



UvA-DARE (Digital Academic Repository)

Ensuring the Visibility and Accessibility of European Creative Content on the World Market

The Need for Copyright Data Improvement in the Light of New Technologies and the Opportunity Arising from Article 17 of the CDSM Directive

Senftleben, M.; Margoni, T.; Antal, D.; Bodó, B.; van Gompel, S.; Handke, C.; Kretschmer, M.; Poort, J.; Quintais, J.; Schwemer, S.

DOI

[10.2139/ssrn.3785272](https://doi.org/10.2139/ssrn.3785272)

Publication date

2022

Document Version

Final published version

Published in

Journal of Intellectual Property, Information Technology and Electronic Commerce Law

License

Other

[Link to publication](#)

Citation for published version (APA):

Senftleben, M., Margoni, T., Antal, D., Bodó, B., van Gompel, S., Handke, C., Kretschmer, M., Poort, J., Quintais, J., & Schwemer, S. (2022). Ensuring the Visibility and Accessibility of European Creative Content on the World Market: The Need for Copyright Data Improvement in the Light of New Technologies and the Opportunity Arising from Article 17 of the CDSM Directive. *Journal of Intellectual Property, Information Technology and Electronic Commerce Law*, 13(1), 67-86. <https://doi.org/10.2139/ssrn.3785272>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Ensuring the Visibility and Accessibility of European Creative Content on the World Market

The Need for Copyright Data Improvement in the Light of New Technologies and the Opportunity Arising from Article 17 of the CDSM Directive

by **Martin Senftleben, Thomas Margoni, Daniel Antal, Balázs Bodó, Stef van Gompel, Christian Handke, Martin Kretschmer, Joost Poort, João Quintais and Sebastian Schwemer**

Abstract: In the European Strategy for Data, the European Commission highlighted the EU's ambition "to acquire a leading role in the data economy." At the same time, the Commission conceded that the EU would have to "increase its pools of quality data available for use and re-use." In the creative industries, this need for enhanced data quality and interoperability is particularly strong (section A). Without data improvement, unprecedented opportunities for monetising the wide variety of creative content in EU Member States and making this content available for new technologies, such as artificial intelligence ("AI") systems, will most probably be lost (section B). The problem has a worldwide dimension. While the US have already taken steps to provide an integrated data space for music as of 1 January 2021, the EU is

facing major obstacles not only in the field of music but also in other creative industry sectors (section C). Weighing costs and benefits (section D), there can be little doubt that new data improvement initiatives and sufficient investment in a better copyright data infrastructure should play a central role in EU copyright policy. The work notification system following from Article 17(4)(b) of the Directive on Copyright in the Digital Single Market may offer an unprecedented opportunity to bundle and harmonize data in a shared EU copyright data repository (section E). In addition, a trade-off between data harmonisation and interoperability on the one hand, and transparency and accountability of content recommender systems on the other, may pave the way for new initiatives (section F).

Keywords: copyright; metadata; cultural diversity; licensing; enforcement; recommender systems; text and data mining

© 2022 Martin Senftleben, Thomas Margoni, Daniel Antal, Balázs Bodó, Stef van Gompel, Christian Handke, Martin Kretschmer, Joost Poort, João Quintais and Sebastian Schwemer

Everybody may disseminate this article by electronic means and make it available for download under the terms and conditions of the Digital Peer Publishing Licence (DPPL). A copy of the license text may be obtained at <http://nbn-resolving.de/urn:nbn:de:0009-dppl-v3-en8>.

Recommended citation: Martin Senftleben, Thomas Margoni et al., Ensuring the Visibility and Accessibility of European Creative Content on the World Market - The Need for Copyright Data Improvement in the Light of New Technologies and the Opportunity Arising from Article 17 of the CDSM Directive 13 (2022) JIPITEC 67 para 1

A. Introduction

1 Since the early days of the digital revolution, the dream of the free flow of information across cultures and continents has been accompanied by the hope that digital rights management (“DRM”) in the area of copyright would maximise the spectrum of available literary and artistic productions (including content for niche audiences), minimise transaction costs, pave the way for ubiquitous and differentiated licensing solutions and allow creative industries to thrive. In reaction to the challenges arising from the digital environment, the 1996 WIPO “Internet” Treaties¹ introduced new international standards against the circumvention of technological measures that are employed to protect copyrighted works, and the removal or alteration of copyright management information.² The 2001 Directive on the Harmonisation of Copyright and Related Rights in the Information Society (“Information Society Directive” or “ISD”)³ transposed these international standards into EU law.

* Martin Senftleben, Professor of Intellectual Property Law and Director, Institute for Information Law (IVIIR), University of Amsterdam, The Netherlands; Of Counsel, Bird & Bird, The Hague; Thomas Margoni, Research Professor of Intellectual Property Law, Centre for IT & IP Law (CiTIP), Faculty of Law, KU Leuven, Belgium; Daniel Antal, Independent Researcher, The Hague, The Netherlands; Balázs Bodó, Associate Professor, Institute for Information Law (IVIIR), University of Amsterdam, The Netherlands; Stef van Gompel, Professor of Intellectual Property, Vrije Universiteit Amsterdam; Associate Professor, Institute for Information Law (IVIIR), University of Amsterdam, The Netherlands; Christian Handke, Associate Professor of Cultural Economics, Erasmus University Rotterdam, The Netherlands; Martin Kretschmer, Professor of Intellectual Property Law and Director, CREATE, University of Glasgow, United Kingdom; Joost Poort, Associate Professor and Co-Director, Institute for Information Law (IVIIR), University of Amsterdam, The Netherlands; João Quintais, Assistant Professor, Institute for Information Law (IVIIR), University of Amsterdam, The Netherlands; Sebastian Schwemer, Associate Professor, Centre for Information and Innovation Law (CIIR), University of Copenhagen, Denmark; Adjunct Associate Professor, Norwegian Research Center for Computers and Law (NRC-CL), University of Oslo, Norway.

1 WIPO Copyright Treaty and WIPO Performances and Phonograms Treaty, adopted in Geneva on December 20, 1996.
 2 Articles 11 and 12 of the WIPO Copyright Treaty; Articles 18 and 19 of the WIPO Performances and Phonograms Treaty.
 3 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, *OJ* 2001 L 167, 10.

2 Besides applications by individual companies, the issue of copyright data management—in the sense of attaching and standardising metadata to works stemming from various authors and producers—has traditionally played a crucial role in the area of collective licensing of creative content. Nowadays, content distribution platforms that operate internationally, such as Spotify, Apple Music, YouTube, Netflix and Getty Images, play a central role as well. With Article 17 of the Directive on Copyright in the Digital Single Market (“Digital Single Market Directive” or “CDSMD”),⁴ the topic receives an important additional dimension.⁵ Article 17 addresses specifically online platforms that allow users to upload and share user-generated content (“UGC”). The collaboration between the creative industry and these platforms—Online Content Sharing Service Providers (“OCSSPs”)⁶—has already led to the creation of content identification systems (and corresponding databases) in the past, and can be expected to foster the establishment of more extensive content libraries and corresponding metadata for the purposes of online content identification and moderation in the future.⁷ Article 17(4)(b) CDSMD requires OCSSPs to make

“in accordance with high industry standards of professional diligence, best efforts to ensure the unavailability of specific works and other subject matter for which the rightholders

4 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, *OJ* 2019 L 130, 92.

5 For a more detailed discussion of the relationship between the data economy and the creative industries depending on copyright, see Valérie-Laure Benabou, in collaboration with Célia Zolynski and Laurent Cytermann, *Droit de la propriété littéraire et artistique, données et contenus numériques*, Paris: CSPLA 2018.

6 See the definition in Article 2(6) CDSMD. For a more detailed discussion, see Axel Metzger and Martin R.F. Senftleben, “Comment of the European Copyright Society: Selected Aspects of Implementing Article 17 of the Directive on Copyright in the Digital Single Market into National Law”, *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 10 (2020), 115–131.

7 For a proposal to use Article 17 CDSMD as a catalyst to build a public repository of public domain works and openly licensed works, see Julia Reda and Paul Keller, “A Proposal to Leverage Article 17 to Build a Public Repository of Public Domain and Openly Licensed Works”, *Kluwer Copyright Blog*, 23 September 2021, available at: <http://copyrightblog.kluweriplaw.com/2021/09/23/a-proposal-to-leverage-article-17-to-build-a-public-repository-of-public-domain-and-openly-licensed-works/>.

have provided the service providers with the relevant and necessary information;...”

3 Evidently, this provision imposes more than a content moderation obligation on OCSSPs.⁸ At the

8 As to the underlying debate on new licensing and content moderation obligations, see Axel Metzger and Martin R.F. Senftleben, “Understanding Article 17 of the EU Directive on Copyright in the Digital Single Market – Central Features of the New Regulatory Approach to Online Content-Sharing Platforms”, *Journal of the Copyright Society of the U.S.A.* 67 (2020), 279 (284-308); Christophe Geiger and Bernd Justin Jütte, “Platform liability under Article 17 of the Copyright in the Digital Single Market Directive, Automated Filtering and Fundamental Rights: An Impossible Match”, *Gewerblicher Rechtsschutz und Urheberrecht International* 70 (2021), 517; Sebastian Felix Schwemer, “Article 17 at the Intersection of EU Copyright Law and Platform Regulation”, *Nordic Intellectual Property Law Review* 2020, 400-435; Martin R.F. Senftleben, “Institutionalized Algorithmic Enforcement – The Pros and Cons of the EU Approach to Online Platform Liability”, *Florida International University Law Review* 14 (2020), 299-328; Martin Husovec and João Pedro Quintais, “How to License Article 17? Exploring the Implementation Options for the New EU Rules on Content-Sharing Platforms under the Copyright in the Digital Single Market Directive”, *Gewerblicher Rechtsschutz und Urheberrecht International* 70 (2021), 325 (325-348); João Pedro Quintais, Giancarlo Frosio, et al. “Safeguarding User Freedoms in Implementing Article 17 of the Copyright in the Digital Single Market Directive: Recommendations from European Academics”, *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 10 (2020), 277-282; Giancarlo Frosio, “Reforming the C-DSM Reform: A User-Based Copyright Theory for Commonplace Creativity”, *International Review of Intellectual Property and Competition Law* 51 (2020), 709 (724-726); Sebastian Felix Schwemer and Jens Schovsbo, “What is Left of User Rights? – Algorithmic Copyright Enforcement and Free Speech in the Light of the Article 17 Regime”, *Intellectual Property Law and Human Rights*, 4th ed., Alphen aan den Rijn: Wolters Kluwer 2020, 569-589; Martin R.F. Senftleben, “Bermuda Triangle: Licensing, Filtering and Privileging User-Generated Content Under the New Directive on Copyright in the Digital Single Market” *European Intellectual Property Review* 41 (2019), 480 (483-484); Martin R.F. Senftleben, Christina Angelopoulos, et al., “The Recommendation on Measures to Safeguard Fundamental Rights and the Open Internet in the Framework of the EU Copyright Reform”, *European Intellectual Property Review* 40 (2018), 149; Christina Angelopoulos, “On Online Platforms and the Commission’s New Proposal for a Directive on Copyright in the Digital Single Market”, 2017, available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2947800; Giancarlo Frosio, “From Horizontal to Vertical: An Intermediary Liability Earthquake in Europe”, *Oxford Journal of Intellectual Property and Practice* 12 (2017), 565-575; Giancarlo Frosio, “Reforming Intermediary Liability in the Platform Economy: A European Digital Single Market Strategy”, *Northwestern University Law Review* 112 (2017), 19;

same time, it gives copyright and neighbouring right holders a strong incentive to provide work-related data, including accurate and up-to-date ownership information. A rightholder who does not provide “the necessary and relevant information” in the sense of Article 17(4)(b) CDSMD cannot benefit from the new content moderation obligation. As a result, infringing user uploads may become available on online platforms.

4 With this new incentive scheme for notifying work-related data, Article 17(4)(b) CDSMD may play a central role as a catalyst bringing about new pan-European copyright data repositories, at least for content shared on OCSSP platforms. Such a catalyst seems crucial. In theory, the digital environment offers unprecedented opportunities for commercialising literary and artistic productions and serving consumers. To this day, however, several practical problems have prevented the creative industries from realising the full potential of copyright data management and digital modes of exploitation. The lack or inaccuracy of metadata prevents or delays the disbursement of royalties. Moreover, inaccurate and incomplete metadata make content hard to find, or license, and, as a result, may contribute to digital piracy.

5 From an economic perspective, it may be said that even if certain content is technically available via legal channels, inaccurate and incomplete metadata may increase search costs for users to such an extent that data problems de facto create incentives to make unauthorised use where copyright enforcement is weak. Alternatively, potential uses of works may simply be forgone due to such transaction costs. In addition to these problems at the level of individual data sets, the lack of interoperability between data management systems and related data libraries forces stakeholders to deal with a highly inefficient, and often inaccurate, piecemeal network of data providers, systems, datasets and standards. It increases all types of transaction costs because it obliges stakeholders to learn about, identify, and deal with various types of metadata, as well as individual terms and modalities of use. The high costs of dealing with inaccurate and incomplete metadata may moreover favour big providers of copyright-intensive products and services who can afford to invest in database building, data cleansing, and who are capable of bearing the costs of lawsuits arising from data-related conflicts. This enhances the risk of economic concentration in the digital content distribution market and a corresponding

Reto M. Hilty and Valentina Moscon (eds.), *Modernisation of the EU Copyright Rules – Position Statement of the Max Planck Institute for Innovation and Competition*, Max Planck Institute for Innovation and Competition Research Paper No. 17-12, Max Planck Institute for Innovation and Competition: Munich 2017.

power imbalance between copyright holders and content distributors, such as online platforms.

B. Need for Improved Copyright Data Management

6 Emerging new technologies that require the use of large repertoires of creative content shed light on the dimension of transaction cost problems in the creative industries and the risk of losing substantial revenue. The situation in the field of AI systems can serve as an example. For a long time, mankind assumed that only humans were capable of creating literary and artistic works. With developments in the field of AI giving birth to a new kind of algorithmic work creation in the realm of cultural creativity, this assumption no longer seems valid. Today, AI systems increasingly assist in the creation of works of art and literature (“AI-assisted works”). Sometimes, on the basis of appropriate training material, they may also be capable of mimicking human literary and artistic productions, such as poems, music and paintings (“AI-generated works”).⁹ The technology

enabling their creative functions is becoming more and more advanced and instead of fully relying on human instructions, contemporary AI systems are becoming increasingly autonomous. Certain types of deep-learning systems may give users the impression of being capable of cultural creation, potentially almost independently, allowing for broad-scale production of cultural objects that eye and ear often fail to distinguish from human creations.¹⁰

7 In this context, however, it must not be overlooked that “artificial creativity” is impossible without source material in a harmonised and interoperable format that can be used for feeding and instructing AI systems. Without machine-readable literary and artistic input stemming from authors of flesh and blood, an AI system has no template for its own processes of mimicking human creativity. Modern data-driven AI often uses Text-and-Data Mining (“TDM”)¹¹ techniques to extract the data needed for machine learning. TDM has emerged as one of the most powerful digital tools in the AI environment which enables the discovery and extraction of patterns, correlations and more generally of (often hidden) knowledge from existing content and data.¹² Both high-tech and creative industries are currently being revolutionised by the advancements in this data-driven type of AI. Techniques that are currently

9 See Ahmed Elgammal, Bingchen Liu et al., “CAN: Creative Adversarial Networks Generating “Art” by Learning About Styles and Deviating from Style Norms”, June 2017, available at: https://www.researchgate.net/publication/317823071_CAN_Creative_Adversarial_Networks_Generating_Art_by_Learning_About_Styles_and_Deviating_from_Style_Norms, 17 (Elgammal and his fellow researchers carried out an experiment to determine whether humans were capable of distinguishing computer-generated art from human art by its appearance. 75% of the research subjects assumed that the computer-generated paintings were created by a human artist). Cf. Dan Burk, “Thirty-Six Views of Copyright Authorship, by Jackson Pollock”, *Houston Law Review* 58 (2020), 263 (270-321); P. Bernt Hugenholtz and João Pedro Quintais, “Copyright and Artificial Creation: Does EU Copyright Law Protect AI-Assisted Output?”, *International Review of Intellectual Property and Competition Law* 52 (2021), 1190 (1212-1213); Martin R.F. Senftleben and Laurens D. Buijtelaa, “Robot Creativity: An Incentive-Based Neighbouring Rights Approach”, *European Intellectual Property Review* 42 (2020), 797-812; Daniel Gervais, “The Machine as Author”, *Iowa Law Review* 105 (2020), 2053; Jane C. Ginsburg and Luke Ali Budiardjo, “Authors and Machines”, *Berkeley Technology Law Journal* 34 (2019), 343 (395-396); Marie-Christine Janssens and Frank Gotzen, “Kunstmatige Kunst. Bedenkingen bij de toepassing van het auteursrecht op Artificiële Intelligentie”, *Auteurs en Media* 2018-2019, 323 (325-327); William T. Ralston, “Copyright in Computer-Composed Music: HAL Meets Handel”, *Journal of the Copyright Society of the U.S.A.* 52 (2005), 281; Shlomit Yanisky-Ravid and Samuel Moorhead, “Generating Rembrandt: Artificial Intelligence, Copyright and Accountability in the 3A Era”, *Michigan State Law Review* (2017), 659 (662); Annemarie Bridy, “The Evolution of Authorship: Work Made by Code”,

Columbia Journal of Law and the Arts 39 (2016), 395 (397); Robert C. Denicola, “Ex Machina: Copyright Protection for Computer-Generated Works”, *Rutgers University Law Review* 69 (2016), 251.

10 The impact that AI is having in the field of IP, and copyright in particular, has been recognised by the European Commission, which has specifically identified a number of ambitious interventions in this area in its recent “IP Action Plan”, see European Commission, 15 November 2020, *Making the Most of the EU’s Innovative Potential – An Intellectual Property Action Plan to Support the EU’s Recovery and Resilience*, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Document COM(2020) 760 final, 12. See also the report by Alexandra Bensamoun and Joëlle Farchy, in collaboration with Paul-François Schira, *Intelligence artificielle et culture*, Paris: CSPLA 2020.

11 The abbreviation “TDM” is used here for text-and-data mining in accordance with the use that has become customary in the domain of copyright. It is not to be confused with “term document matrix” – an important standard organizational form of data describing natural language texts for NLP algorithms.

12 Thomas Margoni, “Computational Legal Methods: Text and Data Mining in Intellectual Property Research”, in: Irene Calboli and Maria Lillà Montagnani (eds.), *Handbook on Intellectual Property Research*, Oxford: Oxford University Press 2021, 487-505.

discussed under the headings of Machine Learning (“ML”), Natural Language Processing (“NLP”) and Deep Neural Networks (“DNN”) require the “training” on vast amounts of content and data in order to achieve reliable results that may finally lead to new scientific and technological advancements, products and services. This information is often deduced, through automated machine-reading processes from books, magazine articles, music works, or films enjoying copyright protection. Not surprisingly, the insatiable appetite of “creative” AI systems for literary and artistic data input is often regarded as a promising new source of revenue for the creative industries.¹³

- 8 The use of copyrighted works as training material for these types of AI applications, however, raises complex questions. When humans learn a new task or skill (e.g. a new language), they usually store the training information (e.g. the textbook rules and examples used to learn the language) as an electrochemical trace in the area of the brain dedicated to language. Humans do not need a copyright exception in order to store that copy. However, it is far from clear that when a computer makes the corresponding digital copy of training material in order to learn a language—or any other task for that matter—that this activity is likewise excluded from the copyright domain.¹⁴ On the contrary, the use of any digital copy, temporary or permanent, in whole or in part, direct or indirect, may amount to the infringement of the right of reproduction laid down in Article 2 ISD.
- 9 The right of reproduction thus constitutes a pivotal element in AI training processes. ML-based systems may require numerous and different types of reproductions: certain copies may be just temporary (the conversion of .pdf into .xml for annotation and enrichment purposes), others may be permanent (the initial creation of *corpora* or databases of training material, or the final storage of said material for replicability, accountability and verifiability

ity of the training process). Some copies may be in whole (such as the initial reproduction of the *corpora*), while other copies may be in part (such as the information stored in the “trained models” which will be used by the AI algorithm to perform the intended task). Finally, some reproductions may be direct and others may be only indirect (again the final “trained models” may contain only partial and modified copies of the original material). Further steps in the AI training process and the distribution and use of the final outcome may involve additional rights that are exclusively reserved to copyright holders, such as the right of distribution and the right of communication to the public. If no exceptions or limitations permit the use of copyrighted material without authorisation,¹⁵ the current formalistic interpretation that the CJEU embraces, especially in relation to the right of reproduction,¹⁶ points towards the conclusion that all these individual acts of use require licenses.¹⁷

- 10 Against this background, appropriate copyright data management and licensing infrastructures are not only desirable to offer the creative industries the opportunity of exploiting the promising new market for AI training data. Improved copyright data management is also indispensable to enable EU high-tech industries to compete with AI system developers in other regions. In Article 3(1) CDSMD, EU legislation has granted a statutory permission to reproduce literary and artistic works for AI training purposes. This limitation of copyright protection, however, only covers TDM in the context of scientific research carried out by eligible research organisations and cultural heritage institutions.¹⁸ Article 4(1) CDSMD supplements this research privilege with a general TDM exemption that can also be invoked by commercial AI system developers. This broader copyright limitation, however, is only applicable as long as copyright holders refrain from reserving their ex-

13 Cf. Paul Covington, Jay Adams and Emre Sargin, “Deep Neural Networks for Youtube Recommendations”, in: *Proceedings of the 10th Acm Conference on Recommender Systems*, RecSys ’16, New York: Association for Computing Machinery 2016, 191-198, available at: <https://doi.org/10.1145/2959100.2959190>; Kurt Jacobson, Vidhya Murali et al., “Music Personalization at Spotify”, in: *Proceedings of the 10th Acm Conference on Recommender Systems*, RecSys ’16, New York: Association for Computing Machinery 2016, 373, available at: <https://doi.org/10.1145/2959100.2959120>.

14 For a more detailed discussion of this point, see Rossana Ducato and Alain Strowel, “Ensuring Text and Data Mining: Remaining Issues with the EU Copyright Exceptions and Possible Ways Out”, *European Intellectual Property Review* 43 (2021), 322-337.

15 Thomas Margoni and Martin Kretschmer, “A Deeper Look Into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology”, *CREATe Working Paper 2021/7*, Glasgow: CREATe Centre 2021.

16 For a critique of this approach, see Martin R.F. Senftleben, “Flexibility Grave – Partial Reproduction Focus and Closed System Fetishism in CJEU, *Pelham*”, *International Review of Intellectual Property and Competition Law* 51 (2020), 751-769.

17 Cf. Ducato and Strowel, *supra* note 17, 322-337, who propose a different interpretation of the relationship between “right” and “infringement” in the realm of Article 2 ISD relying *inter alia* on the “recognisability” test which the CJEU expressed in its *Pelham* decision (CJEU, 29 July 2019, case C476/17, *Pelham*).

18 Cf. the definition in Article 2(1) and (3) CDSMD.

clusive rights under Article 4(3) CDSMD.¹⁹ The need to obtain licenses for commercial applications is thus the rule in EU copyright law; a use permission without prior rightholder authorisation is the exception. With regard to commercial AI training, Article 5(1) ISD only provides a loophole for TDM processes that keep within the confines of transient, temporary copying.²⁰ This restrictive approach may be insufficient for the needs of high-tech firms focusing on AI development. Considering current industry practices, it seems safe to assume that more than temporary takings from copyrighted source material will be necessary in many cases.

- 11 Main international competitors of the EU have chosen approaches that markedly depart from the focus on copyright licensing adopted in Europe. Countries such as the US, Canada, Singapore, South Korea, Japan, Israel or Taiwan have adopted regulatory measures which, in the natural tension between the protection of investments and the promotion of innovation, have opted for broader copyright limitations arguably favouring the latter over the former. The specific measures that have been adopted in order to gauge the proper balance have evolved from, and thus mirror, the domestic legal culture and characteristics. In the US, for instance, TDM and ML analyses are routinely considered to be transformative uses and as such to constitute fair use which is permissible without the prior authorisation of the rightholder and which does not generate claims for fair compensation. This means that using protected works not as works but as input data to extract in-

formation that will be used to create new knowledge—so called non-consumptive or non-expressive uses²¹—is considered a free activity that does not require licensing efforts. Japan is another interesting example as its copyright law can be considered closer to continental-European models. Instead of a broad standard (i.e. fair use), Japanese copyright legislation provides for a list of exceptions and limitations that resembles to a certain degree the approach taken in Article 5 ISD. Japan has implemented in its copyright legislation a broad TDM exception in 2009. This provision refrains from precluding commercial users from invoking the TDM exception.²² The US and Japan are interesting examples because, while belonging to different copyright traditions, they both have thriving creative and cultural industries as well as a highly competitive high-tech sector in the field of AI.

- 12 Considering this global scenario, it is of particular importance to establish efficient copyright data creation, management and licensing infrastructures, and employ Article 17(4)(b) CDSMD as a tool to amass copyright metadata that can help to achieve this goal. In the current policy debate, creative industry representatives in European countries often express a preference for a restrictive approach that only leaves room for narrow copyright exceptions. They fear that a more flexible solution would allow the high-tech industry to exploit copyrighted source material for AI training purposes without sharing the benefits that accrue from the development of AI products and services on this basis. This approach may disadvantage EU-based high-tech industries in comparison with their peers in other legal systems that are willing to favour the high-tech sector. The need to obtain an authorisation to train AI algorithms on vast amounts of data—including copy-

19 For a discussion of opt-out systems as tools to reduce the impact of use privileges on the commercialisation of the work, see Martin R.F. Senftleben, “How to Overcome the Normal Exploitation Obstacle: Opt-Out Formalities, Embargo Periods, and the International Three-Step Test”, *Berkeley Technology Law Journal Commentaries* 1, No. 1 (2014), 1-19.

20 CJEU, 16 July 2009, case C-5/08, Infopaq/Danske Dagblades Forening, para. 56-58; CJEU 17 January 2012, case C-302/10, Infopaq II, para. 36, 44 and 51-56. Cf. Christophe Geiger et al., “Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?”, *International Review of Intellectual Property and Competition Law* 49 (2018), 814 (814-844); Thomas Margoni, “AI, Machine Learning and EU Copyright Law: Who owns AI?”, *Annali Italiani del Diritto d’Autore, della Cultura e dello Spettacolo XXVII* (2018); 281 (281-304); Rossana Ducato and Alain Strowel, “Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to ‘Machine Legibility’”, *International Review of Intellectual Property and Competition Law* 50 (2019), 649; Eleonora Rosati, “An EU Text and Data Mining Exception for the Few: Would it Make Sense?”, *Journal of Intellectual Property Law and Practice* 13 (2018), 429 (429-430); Andres Guadamuz and Diane Cabell, “Data Mining in UK Higher Education Institutions: Law and Policy”, *Queen Mary Journal of Intellectual Property* 4 (2014), 3 (3-29).

21 Matthew Sag, “Copyright and Copy-Reliant Technology”, *Northwestern University Law Review* 103 (2009), 1607-1682; Ian Hargreaves, *Digital Opportunities – A Review of Intellectual Property and Growth*, London: UK Department for Business, Innovation and Skills, 18 May 2011.

22 The Japanese Copyright Act envisages an exception for TDM that is not limited to non-commercial or to research only purposes, see Article 47-septies Japanese Copyright Act reported and discussed in Lucie Guibault and Thomas Margoni, “Legal Aspects of Open Access to Publicly Funded Research”, in: OECD (ed.), *Enquiries Into Intellectual Property’s Economic Impact*, Chapter: 7, OECD 2015, 373-414, 396 available at: <https://www.oecd.org/sti/ieconomy/intellectual-property-economic-impact.htm>. See also Marco Caspers, Lucie Guibault et al., *Future TDM – Baseline Report of Policies and Barriers of TDM in Europe*, Amsterdam: Institute for Information Law 2016, 75-76; Tatsuhiro Ueno, “The Flexible Copyright Exception for ‘Non-Enjoyment’ Purposes Recent Amendment in Japan and Its Implication”, *Gewerblicher Rechtsschutz und Urheberrecht International* 70 (2021), 145-152.

righted works—constitutes an additional cost factor in the form of transaction costs and licensing fees. When the costs involved are too high, it will negatively impact the ability of the EU’s AI sector to compete on the world market and consequently reduce the potential economic value of licensing content for training purposes.²³ If, however, information on the reservoir of available European training material and copyright holders entitled to grant a license is missing or incomplete, the conclusion of licensing agreements is beyond reach from the outset and the creative industries in the EU will lose income that could have come from the use of copyrighted works for AI training purposes.

- 13 Against this background, the concern must be taken seriously that, despite legislation seeking to ensure revenue streams on the basis of licensing obligations, the creative industries in the EU may fail to reap benefits that could accrue from the use of copyrighted material in the AI sector simply because ownership and repertoire information is not available. In terms of regulatory competition, foreign countries opting for less strict regulatory solutions and less licensing and rights clearance obligations may also appear more attractive to high-tech businesses. The EU may thus be confronted with a double failure of the selected regulatory design: neither new income for creative industries nor sufficient investment in promising new high-tech products and services.
- 14 Appropriate solutions for copyright data creation and management in the EU, however, may change the equation. Enhanced cooperation between high-tech companies and the creative industries on the basis of licensing agreements, mutually-agreed use protocols and safeguards against algorithms that disregard competition and media regulations, may even increase the quality and customisation of AI input. Benefits flowing from enhanced cooperation and better input for AI training may compensate the costs arising from an obligation to obtain licenses while, at the same time, ensuring that the benefits

of copyright-based AI training are fairly shared. For this positive scenario to take shape, however, it is indispensable to have a well-functioning copyright data infrastructure in place that offers comprehensive, up-to-date ownership and repertoire information across EU Member States. As a legislative tool that binds copyright owners in all EU Member States and generates relevant data streams in the whole EU territory, Article 17(4)(b) CDSMD can play an important role in this respect. Imposing the obligation on rightholders to provide “relevant and necessary information” with regard to works and other protected subject matter, Article 17(4)(b) CDSMD creates an important legal mechanism to collect ownership and repertoire information. When notifications of works from all corners of the Union are bundled and harmonised, the resulting overarching database makes works and copyright holders visible. In this way, it offers high-tech companies looking for AI training material valuable information on the spectrum of available works in the EU and a solid basis for identifying relevant rightholders.

- 15 Discussing the need for copyright data creation, improvement and harmonisation, the increasing use of automated content recommender systems must be factored into the equation as well.²⁴ Various providers of digital services, including Spotify,²⁵ Netflix²⁶ and YouTube, employ content recommender systems to a growing extent to recommend copyrighted content to users.²⁷ Copyright data improvement also has an important role to play in relation to these systems, e.g., in relation to the visibility of niche repertoires and the enhancement of cultural diversity.²⁸

24 For a broader discussion of new trends in the use of AI tools, including recommender systems, see Juliette Denis and Joëlle Farchy, *La culture des données: Intelligence artificielle et algorithmes dans les industries culturelles*, Paris: Transvalor - Presses des mines 2020.

25 See <https://towardsdatascience.com/how-spotify-recommends-your-new-favorite-artist-8c1850512af0>.

26 See https://medium.com/@springboard_ind/how-netflix-recommendation-engine-works-bd1ee381bf81.

27 Such system is in the context of the proposed Digital Services Act (“DSA”) defined as “a fully or partially automated system used [by an online platform] to suggest in its online interface specific information to recipients of the service, including as a result of a search initiated by the recipient or otherwise determining the relative order or prominence of information displayed.”, see Article 2(o) Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and Amending Directive 2000/31/EC, COM(2020) 825 final.

28 A different but related issue relates to filter bubbles in the

23 For a critique of the approach taken in the EU, see Christophe Geiger, “The Missing Goal-Scorers in the Artificial Intelligence Team: Of Big Data, the Right to Research and the Failed Text-and-Data Mining Limitations in the CSDM Directive”, in: Martin R.F. Senftleben, Joost Poort et al. (eds.), *Intellectual Property and Sports - Essays in Honour of Bert Hugenholtz*, The Hague/London/New York: Kluwer Law International 2021, 383-394; Christian Handke, Lucie Guibault and Joan-Josep Vallbé, “Is Europe Falling Behind in Data Mining? Copyright’s Impact on Data Mining in Academic Research”, in: Birgit Schmidt and Milena Dobrova (eds.), *New Avenues for Electronic Publishing in the Age of Infinite Collections and Citizen Science: Scale, Openness and Trust - Proceedings of the 19th International Conference on Electronic Publishing*, IOS 2015, 120-130.

Without appropriate metadata that enhance the visibility of European content for automated recommender systems, the lack of niche repertoire recommendations may be due to inaccurate or missing data rather than being the result of a discriminatory mainstream orientation of the content recommender system. In this context, however, the lack of transparency of recommender systems, in particular with regard to the parameters used to select content and target consumers, prevents the identification of data issues and the development of appropriate solutions. The (proposed) legal framework in the EU addresses only certain aspects of this dilemma.²⁹

C. Herculean Task of Copyright Data Improvement

- 16 A scenario with mutual benefits for creative and high-tech industries, however, will only arise if the considerable problems in the field of copyright

context of entertainment recommender systems, see e.g. Martin Koppe, “Do algorithms keep playing the same old song?”, *CNRS News*, 27.11.2021, available at: <https://news.cnrs.fr/articles/do-algorithms-keep-playing-the-same-old-song>.

- 29 The proposed Digital Services Act, *supra* note 27, stipulates in Article 29(1) that very large online platforms using recommender systems “shall set out in their terms and conditions, in a clear, accessible and easily comprehensible manner, the main parameters used in their recommender systems, as well as any options for the recipients of the service to modify or influence those main parameters that they may have made available, including at least one option which is not based on profiling“ within the meaning of Article 4(4) GDPR. For a critique on the proposed opt-out, see also European Data Protection Supervisor, *Opinion 1/2021 on the Proposal for a Digital Services Act*, 10 February 2021, https://edps.europa.eu/system/files/2021-02/21-02-10-opinion_on_digital_services_act_en.pdf. Furthermore, under certain circumstances relating to significant systemic risks, very large online platforms may be obliged to adjust their content recommender systems in line with Article 27(1)(a) DSA. Importantly, however, this transparency and opt-out obligation, within its DSA context, only relates to hosting services. Cf. Article 2(h) DSA. Transparency of copyright recommender systems appears neither to be addressed in the recently proposed *Artificial Intelligence Act* which focusses on high-risk AI systems, see Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), and Amending Certain Union Legislative Acts, COM(2021) 206 final. See Sebastian Felix Schwemer, “Recommender Systems in the EU: from Responsibility to Regulation?”, *FACCTRec Workshop '21*, held from 27 September to 1 October 2021 in Amsterdam, paper available at: <https://ssrn.com/abstract=3923003>.

data creation and management can be overcome. To better illustrate data obstacles in European creative industries, the situation in the music sector can serve as a starting point.

I. Experiences in the Music Industry

- 17 The music segment of the creative industry offers several well-known examples of data infrastructures, such as the Common Information System (“CIS”) of the International Confederation of Societies of Authors and Composers (“CISAC”). With its different nodes in several regions of the world, the CIS-Net system and accompanying standards constitute a global tool seeking to facilitate music licensing and the distribution of revenues.³⁰ In terms of data standardisation, the International Standard Work Code (“ISWC”) of the music publishing industry,³¹ the International Standard Recording Code (“ISRC”) of the recording industry, the Interested Party Information (“IPI”) number, and the International Standard Name Identifier (“ISNI”) offer prime examples of existing initiatives to enable the exchange of accurate data related to the identification of repertoire or related to the mitigation of ex post transaction costs that arise in relation to the operation of licensing agreements.
- 18 At the same time, these examples reveal data deficiencies and interoperability problems arising from different sets of metadata and different approaches to data identification and verification. To this day, initiatives to harmonize ISWC and ISRC metadata and incorporate them into a single, comprehensive database have failed. In the EU, former Commissioner Neelie Kroes launched a working group to stimulate the establishment of a Global Repertoire Database (“GRD”) in 2008. While the working group participants, including producers, collective management organisations (“CMOs”) and distribution platforms, arrived at recommendations on the way

30 See <https://www.cisac.org/What-We-Do/Information-Services/CIS-Net>.

31 ISWC has been developed by CISAC, in collaboration with ISO, as “a unique, permanent, and internally recognized reference number for the identification of musical works”. As an example of a further unique identifier system, see also GRiD (Global Release Identifier) which has been developed by IFPI. Cf. Ariel Katz, “The Potential Demise of Another Natural Monopoly: New Technologies and the Administration of Performing Rights”, *Journal of Competition Law and Economics* 1 (2005), 276.

forward,³² the project was abandoned in 2014.³³ Other unsuccessful attempts include the International Music Joint Venture in 2000, which was formed by several CMOs in Europe and North America, and a project initiated by the World Intellectual Property Organization (“WIPO”) aiming at the establishment of a common rights database in 2011.³⁴

- 19 In the US, by contrast, a new initiative to form a comprehensive database followed from the 2018 Music Modernization Act (“MMA”).³⁵ In Title I, the MMA establishes the Mechanical Licensing Collective (“MLC”) as a one-stop shop for obtaining music licenses. For this new licensing body to function properly, it is necessary to have an authoritative and comprehensive database of music rights in place.³⁶ The MLC seeks to achieve this goal by working closely together with major providers of music streaming services, in particular Apple and Spotify.³⁷ The new licensing hub offers a US-wide platform for licence administration, enforcement and royalty processing as of 1 January 2021.³⁸

32 Cf. Mark Isherwood, “Global Repertoire Database”, presented at: World Intellectual Property Organization, *Enabling Creativity in the Digital Environment: Copyright Documentation and Infrastructure*, WIPO Meeting wipo_cr_doc_ge_11, 13-14 October 2011, Geneva: WIPO 2011, available at: https://www.wipo.int/meetings/en/2011/wipo_cr_doc_ge_11/prov_program.html.

33 Cf. Paul Resnikoff, “Global Repertoire Database Declared a Global Failure”, *Digital Music News*, 10 July 2014, available at: <https://www.digitalmusicnews.com/2014/07/10/global-repertoire-database-declared-global-failure/>; Sebastian Felix Schwemer, *Licensing and Access to Content in the European Union. In Licensing and Access to Content in the European Union: Regulation between Copyright and Competition Law*, Cambridge: Cambridge University Press 2019, 68-73.

34 Schwemer, *supra* note 33, 69-70.

35 House Report 1551, Pub. L. 115-264.

36 Cf. Frank Lyons, Hyojung Sun et al., *Music 2025 - The Music Data Dilemma: Issues Facing the Music Industry in Improving Data Management*, Newport: UK Intellectual Property Office 2019, available at: <https://www.gov.uk/government/publications/music-2025-the-music-data-dilemma>, 34.

37 See <https://www.appleworld.today/blog/2019/11/18/apple-spotify-to-fund-new-music-royalties-collective>.

38 See <https://www.themlc.com/press/mechanical-licensing-collective-begins-full-operations-envisioned-music-modernization-act>. As to the underlying planning and preparations, see U.S. Copyright Office Library of Congress, MLC Comments in Reply to the Designation Proposal of the American Music Licensing Collective, Inc., Docket No. 2018-11, 21, available at: <https://>

- 20 This recent US initiative shows that—despite general metadata infrastructures, such as the CIS-Net system and the ISWC/ISRC standards—a strong need is felt in the music industry to combine, streamline and improve rights databases and establish overarching licensing platforms. New initiatives in Europe point in the same direction.³⁹ The Technical Online Working Group Europe (“TOWGE”) brings together a large group of European CMOs, music publishers and rights agencies developing a digital royalty processing system. TOWGE is based on a small group of direct licensors reporting back to local societies.⁴⁰ An initiative with similar objectives has been taken by the Finnish CMO Teosto. A collaboration between Teosto and the start-up company Mind Your Rights has led to the “Concertify” platform seeking to provide—on top of existing industry structures—an efficient and transparent cross-border copyright licensing system. Concertify allows artists, copyright holders, including CMOs, music publishers and event organisers to interact directly by using modules, such as a module for setlist reporting.⁴¹ With the support of the Slovak Art Council, a collaboration between the CMO SOZA and various stakeholders has led to the creation of a prototype for a comprehensive data and metadata database of the Slovak music repertoire. The consortium also created the prototype of a “Listen Local” recommender system that meets the requirements of the trustworthy AI recommendations of the High-Level Working Group on AI.⁴² The accompanying feasibility study highlighted and quantified the problems that arise from incomplete copyright data in existing databases and commercial AI-solutions. For example, it demonstrated that at least 15% of Slovak, Estonian, Hungarian and Dutch works are unlikely to be ever exploited due to data

[bw-98d8a23fd60826a2a474c5b4f5811707-bwcore.](https://s3.amazonaws.com/photos/Proposed_MLC_-_Reply_Comments.pdf)

s3.amazonaws.com/photos/Proposed_MLC_-_Reply_Comments.pdf.

39 For a discussion of further data integration and harmonisation opportunities in the EU, see Norbert Gronau and Martin Schaefer, “Why Metadata Matters for the Future of Copyright”, *European Intellectual Property Review* 43 (2021), 488-494; Martin Schaefer, “Why Metadata Matter for the Future of Copyright”, *Kluwer Copyright Blog*, 27 November 2020, available at: <http://copyrightblog.kluweriplaw.com/2020/11/27/why-metadata-matter-for-the-future-of-copyright/>.

40 See <https://www.digitalmusicnews.com/2019/07/26/towge-digital-royalty-group/>.

41 See <https://www.mindyourrights.fi/>.

42 See <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>.

problems.⁴³ In the area of standardisation, the work of Digital Data Exchange (“DDEX”) is of particular interest. The DDEX system has continuously been expanded to all aspects of the digital music value chain. At the interface between ISWC and ISRC, it provides linkages between work and recording data.⁴⁴

II. Steps Taken in Other Creative Industry Segments

21 Other sectors of the creative industry are facing similar data problems and have embarked on initiatives for data improvement, harmonisation and combination as well. In the field of book publishing, industry initiatives, such as the establishment of different e-book platforms and catalogues, play an important role. Flickr and Google Images offer a search option for material covered by a creative commons licence.⁴⁵ Another example is the Entertainment Identifier Registry (EIDR), which is a universal unique identifier system for movie and television assets based on DOI technology.⁴⁶ As to standardisation, the International Standard Book Number (“ISBN”), the International Standard Serial Number (“ISSN”) for journals, the International Standard Music Number (“ISMN”) for notated music, and the International Standard Audiovisual Number (“ISAN”) for audiovisual works can serve as examples. Moreover, the standardisation work of the international EDItEUR group—leading to the “ONIX” family of standards⁴⁷—is important in the field of books, e-books and serials.⁴⁸ With regard to the digital environment, the International DOI Foundation provides the aforementioned Digital Object Identifier (“DOI”) services and registration: a technical and social infrastructure for the registration and use of persistent interoperable identifiers for use on digital networks, including identifiers for literary and artistic works.⁴⁹

22 In the area of visual arts, CISAC’s Visual Arts Council has extended its initial work on the right of resale and established an online licensing hub⁵⁰ under the umbrella of the International Council of Creators of Graphic, Plastic and Photographic Arts (“CIAGP”).⁵¹ OnLineArt (“OLA”) is a one-stop shop for obtaining licenses for worldwide online use of works of visual art currently encompassing works of 60,000 artists.⁵² Existing initiatives in the visual arts sector—in particular museums and other cultural heritage institutions digitising works in their holdings—have substantially extended the data coverage of works of fine art; however, the situation in the field of photography and illustrations is much less transparent.⁵³ Major visual arts libraries, such as Getty Images, may consistently use data management tools. The costs of properly documenting individual works, however, may be prohibitively high for smaller providers of photography and illustrations in the light of the low average value of individual works.⁵⁴ In comparison with the status quo reached in the field of music, the process of harmonising, attaching and bundling (meta-)data still seems in its infancy in the area of visual arts.

III. Supportive New Technologies

23 In the discussion on copyright data improvement, it is important to note that the lack of high quality, publicly accessible metadata for copyrighted material also prompted intense innovation among technology developers. Existing initiatives show that new technologies, in particular AI and blockchain, may support the streamlining and improvement of copyright data. The aforementioned Concertify platform, for instance, is the result of a collaboration between Teosto and the start-up company Mind Your Rights. The nucleus of the Concertify system for efficient and transparent cross-border copyright licensing was a setlist app which Mind Your Rights had initially developed for Teosto to facilitate setlist

43 Daniel Antal, *Feasibility Study On Promoting Slovak Music in Slovakia and Abroad*, The Hague: Reprex 2020, available at: https://reprex.nl/publication/listen_local_2020/.

44 See <https://ddex.net/about-ddex/purpose/>.

45 See <https://www.theverge.com/2020/8/31/21408305/google-images-photo-licensing-search-results> (Google Images) and <https://www.flickr.com/creativecommons/> (Flickr).

46 See <https://www.eidr.org/>.

47 See <https://www.editeur.org/8/ONIX/>.

48 See <https://www.editeur.org/2/About/#Intro>.

49 See <https://www.doi.org/>.

50 See <https://www.cisac.org/What-We-Do/Creators-Relations/CIAGP>.

51 See <http://www.ciagp.org/>.

52 See <https://onlineart.info/>.

53 For a closer analysis of the particular situation and dynamics in the visual arts sector, see the report by Tristan Azzi and Yves El Hage, *Les métadonnées liées aux images fixes*, Paris: CSPLA 2021.

54 Cf. Richard A. Posner, “Transaction Costs and Antitrust Concerns in the Licensing of Intellectual Property”, *John Marshall Review of Intellectual Property Law* 4 (2005), 325.

reporting on the basis of blockchain technology.⁵⁵ Similarly, ASCAP, SACEM and PRS launched a partnership⁵⁶ to “prototype a new shared system of managing authoritative music copyright information using blockchain technology.”⁵⁷ The concept of the project is to develop a blockchain-based solution built on IBM’s Hyperledger Fabric that links and manages two standards for copyright-protected content used for music recordings: the International Standard Recording Code (ISRC) and the International Standard Work Code (ISWC). The link between these data would improve royalty matching and licensing. The ultimate goal of the project is to enable a “shared, decentralized database of musical work metadata with real-time update and tracking capabilities.”⁵⁸

- 24 These examples reflect initiatives to employ distributed ledger (blockchain) technology as a technological architecture for creating and operating shared metadata resources in highly fragmented domains of literary and artistic production. The underlying projects seek to recognise and respond to the metadata issues in the area of copyright. The initiatives, however, may stem from tech companies outside the literary and artistic field—a fact that may indicate structural problems preventing the incumbent creative industries from embracing and fully developing the potential of new technologies. Substantial further innovation in the field was clearly limited by the lack of high quality, comprehensive metadata, which prompted some start-ups to experiment with bottom-up, collaborative metadata pooling, similar to the efforts made for establishing Wikidata.⁵⁹

IV. Different Settings for Data Improvement

- 25 The described experiences with existing data infrastructures and current initiatives to arrive at

55 See <https://www.mindyourrights.fi/>.

56 See <https://www.ascap.com/press/2017/04-07-ascap-sacem-prs-blockchain>.

57 See <https://societe.sacem.fr/en/press-resources/per-publication/press-releases/ascap-sacem-and-prs-for-music-initiate-joint-blockchain-project-to-improve-data-accuracy-for-rightsholders>.

58 Id. See also <https://www.ascap.com/press/2017/04-07-ascap-sacem-prs-blockchain>.

59 Cf. Balász Bodó, Daniel Gervais and João Pedro Quintais, “Blockchain and Smart Contracts: The Missing Link in Copyright Licensing?”, *International Journal of Law and Information Technology* 26 (2018), 311-336.

better results shed light on different settings for the improvement and harmonisation of copyright (meta-)data. The initiative to harmonise, combine and enhance the coverage of work-related data may come from different actors in the public and private sphere, and employ different tools of public and private law:

- *legislation*: the MLC, for instance, is the result of US legislation that explicitly mandates the establishment of a nationwide licensing hub for mechanical music rights. In the EU, Article 17(4)(b) CDSMD, indirectly, may have similar effects if the new obligations to license user-uploaded content and exchange work-related data for content moderation purposes leads to shared data standards and content identification libraries. In addition, the 2014 Directive on Collective Management of Copyright and Related Rights (“Collective Rights Management Directive” or “CRMD”)⁶⁰ incentivizes CMOs to cooperate in licensing hubs for multi-territorial licensing of online rights in musical works and adopt voluntary industry standards to improve efficiency in the exchange of data. Any legislation at national or EU level for the improvement of copyright data management, however, must observe Article 5(2) of the Berne Convention for the Protection of Literary and Artistic Works (“BC”), which prohibits subjecting the enjoyment and exercise of copyright to mandatory formalities, such as registration requirements;⁶¹
- *public institutions*: impulses for the further development of the copyright data infrastructure may also arise from non-legislative initiatives taken by national, European or international public bodies. The 2008 GRD working group, for instance, came together under the auspices of former Commissioner Neelie Kroes. WIPO initiated the aforementioned 2011 project for the establishment of a common rights database and has embarked on surveys on voluntary registration systems for copyright and related rights in 2005, 2010 and 2021;⁶²

60 Directive 2014/26/EU of the European Parliament and of the Council of 26 February 2014 on collective management of copyright and related rights and multi-territorial licensing of rights in musical works for online use in the internal market, *OJ* 2014 L 84, 72.

61 For an in-depth analysis of the impact of this international ban on formalities, see Stef van Gompel, *Formalities in Copyright Law: An Analysis of Their History, Rationales and Possible Future*, Alphen aan den Rijn: Kluwer Law International 2011.

62 WIPO Survey of National Legislation on Voluntary Registration Systems for Copyright and Related Rights, prepared by

- *private entities*: the initiatives that have led to TOWGE, the Concertify platform and SOZA's Listen Local platform show that private entities, in particular CMOs, may play a decisive role in the further harmonisation and combination of copyright-related data. In addition, individual companies, such as Apple and Spotify, may obtain a market position that allows them to bring together an unprecedented volume of data and establish de facto data standards with a major impact on the sector. External technology start-ups also invest heavily in solutions based on blockchain or related technologies.

26 For the analysis of copyright data management issues, it is important to bear these different settings in mind. To arrive at a substantial improvement of the copyright data infrastructure, it may be necessary to combine public and private initiatives and seek to offer both legislative and market incentives. The legislation-made MLC initiative in the US, for instance, relies on Apple and Spotify as central sponsors and data providers. A similar, large-scale public/private partnership may be necessary to allow European creative industries to compete at eye level with data and licensing improvement on the other side of the Atlantic.

V. Sector-Specific Stumbling Blocks

27 For the success of European initiatives, however, it is also important to consider potential stumbling blocks and corrosive dynamics which large-scale data improvement projects may unleash in the creative industry sector:

- *rivalry between small and big players*: small players and repertoire holders may perceive the establishment of overarching, comprehensive data infrastructures and licensing hubs in the creative industries as a threat. For example, small European CMOs may fear to be left behind⁶³

the Secretariat, SCCR/13/2, November 9, 2015, available at: https://www.wipo.int/meetings/en/doc_details.jsp?doc_id=52829; WIPO Second Survey on Voluntary Registration and Deposit Systems (2010), available at: https://www.wipo.int/copyright/en/registration/registration_and_deposit_system_03_10.html; and WIPO Survey on Voluntary Copyright Registration Systems: Final Report, prepared by Stef van Gompel and Saule Massalina, Amsterdam, 23 April 2021, available at: https://www.wipo.int/edocs/mdocs/mdocs/en/wipo_crr_ge_2_21/wipo_crr_ge_2_21_report.pdf.

63 The risk of a “de facto copyright register in the hands of dominant platforms” was also identified by Germany in its statement accompanying the Council vote on the CDSM Directive. See Schwemer, *supra* note 8, 400-435.

when major European CMOs take joint initiatives and organise data and licensing processes in a way that enhances the visibility and availability of their content—potentially at the expense of repertoire administered by other CMOs which do not have comparable tools to enhance content visibility and availability.⁶⁴ At the global level, individual companies with considerable market power, such as Apple, Spotify, YouTube and Netflix, may establish individual data standards that require European rightholders to deal with different data systems for the purposes of distributing content and monitoring the volume of use. European artists and music distributors may also fear being left behind. In fact, they may lose visibility and market shares on the world market. With the MMA, the US managed to establish a licensing hub in collaboration with US-based streaming services. If this infrastructure becomes a central data resource in the sector, insufficient weight may be lent to non-US (niche) repertoire;

- *fear of losing traditional gatekeeper position*: in sectors with a less developed data infrastructure, such as the field of visual arts, traditional content gatekeepers—holders of individual work libraries, including CMOs—may feel uneasy about initiatives to systematically attach metadata to copyrighted content and include resulting data sources in a comprehensive database and licensing infrastructure. Once a comprehensive and authoritative platform for rights clearance is in place, traditional “middlemen” in the rights clearance process may fear that they become obsolete. The creation of non-harmonised and non-interoperable coding systems and data silos may be part of a survival strategy seeking to preserve a position on the content market, which a more efficient, overarching system for copyright data management may put at risk;
- *path dependence*: stakeholders are likely to have invested substantially in their own proprietary, and often incompatible (meta-)data systems. This investment in individual data infrastructures causes considerable switching costs in case an overarching, harmonised standard is set. This provides a strong disincentive to support initiatives to establish a common, harmonised data standard that requires changes to pre-existing individual data management systems.

28 This outline of problems arising from data harmonisation and improvement projects sheds light on

64 Cf. Lucie Guibault and Stef van Gompel, “Collective Management in the European Union”, in: Daniel Gervais (ed.), *Collective Management of Copyright and Related Rights*, 3rd ed., Alphen aan den Rijn: Kluwer Law International 2015, 139 (172).

central obstacles to the establishment of integrated data spaces which the European Commission also highlighted in its *European Strategy for Data*.⁶⁵ In this Communication, the Commission referred not only to insufficient data quality and interoperability as problem drivers but also to imbalances in market power, a lack of trust and insufficient economic incentives as obstacles to initiatives seeking to ameliorate and finally overcome the problematic status quo.⁶⁶

D. Costs and Benefits

29 Considering difficulties and obstacles, it becomes apparent that the improvement of the copyright data infrastructure in the EU is not an easy task. As a highly complex endeavour, it can hardly be accomplished without substantial investment in metadata creation and improvement, technical data management infrastructure and harmonisation initiatives. The foregoing analysis already offers first insights into the costs that an initiative to improve copyright data may entail in different creative industry sectors.

I. Considerable Investment Necessary

30 With regard to the overall costs of setting up and maintaining a comprehensive copyright data management system, the music industry examples again provide some indications. Reportedly, the European GRD initiative that had commenced in 2008, finally collapsed after an investment of £8 million because the CMOs involved could no longer agree on the funding of the project.⁶⁷ The MLC project in the US rests on a start-up investment of \$33.5 million.⁶⁸ After the start-up phase, MLC expenditures are ex-

pected to average \$30 million annually and amount to \$227 million from 2021 to 2028.⁶⁹

31 According to these figures, there might be a substantive gap between the investment which interested parties in the EU, such as CMOs, are willing to make, and the budget that would be necessary to establish a comprehensive data infrastructure and, if this is desired, run a licensing hub. Before leaning too heavily on cost estimates made in a US context, however, it is important to note that MLC calculations were based on data input from only two central sources: iTunes (Apple Music) and Spotify. Given the cultural diversity and wide variety of copyright data sources in the EU, a European data integration project (not relying exclusively on US-based Apple and Spotify data) would probably require an even larger investment in the start-up phase and following years.

32 Looking at the visual arts sector, an additional cost dilemma comes to the fore: the individual costs to be made in respect of each individual content item. In the field of photography, for instance, databases would have to contain an extremely high number of works. In many cases, these works will have a relatively low average licensing value. This constellation raises the problem that, even if a harmonised data format and a central data recording system become available, the required investment in metadata entry and maintenance may still not come forward because the revenue accruing from visibility and “findability” in the comprehensive database can hardly be expected to outweigh the costs of data entry. The expected market value does not justify the time and money that would have to be spent for each individual content item. Hence, the mere existence of a comprehensive and authoritative data infrastructure in a given sector does not automatically ensure that all rightholders provide the data necessary to maintain data accuracy and completeness. Revisiting the potential discrepancy between the interests of small and big players, continuous data entry and maintenance may be less burdensome for holders of big work libraries in the light of economies of scale. For instance, it is conceivable that holders of big repertoires are able to switch from manual data entry to the use of automated or machine-learning systems, which substantially reduce the cost per unit.

33 Finally, it is to be noted that “costs” can also be understood in a broader sense. Instead of confining the analysis to monetary aspects, it is important to con-

65 European Commission, 19 February 2020, “A European Strategy for Data”, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Document COM(2020) 66 final, 1, available at: <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy>.

66 European Commission, id. 7-8.

67 See <https://completemusicupdate.com/article/prs-confirms-global-repertoire-database-cannot-move-forward-pledges-to-find-alternative-ways/>.

68 See <https://www.appleworld.today/blog/2019/11/18/apple-spotify-to-fund-new-music-royalties-collective>.

69 U.S. Congressional Budget Office Cost Estimate, S. 2823 – Music Modernization Act, as reported by the Senate Committee on the Judiciary on 12 September 2018 (revised version of 17 September 2018), 3, available at <https://www.cbo.gov/system/files/2018-09/s2823.pdf>.

sider broader cultural repercussions, in particular the impact of standardised data formats and comprehensive copyright data systems on *cultural diversity*, recognition and attribution (in the sense of the moral rights enjoying protection under international copyright law and the national copyright systems of EU Member States) and the visibility and availability of the full spectrum of European creative works. In the case of photography, for instance, the commercial value of a work for rightholders, as explained, will often be smaller than the cost of documenting the work. The outlined problem raises concerns about large economies of scale favouring large repertoire owners who can automate the documentation and indexation process. Considering this potential problem, it becomes apparent that the burden of documenting and promoting content in large, supra-national content repositories should not increase data management burdens to such an extent that it becomes unprofitable for smaller entities to comply with data standards and data entry requirements. Otherwise, the measures taken to improve copyright data management may discriminate against holders of small repertoires—and potentially even against smaller national repertoires in the EU—and reduce the cultural diversity which the improved data system is intended to reflect.

II. Benefits Accruing from Improved Copyright Data

- 34 Benefits that can be expected to flow from an improved data management infrastructure are enhanced licensing opportunities, more efficient enforcement of rights, the reduction of royalty losses and the enhancement of access of high-tech industries to copyright data. Conversely, missing or inaccurate copyright metadata can lead to various types of welfare losses:
- a) work is not found and therefore not licensed. That is, the licensing transaction does not take place, depriving both rightholders and consumers of the potential welfare gains (producer surplus and consumer surplus) which a transaction would generate in the counterfactual of accurate metadata;
 - b) work is found or the potential licensee is aware of the work, but information to license is missing. This may result in two outcomes:
 - i) the work is not used/consumed, as under (a);
 - ii) the work is pirated/used without a license. In this case, all welfare effects of the transaction are generated on the demand side, while rightholders do not benefit;
 - c) The work is found and licensed, but no proper remuneration is provided to rightholders as a consequence of the inaccurate metadata, i.e., licensing revenues are collected but do not reach the rightholders due to metadata issues.
- 35 Missed licencing and remuneration opportunities not only entail so-called static welfare losses; there can be dynamic effects as well. Efficient licensing can enable more creators to draw on existing copyrighted works, reducing the costs of follow-on creativity. Secondly, smaller markets for copyrighted works and greater costs of licensing will entail lower incentives to invest in innovative complementary goods and services (e.g. innovative ways of disseminating copyrighted works online or innovative recommender systems). Thirdly, high transaction costs, legal uncertainty, competition from unlawful competitors, market concentration and barriers to entry that result from (the requirement to incur) sunk costs can inhibit innovation. Efficient licensing systems—including metadata—can mitigate these issues. An obvious remedy, therefore, would be to correct and complete the metadata.
- 36 In addition, the aforementioned cultural dimension must be taken into account. Better visibility and availability of European cultural productions on the world market and the (possibly even more important) domestic European market offers important benefits. To the extent that EU creative industries do not have their own comprehensive repertoire databases, they depend on the configuration of content recommendation and licensing systems developed elsewhere. This entails the risk of insufficient influence on the promotion, sales and distribution process.⁷⁰ In theory, the repertoire databases of Apple Music, Spotify, YouTube or Deezer, for instance, may offer all providers of cultural content similar opportunities to reach out to end consumers. In practice, however, the visibility and success of a work will depend, *inter alia*, on the way in which these providers organise work- and creator-related (meta-)data and generate recommendations for end consumers. This implies that European content producers depend heavily on metadata and recommendation systems that have been developed by powerful individual companies. In the field of music, the MLC initiative that follows from US legislation may strengthen this trend. As the MLC database has been established with a focus on the US market

⁷⁰ As to existing legislation seeking to enhance the visibility and prominence of European content, see Article 13(1) of the Audiovisual Media Services Directive 2010/13/EC, as amended by Directive 2018/1808/EU.

and in collaboration with Apple and Spotify, European content is unlikely to occupy the centre stage.

- 37 A further risk arises from the diversity of European content in terms of cultural backgrounds and languages. There are various cultural and media policy tools employed in Europe—mainly introduced in national law, such as various local content regulations (for example, in the form of radio or television “quotas” or programming guidelines set for public broadcasters). These instruments aim at the development of local audiences for local content. For these instruments to be efficient and measurable, usable and timely metadata are necessary. Descriptive metadata, however, are usually connected with natural languages. The costs of documenting in smaller European languages relative to the expected sales value can be significantly higher for language groups with fewer potential buyers. This creates an incentive to replace higher cost-to-market repertoires from smaller language groups with (translations of) lower cost-to-market repertoires from large language groups, such as works for English-speaking audiences, in unregulated markets. It also creates incentives to bypass regulations, like in television or radio broadcasting streams, when neither the regulated programmer nor the public authorities measuring local content guidelines have high quality data available.⁷¹

E. Article 17 CDSMD as a Catalyst

- 38 Considering the described complexity of data improvement initiatives, the various factors impacting data creation and management, and the different dynamics, costs and benefits in individual sectors of the creative industry, Article 17 CDSMD can hardly be expected to solve all dilemmas surrounding copyright data in the EU. Nonetheless, the provision—in particular the mechanism of notifying works under Article 17(4)(b)—seems to offer an unprecedented opportunity for data improvement, in particular with regard to those categories of creative content that feature prominently on OCSSP platforms: music, film, photography and other forms of visual art.⁷²

71 Cf. Daniel Antal, Amelia Fletcher and Peter L. Ormosi, “Music Streaming: Is It A Level Playing Field?”, *Competition Policy International* 2021, 23 February 2021, available at: www.competitionpolicyinternational.com.

72 Cf. the OCSSP definition in Article 2(6) CDSMD. As to the underlying user activity of sharing literary and artistic works, see Martin R.F. Senftleben, “User-Generated Content – Towards a New Use Privilege in EU Copyright Law”, in: Tanya Aplin (ed.), *Research Handbook on IP and Digital Technologies*, Cheltenham: Edward Elgar 2020, 136-162; Jean-Paul Triaille, Séverine Dusollier et al., *Study on the Application of Directive*

I. Tapping Into the Data Flow From Rightholders to OCSSPs

- 39 As already explained above, Article 17(4)(b) requires OCSSPs to make best efforts to ensure the unavailability of works and other protected subject matter for which rightholders have provided OCSSPs with “relevant and necessary” information. This notification mechanism generates a data flow from rightholders to OCSSP platforms, covering any unlicensed content that rightholders want to have removed from the platforms.⁷³ The notification of works gives rightholders the opportunity to ensure the application of measures to block and remove infringing content. “[R]elevant and necessary information” in the sense of Article 17(4)(b) can be expected to go beyond mere work-related data. A copyright owner sending information must inform the OCSSP about their identity, address and contact details, and the nature and (territorial) scope of the rights that are asserted. Article 17(8) CDSMD stipulates that OCSSPs should “provide rightholders, at their request, with adequate information on the functioning of their practices with regard to the cooperation referred to

2001/29/EC on Copyright and Related Rights in the Information Society, Study prepared by De Wolf & Partners in collaboration with the Centre de Recherche Information, Droit et Société (CRIDS), University of Namur, on behalf of the European Commission (DG Markt), Brussels: European Union 2013, 457-510; Steven D. Jamar, “Crafting Copyright Law to Encourage and Protect User-Generated Content in the Internet Social Networking Context”, *Widener Law Journal* 19 (2010), 843; Natali Helberger, Lucie Guibault et al., *Legal Aspects of User Created Content*, Amsterdam: Institute for Information Law 2009; Mary W.S. Wong, “Transformative User-Generated Content in Copyright Law: Infringing Derivative Works or Fair Use?”, *Vanderbilt Journal of Entertainment and Technology Law* 11 (2009), 1075; Edward Lee, “Warming Up to User-Generated Content”, *University of Illinois Law Review* 2008, 1459; Betty Buckley, “SueTube: Web 2.0 and Copyright Infringement”, *Columbia Journal of Law and the Arts* 31 (2008), 235; Tom W. Bell, “The Specter of Copyism v. Blockheaded Authors: How User-Generated Content Affects Copyright Policy”, *Vanderbilt Journal of Entertainment and Technology Law* 10 (2008), 841; Steven Hechter, “User-Generated Content and the Future of Copyright: Part One – Investiture of Ownership”, *Vanderbilt Journal of Entertainment and Technology Law* 10 (2008), 863; Greg Lastowka, “User-Generated Content and Virtual Worlds”, *Vanderbilt Journal of Entertainment and Technology Law* 10 (2008), 893; OECD, 12 April 2007, *Participative Web: User-Created Content*, Doc. DSTI/ICCP/IE(2006)7/Final, available at: <https://www.oecd.org/sti/38393115.pdf>.

- 73 Strictly speaking, data flows following from the practical implementation of Article 17(4)(b) CDSMD do not cover information on (i) licensed content; and (ii) unlicensed content, in respect of which rightholders refrain from actively enforcing their rights under Article 17 CDSMD.

in paragraph 4.” Without contact information, this reporting duty cannot be fulfilled. In the context of the complaint and redress mechanism following from Article 17(9) CDSMD, rightholders “shall duly justify the reasons for their [content blocking] requests.” Rightholders are thus under an obligation to substantiate their claims. Evidently, the information exchange between rightholders and OCSSPs is intended to create not only up-to-date libraries of fingerprints or other reference information to identify works, but also an accurate and constantly updated collection of data concerning rights ownership and contact information. Otherwise, OCSSPs can hardly report on content moderation practices and invite rightholders to substantiate blocking requests in the framework of complaint procedures.

- 40 Considering these proportions of the data flow and the need for up-to-date information on protected works, rights ownership, and the nature and scope of rights, the specific opportunity arising from Article 17(4)(b) CDSMD becomes manifest: if all notifications that are sent to OCSSPs across EU Member States are collected and bundled in a central EU copyright data repository, the accumulation of EU copyright data can lead to an unprecedented data reservoir that outperforms pre-existing data silos of CMOs, rightholders and distribution platforms.⁷⁴ As the described cooperation between rightholders and OCSSPs—enabling content moderation reporting and collaboration in complaint cases—requires that the information on rights and rights ownership be updated continuously, the central EU copyright data repository fed by Article 17(4)(b) notifications can be expected to achieve a relatively high degree of data currentness and accurateness.
- 41 To establish this EU copyright data repository, it is necessary to tap into Article 17(4)(b) notifications. Instead of sending “relevant and necessary information” only to OCSSPs, rightholders would have to make this information available, in parallel, to a central institution administering the EU copyright data repository.⁷⁵ This data aggregation mechanism could

74 In the legislative process leading to the adoption of Article 17 CDSMD, Germany suggested in this vein “public, transparent notification procedures” as a potential concept to “counteract a de facto copyright register in the hands of dominant platforms.” see Council of the European Union, Statement by Germany, (5 April 2019), point 5, 4, <https://data.consilium.europa.eu/doc/document/ST-7986-2019-ADD-1-REV-2/en/pdf>.

75 As to this additional data transmission obligation, the question of enforcement arises. What should be the consequence of not reporting? A ban on content blocking may be problematic if it leads to delays. If rightholders directly engage with OCSSPs, the latter can directly act upon the received information. If they must wait until the information

overcome the traditional resistance of central gatekeepers, such as CMOs, to share valuable information on works and copyright owners. Arguably, the incentive to block infringing content uploads with Article 17(4)(b) notifications is strong enough to make use of the notification system, even if notified information is also included in an overarching EU database. At the same time, the bundling and harmonisation of copyright metadata in an open format EU repository would lead to data access and transparency for all OCSSPs—regardless of their size—and other interested users (including other online platforms). As a result, big OCSSPs with broader access to copyright data because of more comprehensive activities are less likely to become new gatekeepers with competitive advantages because of superior knowledge of works and copyright owners. The larger copyright data flow to big OCSSPs would automatically enrich the EU data repository as well. The information will thus be available to all interested OCSSPs and other potential users.

II. Implementation Templates and Data Interoperability

- 42 A template for legislation that would ensure this redirection of copyright (meta-)data to a central data collection point can be found in Article 3(6) of the 2012 Orphan Works Directive⁷⁶ (regarding information on the use of orphan works) and Article 10(1) CDSMD (regarding information on the use of out-of-commerce works). Article 3(6) of the 2012 Orphan Works Directive provides:

*“Member States shall take the necessary measures to ensure that the information referred to in paragraph 5 [information on diligent searches, orphan work use, orphan work status and contact information of cultural heritage institutions] is recorded in a single publicly accessible online database established and managed by the Office for Harmonization in the Internal Market (‘the Office’) in accordance with Regulation (EU) No 386/2012.”*⁷⁷

is registered at EUIPO, this may be different. For orphan works, the “penalty” is not being able to use the work in accordance with the use privilege prescribed in the Orphan Works Directive.

76 Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on certain permitted uses of orphan works, *OJ* 2012 L 299, 5.

77 Article 10(1) CDSMD includes a similar requirement for information on out-of-commerce works: “Member States shall ensure that information from cultural heritage institutions, collective management organisations or relevant public authorities, for the purposes of the identification of the out-of-commerce works or other subject matter, cov-

- 43 Interestingly, Article 3(6) of the Orphan Works Directive and Article 10(1) CDSMD also mention the institution that could take care of the central EU copyright data repository: the European Union Intellectual Property Office in Alicante (“EUIPO”), known as Office for Harmonization in the Internal Market until 23 March 2016.
- 44 To achieve data interoperability, the legal obligation to send Article 17(4)(b) CDSMD notifications not only to OCSSPs but also to the EUIPO could be accompanied by an additional obligation to provide the data in a specific, standardised format. In this way, Article 17(4)(b) could be employed as a vehicle to tackle not only issues of data accuracy and recentness, but also the problem of data interoperability and data harmonisation. One could also think of imposing an obligation on OCSSPs to accept notifications in the standardised format used by the EUIPO. In this way, the parallel data transmission obligation would have the benefit for rightholders of creating one data submission standard that is generally accepted and allows the universal application of notifications. Rightholders would no longer have to deal with data submission standards that may vary from OCSSP to OCSSP.

III. No Conflict With International Prohibition of Formalities

- 45 The international prohibition of formalities following from Article 5(2) of the Berne Convention for the Protection of Literary and Artistic Works (“BC”) need not constitute an insurmountable obstacle in this respect. According to Article 5(2), “[t]he enjoyment and the exercise” of the rights granted in Article 5(1) BC shall not be subject to any formality. Article 5(1) covers the rights which the laws of Berne Union countries “do now or may hereafter grant to their nationals, as well as the rights specially granted by this Convention.” As Stef van Gompel explains in his in-depth analysis of the scope of the prohibition following from Article 5(2) BC, the ban on formalities:

ered by a licence granted in accordance with Article 8(1), or used under the exception or limitation provided for in Article 8(2), as well as information about the options available to rightholders as referred to in Article 8(4), and, as soon as it is available and where relevant, information on the parties to the licence, the territories covered and the uses, is made permanently, easily and effectively accessible on a public single online portal from at least six months before the works or other subject matter are distributed, communicated to the public or made available to the public in accordance with the licence or under the exception or limitation. The portal shall be established and managed by the European Union Intellectual Property Office in accordance with Regulation (EU) No 386/2012.”

“includes formalities relating to the coming into existence, the maintenance and the enforcement of copyright. The Berne prohibition on formalities does not extend to formalities that regulate the extent of protection or the means of redress afforded to authors to protect their rights. This suggests that formalities are allowed if they establish the manner of exercising copyright, but not if their non-compliance renders the exercise of rights completely impossible.”⁷⁸

- 46 Within this matrix, the notification system following from Article 17(4)(b) CDSMD falls within the category of permissible formalities concerning the “manner of exercising copyright” and the regulation of “the extent of protection.” By stipulating that OCSSPs perform an act of communication to the public, or an act of making available to the public, when they give the public access to protected works that have been uploaded by users, Article 17(1) CDSMD establishes a direct, primary liability of online platforms⁷⁹ in an area that, traditionally, has been regulated from the perspective of secondary liability for infringing content uploads.⁸⁰ Quite clearly, the

78 Van Gompel, *supra* note 61, 212.

79 For a more detailed discussion of the nature of the right recognized in Article 17 CDSMD, see Husovec and Quintais, *supra* note 8, 325-348.

80 Cf. Matthias Leistner, “European Copyright Licensing and Infringement Liability Under Art. 17 DSM-Directive Compared to Secondary Liability of Content Platforms in the U.S. – Can We Make the New European System a Global Opportunity Instead of a Local Challenge?”, *Zeitschrift für Geistiges Eigentum/Intellectual Property Journal* 26 (2020), 123-214; Stefan Kulk, *Internet Intermediaries and Copyright Law – Towards a Future-Proof EU Legal Framework*, Utrecht: University of Utrecht 2018; Martin R.F. Senftleben, “Content Censorship and Council Carelessness – Why the Parliament Must Safeguard the Open, Participative Web 2.0”, *Tijdschrift voor Auteurs-, Media- & Informatierecht* 2018, 139 (139-140); Martin Husovec, *Injunctions Against Intermediaries in the European Union – Accountable But Not Liable?*, Cambridge: Cambridge University Press 2017; Christina Angelopoulos, *European Intermediary Liability in Copyright: A Tort-Based Analysis*, Alphen aan den Rijn: Kluwer Law International 2016; Martin R.F. Senftleben, “Breathing Space for Cloud-Based Business Models – Exploring the Matrix of Copyright Limitations, Safe Harbours and Injunctions”, *Journal of Intellectual Property, Information Technology and E-Commerce Law* 4 (2013), 87 (87-90 and 94-95); Thomas Hoeren and Silviya Yankova, “The Liability of Internet Intermediaries – The German Perspective”, *International Review of Intellectual Property and Competition Law* 43 (2012), 501; Rita Matulionyte and Sylvie Nérisson, “The French Route to an ISP Safe Harbour, Compared to German and US Ways”, *International Review of Intellectual Property and Competition Law* 42 (2011), 55; Miguel Peguera, “The DMCA Safe Harbour and Their European Counterparts: A Comparative Analysis of Some Common Problems”, *Columbia Journal of Law and the Arts* 32 (2009), 481; Christiaan

more detailed specification of this exclusive right, including the option to escape liability with best efforts to obtain licenses and apply content filters (Article 17(4)(a) and (b) CDSMD), regulate “the extent of protection.”⁸¹ The fact that rightholders are obliged to provide “relevant and necessary information” under Article 17(4)(b) shows that the provision establishes a specific “manner of exercising copyright.”⁸² In any event, in situations where no authorisation has been granted to OCSSPs, rightholders can still enforce their rights against individual uploaders.⁸³ Instead of rendering the exercise of rights impossible, Article 17(4)(b) thus offers rightholders an additional possibility to ensure the unavailability of their works on OCSSP platforms.

- 47 On balance, the notification system following from Article 17(4)(b) is a permissible formality because it enhances the extent of protection and regulates the manner of exercising copyright in the specific context of cooperation with OCSSPs. Against this background, it is possible to extend the notification mechanism and add an obligation to send notifications not only to OCSSPs but also to a central EU data collection point that could be established at the EUIPO. The prohibition of formalities in Article 5(2) BC does not preclude the introduction of this data improvement mechanism in the EU.

IV. Extension to Right of Reproduction

- 48 Before painting an overly positive picture and presenting Article 17(4)(b) notifications as the ultimate cure for copyright data issues in the EU, however, it is important to point out that the aggregation of Article 17(4)(b) data is only one piece of a more complex puzzle. As explained, this piece seems important and promising enough to take the described steps towards an overarching EU data repository. Nonetheless, it is important to add several nuances and warn against exaggerated expectations.

Alberdingk Thijm, “Wat is de zorgplicht van Hyves, XS4All en Marktplaats?”, *Ars Aequi* 2008, 573; Matthias Leistner, “Von “Grundig-Reporter(n) zu Paperboy(s)” Entwicklungsperspektiven der Verantwortlichkeit im Urheberrecht”, *Gewerblicher Rechtsschutz und Urheberrecht* 2006, 801.

81 Cf. van Gompel, *supra* note 61, 212.

82 Cf. van Gompel, *supra* note 61, 212.

83 Article 17(2) CDSMD merely exonerates non-commercial uploaders whose activities do not generate significant revenues from liability for copyright infringements in situations where an OCSSP has obtained authorisation, for instance through a licensing agreement.

- 49 First, the regulatory framework of Article 17 CDSMD focuses on the right of communication to the public and acts of making available to the public. This follows clearly from Article 17(1) and (2) CDSMD.⁸⁴ Accordingly, the notification mechanism arising from Article 17(4)(b) CDSMD concerns these exclusive rights. While the right of communication to the public and the right of making available to the public are central to online platforms and various other forms of digital services, new technologies offering promising revenue prospects may require rights clearance in the area of the right of reproduction instead. The use of copyrighted material for AI training purposes (discussed in section B above) can serve as an example. As the text and data mining provisions in Articles 3 and 4 CDSMD show, the right of reproduction⁸⁵ occupies centre stage in this context.

- 50 However, the question arises whether an EU data repository fuelled by data from Article 17(4)(b) notifications is capable of providing useful information for work identification and rights clearance initiatives in new technology areas, such as the AI sector, that require information on reproduction rights. The answer to this question depends on the expression “relevant and necessary information” in Article 17(4)(b). For the purpose of ensuring the unavailability of protected works on OCSSP platforms, it is relevant and necessary to know who is entitled to prohibit the sharing of user-uploaded content because they hold the rights of communication and making available to the public. As the EU data repository enhances the visibility of protected works and increases licensing opportunities, however, it may make sense for copyright holders to provide information on a broader spectrum of exclusive rights and include ownership information covering reproduction rights as well. The mere fact that ownership and repertoire information notified under Article 17(4)(b) CDSMD will make its way into the EU copyright data repository may lead to “enriched” notifications that go beyond the information that is strictly “relevant and necessary” in the OCSSP platform context. As pointed out above, Article 17(4)(b) CDSMD may have the effect of a catalyst that sets in motion a broader process of copyright data aggregation. This broader process may capture additional exclusive rights, such as the right of reproduction.

84 See also Article 3(1) and (2) ISD. For a discussion of the relationship between Article 17(1) and (2) CDSMD on the one hand, and Article 3(1) and (2) ISD on the other, see Husovec and Quintais, *supra* note 8, 325-348.

85 Article 2 ISD.

V. Extension to Data Reflecting Nature and Contents of Works

51 Second, rightholders notify work-related information under Article 17(4)(b) CDSMD for the purpose of detecting unauthorised user uploads on OCSSP platforms. The notification data serves the purpose of identifying works and infringing copies.⁸⁶ Given this focus, Article 17(4)(b) notifications may fail to provide insights into the nature and contents of the work itself (such as information on the genre, theme and subject, language and other metadata of the work). A prospective user looking for a specific type of work, such as an AI developer looking for a specific category of music, text or images, may thus find the information that can be derived from Article 17(4)(b) notifications unsatisfactory. However, this need not be the final word on the matter. Again, it is to be considered that, as source material for an EU data repository, Article 17(4)(b) notifications would lead to enhanced visibility of work repertoires and broaden licensing opportunities for copyright holders. Arguably, these benefits provide a strong incentive for copyright holders to go beyond data for work identification purposes and enrich notifications with additional data reflecting the nature and contents of the work. When the institution administering the EU data repository is included in the stakeholder dialogue following from Article 17(10) CDSMD, the discussion of best practices can address the need for copyright data improvement and support the evolution of appropriate notification standards, including data enrichment besides harmonisation and interoperability issues, to maximise beneficial effects of the bundling of Article 17(4)(b) notifications.

F. Conclusion

52 To enhance the visibility and accessibility of the European repertoire and allow the creative industries to benefit from new licensing opportunities in the field of new technologies, it is important to arrive at a comprehensive database with a focus on European content, including smaller and less-known repertoires reflecting the full cultural diversity across EU Member States. An improved copyright data infrastructure is likely to enhance licensing, enforcement and royalty opportunities for creative industries.

86 As to the functioning of content identification tools and the data required for this process, see the report by Jean-Philippe Mochon and Alexis Goin, in collaboration with the Haute autorité pour la diffusion des oeuvres et la protection des droits sur Internet (Hadopi) and the Centre national du cinéma et de l'image animée (CNC), *Les outils de reconnaissance des contenus et des oeuvres sur les plateformes de partage en ligne II*, Paris: CSPLA 2021.

tries. This added value is a core argument in the cost-benefit analysis that can tip the scales in favour of new efforts to create and harmonise metadata. At the same time, a central EU copyright data repository could provide developers of new technologies, such as AI system developers, broad access to diverse data resources. As a counterweight to initiatives in other regions, such as the MLC in the US, it can be expected to allow European creative industries to innovate and emancipate themselves from other data infrastructures and related content distribution and recommendation systems. It may also prevent a non-European bias in globally dominant AI systems trained on copyright data.

53 The foregoing discussion, however, also reflects the considerable obstacles on the way to more comprehensive and accurate European copyright (meta-) data. In addition to substantial financial resources that will be necessary, a key to new and successful initiatives lies in the creation of appropriate incentives for the creative industries, providers of digital content distribution services and high-tech companies in the field of AI to jointly develop solutions. For a trade-off across these industry sectors, the analysis provides an important starting point. The requirement of providing “relevant and necessary information” for the blocking of infringing content in Article 17(4)(b) CDSMD offers room for establishing an obligation to provide data concerning the protected work, the nature and scope of exclusive rights, and the identity and contact details of the rightholder in a standardised and interoperable format. If all Article 17(4)(b) notifications that are sent to OCSSPs across EU Member States are collected and bundled in a central EU copyright data repository, the accumulation of EU copyright data could lead to an unprecedented data reservoir that outperforms pre-existing data silos of CMOs, rightholders and distribution platforms.

54 All industry branches involved—the creative industries, the providers of online platforms and the high-tech industry—could benefit from an improved and harmonised EU data infrastructure. Content distribution platforms and AI companies may have a particular interest in rules that make copyright enforceability and remuneration obligations conditional on the provision of metadata in a specific, interoperable format. To achieve this goal, it could be said that information on protected literary and artistic creations is only “relevant” in the sense of Article 17(4)(b) when it is provided in a form that allows content moderation systems to read it.⁸⁷ At the core of

87 As to the use of the requirement of “relevant and necessary information” as a tool to promote specific notification standards, see Martin R.F. Senftleben and Christina Angelopoulos, *The Odyssey of the Prohibition on General Monitoring Obligations on the Way to the Digital Services Act: Between Article*

these considerations lies the more general principle that rights must be clearly drawn to be enforceable. In this vein, it can be posited that rightholders must provide interoperable, accessible information to benefit from enhanced enforcement opportunities. In addition, it could be said that Article 17(4)(b) notifications should be detailed and rich enough to allow an EU data repository to enhance the visibility of the European repertoire in a meaningful way and broaden licensing opportunities for copyright holders. This objective may require Article 17(4)(b) notifications that cover a broad spectrum of exclusive rights—not only the rights of communication and making available to the public but also reproduction rights—and metadata reflecting the nature and contents of notified works.

- 55 In sum, new approaches in the area of copyright data improvement can evolve from a trade-off addressing interoperability and transparency interests. On the one hand, the interest of online content distributors and AI trainers in standardised and interoperable data formats could be recognised. On the other hand, transparency and accountability in respect of algorithmic content selection, moderation and recommendation systems should be ensured to pave the way for the eradication of systems that may disadvantage small and lesser-known enterprises and repertoires or creators with specific racial, ethnic or other minority backgrounds. To make this incentive scheme for collaboration attractive to a broad spectrum of copyright holders, further research is necessary to develop appropriate solutions not only for big companies but also for independent labels and other SMEs in the creative industries. In addition, it remains an open question whether the prospect of enhanced collaboration in the area of interoperability and transparency would also be sufficient to convince central gatekeepers, in particular CMOs, to contribute to fully standardised and interoperable copyright metadata. As pointed out above, the fear of losing their exclusive position in controlling relationships with their members may trigger resistance against injecting data into a fully standardised copyright data system. A central data accumulation system built on Article 17(4)(b) CDSMD offers an important data improvement opportunity against this background.

15 of the E-Commerce Directive and Article 17 of the Directive on Copyright in the Digital Single Market, Amsterdam: Institute for Information Law/Cambridge: Centre for Intellectual Property and Information Law 2020, 31, available at: <https://ssrn.com/abstract=3717022>.