mandatory-tdm-exception/> [accessed on 5 December 2019].

³⁰ Ibid.

³¹ Ibid.

³² Legally Speaking: “The EU’s Controversial Digital Single Market Directive”, Communications of the ACM, Viewpoints, Pamela Samuelson, November 2018, Vol. 61, No. 11,

<https://cacm.acm.org/magazines/2018/11/232195-the-eus-controversial-digital-single-market-directive/fulltext>, (5 December 2019), p. 23

³³ In accordance with Article 258 and Article 260 TFEU respectively.

³⁴ See e.g. Section 1-3 of the Norwegian Act No. 120 of December 22, 2018, on Amendments to the Copyright Act, etc. (Portability of Online Content Services, etc.); Section 29A(2) of the (UK) Copyright Designs and Patents Act 1988; and Article L.122-5-10 of the French Intellectual Property Code.

³⁵ At least not found in Norwegian, French or UK legislation.

³⁶ Legally Speaking: “The EU’s Controversial Digital Single Market Directive”, Communications of the ACM, Viewpoints, Pamela Samuelson, November 2018, Vol. 61, No. 11, <https://cacm.acm.org/magazines/2018/11/232195-the-eus-controversial-digital-single-market-directive/fulltext>, (5 December 2019), p. 22.

³⁷ TFEU Article 267a.

³⁸ Analytical table comparing different TDM legislation, White Paper — Open Science in a

sector. This exception is a foreign construct to many Member States' copyright laws.³⁵

Even if one applies a liberal interpretation of Article 4 to the restrictions to the commercial utilization of TDM, is the impact as significant in reality? Arguably, the research sector is more heavily affected by the requirement of lawful access and the potential issues related to the licensing of content to be mined since commercial TDM is often focused in areas of online analytics (such as retail analytics) which are related to consumer movements and trends gained through the use of cookies, plug-ins or social media. However, the risk of copyright infringement on copyrighted portions of big data will still be felt by innovators, especially start-ups which have limited access to financial resources.³⁶

Ultimately, due to the fact that the EU legislator wished to impose a number of restrictions on TDM for commercial purposes, that must be taken into account when the Member State legislators are reviewing the "result to be achieved" when transposing the DSM Directive into domestic law. A development of the exceptions in Member State jurisdictions by national courts or government authorities can be acceptable, but is also restricted by the material content of the EU law being applied. There are two points that must be made in relation this type of development of EU law.

First, the exceptions found in Articles 3 and 4 are quite clear. The wording of the exceptions does not leave much scope for interpretation and therefore a wider interpretation would more likely be found of being in violation of the legal obligation flowing from the DSM Directive. Second, the CJEU is the ultimate interpreter of EU law.³⁷ Based on these considerations, it is clear that domestic courts must walk a thin line if they aim to broaden the scope of the exceptions for TDM.

In our view, the bottom line is that "optimal" transposition of the exceptions for TDM will be unlikely contribute to the state of the art of TDM, unless the Member States wants to run the risk of violating EU law. Could a brave Member State, such as e.g. the IT-focused state of Estonia, be a candidate for this task? For example, we could envisage a mischievous interpretation, aligning exceptions and limitations with a general US-style fair use doctrine, set out in 17 U.S. Code § 107, to encompass TDM activities for all

purposes. Alternatively, a bold step towards a Japanese-style exception which even from 2009, authorised broad TDM activities through the creation of an exception to a copyright holder's exclusive rights, for information analysis, comparison or classification or statistical analysis, with no restriction on beneficiaries.³⁸ The Japanese TDM legislation was further updated in 2019 to permit additional flexibility and legal certainty for innovators, and to enhance the already TDM-favourable environment. It addresses the potential risks that copyright poses for innovation by permitting all users the right to: (i) analyse and understand copyrighted works for machine learning purposes;³⁹ (ii) make and retain incidental electronic copies of works;⁴⁰ and (iii) use copyrighted works for data verification.⁴¹

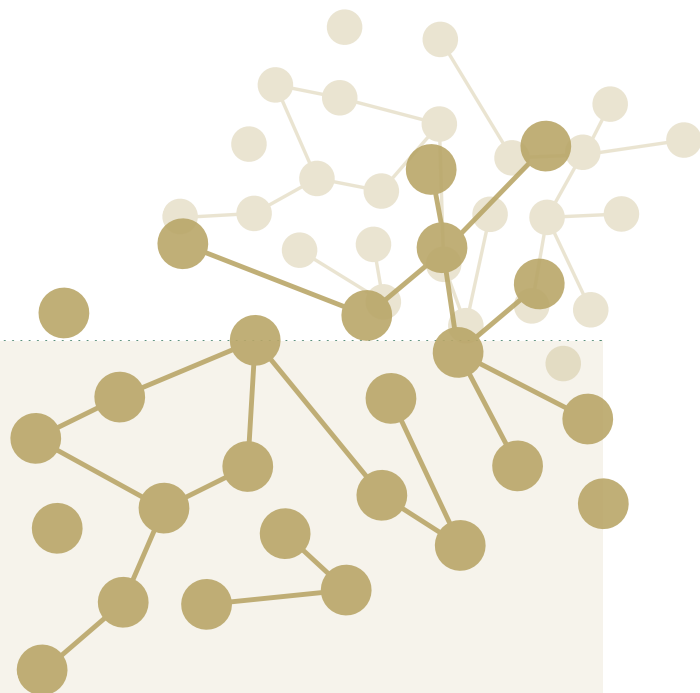
On the other hand, are there obvious limits to what such an act of mischievous transposition act by a Member State would bring about? The CJEU would also be bound by the clear wording of the DSM Directive's TDM exceptions, and any modification to Articles 3 and 4 rather be expected to occur at the EU legislative level. Furthermore, the question of what an optimal transposition of EU law is would depend on whom one wants it to be optimal for. Here, we focus on promoting European innovators in order to further the Digital Agenda. Thus, if you take the position of the developers and users of TDM technology, a wider exception would naturally be welcomed. The optimal transposition for the rightsholder is securing the highest degree of copyright protection for him/her. In the case of TDM exceptions under the DSM Directive, it appears that the interests of the rightsholders have been taken into account to a much larger extent than the TDM innovators which, while being encouraging and reassuring for creators of original content in Europe, seems less-aligned to the digital agenda of the European Commission and the drive to make the EU a leading region at the forefront of global innovation.

Digital Republic — Strategic Guide, Open Edition Press, 2017. For the official provision (in Japanese) visit: http://www.mext.go.jp/b_menu/houan/kakutei/detail/1405213.htm [accessed on 5 December 2019].

³⁹ Article 30-4 of the Japanese Copyright Act (2018 Amendment) For the official provision visit: http://www.mext.go.jp/b_menu/houan/kakutei/detail/1405213.htm (in Japanese only).

⁴⁰ *Ibid*, at Article 47-4.

⁴¹ *Ibid*, at Article 47-5.



2.7.2 To make peace with the current legislation

While it may have been desirable for the EU legislature to implement a TDM exception akin to the provisions that one can find in Japan, or by creating a fair use/fair dealing model for TDM (as can be seen in other innovative economies such as US, Canada, Israel),⁴² we now have the DSM Directive, which while not perfect for ensuring the furtherance of innovation, it does bring about a start in regards to TDM as it is.

It is important for Member States and innovators to maximise the benefits of this legislation, even with its caveats and restrictions. This is even more so given the time that it takes for the EU to legislate in such areas – the last copyright reform for the digital environment took place in 2001⁴³ and we have now been presented with the DSM Directive almost two decades later. Even in Japan, it took a decade for the already-favourable TDM provisions to be revised in accordance with current technological advances. It therefore seems that we will need to make peace with the EU's current TDM provisions, as any further changes are unlikely to be forthcoming in the immediate to near future. This is the classic challenge with law and the speed at which technology progresses – even during the time of the legislative process, and then the implementation of EU laws by Member States, the legal provisions that have been so hotly debated and carefully drafted can quickly become redundant or out of date. The DSM Directive is not to only act facing this dilemma.

It therefore falls on Member States' national laws, courts and practitioners to find appropriate and innovative ways to apply law to new facts and circumstances which arise faster than the creation of new or updated legislation. This will be a difficult task when keeping in mind the lack of margin for manoeuvre when transposing Articles 3

and 4 of the DSM Directive into domestic laws, as pointed out above. Indeed, while a broad transposition of the DSM Directive may be a risky step for Member States, we can look to practice in Europe and hope that national courts and stakeholders take a sensible and purposive approach to TDM and copyright under Articles 3 and 4 of the DSM Directive, so that innovation is neither hindered nor prohibited in the region.

3. CONCLUSION

There are both positive and negative aspects to the TDM exceptions in the DSM Directive which indicate how the EU has approached technological advances and innovation generally through legislation, case law and for TDM specifically.

On one hand, the EU has a strong commitment to the Digital Agenda and wishes to push the EU to the forefront on a global scale but has failed to create an all-encompassing copyright framework for TDM. Given the nature of TDM, the value is in the collection of several sources of work gathered from big data – there is no single victim of copyright infringement because there is little value in one single piece of work or one sole extract. Arguably, the value gained from TDM arises from the analysis of several works or extracts when taken together as a collective, with minimal human intervention or harm to a rightsholders moral or economic rights.

However, it would be unjust to state that the EU approach to TDM and copyright constitutes a full hindrance to innovation in the region as there are several factors which must work together to create a positive environment for start-ups and innovators within the research and non-research sectors, such as access to funding, available

⁴² Geiger C., Frosio G., Bulayenko O., "Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data? Legal Analysis and Policy Recommendations, IIC (2018) 49: 814. <https://doi.org/10.1007/s40319-018-0722-2> [5 December 2019], p. 835.

⁴³ The InfoSoc Directive.

⁴⁴ Maryam Mazraei, Founder of Autopsy (<https://www.getautopsy.com/>), interview of 5 April 2019.

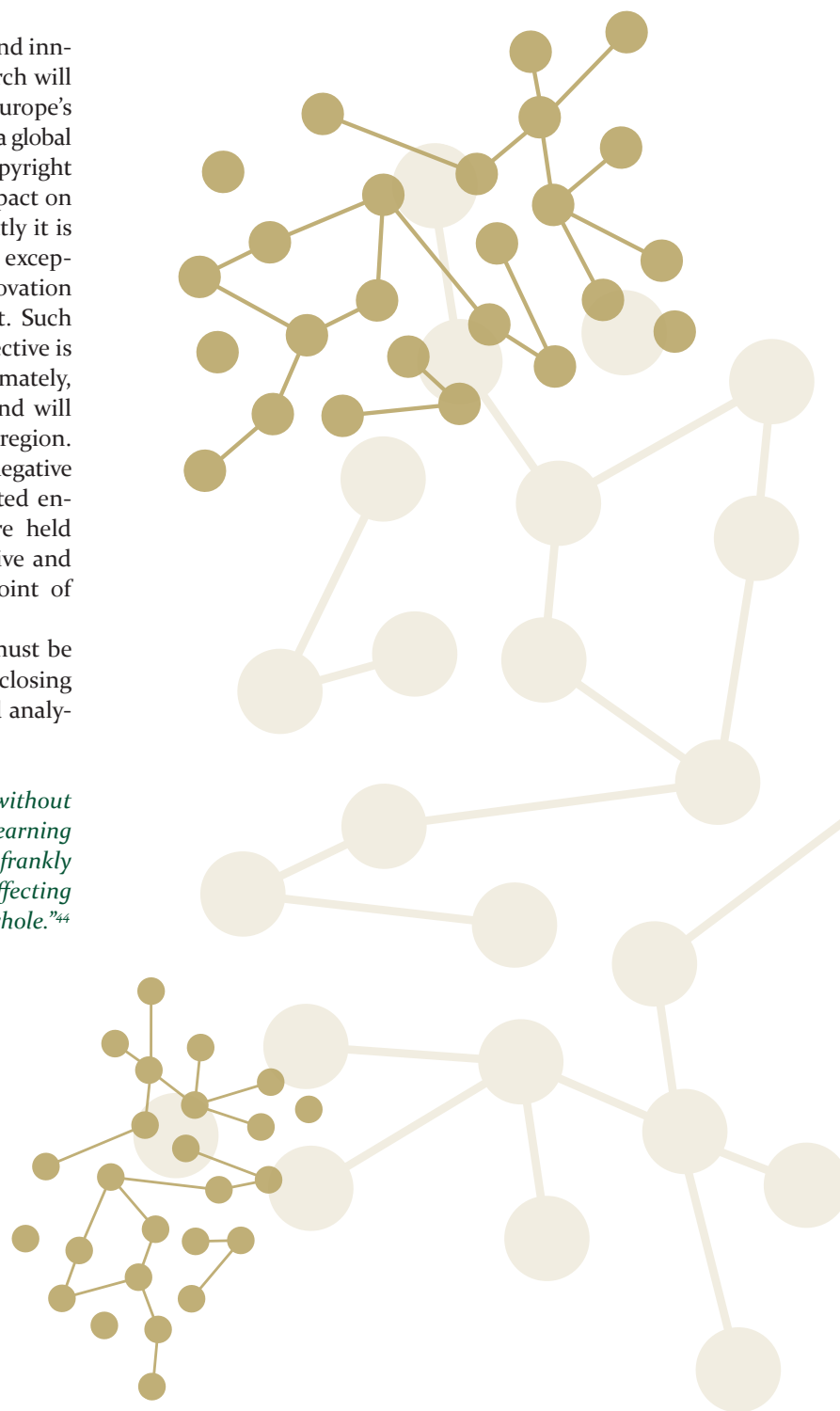
talent and knowledge sharing, as well as an appropriate copyright framework.

At the time of writing, the exact future of TDM and innovation in Europe is unknown, and further research will be required over the next few years to ascertain Europe's market share for TDM, AI and machine learning on a global scale. An analysis as to whether the TDM copyright environment in Europe has had a measurable impact on Europe's success must also be conducted - currently it is hard to confirm whether the TDM limitations and exceptions will be a furtherance or a hindrance to innovation within the EU as the DSM Directive is so recent. Such issues must be monitored closely as the DSM Directive is transposed into Member State's national laws. Ultimately, Europe's approach to copyright alone has not, and will not, continue to be a threat to innovation in the region. Such impact, to the extent it exists in a positive or negative sense, is likely a result of the EU's highly regulated environment generally, where individual rights are held above those of start-ups or tech-giants - a positive and negative consequence depending on the viewpoint of each stakeholder.

What is clear from this article is that Europe must be mindful to the future and to our innovators. In a closing remark from the Founder of a UK-based data and analytics company:

*"To not have the freedom to access information without infringing on IPRs data science and machine learning would be detrimental to our business and quite frankly stop, or make innovation extremely hard, thus affecting the European tech and start-up economy as a whole."*⁴⁴

Something which surely the EU wishes to avoid.



Charlotte Gerrish

Founding Lawyer of Gerrish Legal. Solicitor of England & Wales, and Ireland and registered at the Paris Bar (France). Charlotte specializes in IP/IT, privacy and commercial law and recently obtained an LLM-degree in Intellectual Property Law from Uppsala University specializing in copyright issues and technology.



Anders Skavlan

Legal Consultant at Gerrish Legal. Working within the field of commercial law and privacy law. Finalizing a LLM-degree in European and international law from Europa-Institut in Saarbrücken as well as a Master of Laws degree from University of Bergen.