

Faculté de droit et de criminologie

The legal Evolution of the Text and Data Mining Copyright Exceptions in the European Union

From fragmentation to uniformity?

Author: Stéphan André

Promoter: Alain Strowel

Academic Year 2021-2022

Finality European Law

Plagiarism and serious methodological error

Plagiarism, even of non-copyrighted texts, leads to the application of section 7 of articles 87 to 90 of the General Regulations for Studies and Examinations.

Plagiarism consists of using ideas, a text or a work, even partially, without mentioning the name of the author and the source at the exact time and place of each use*.

In addition, the literal reproduction of passages from a work without placing them in quotation marks, even if the author and source of the work are mentioned, constitutes a serious methodological error that can lead to failure.

* On this subject, see notably <http://www.uclouvain.be/plagiat>.

Table of contents

Introduction	1
Part I: The technicity of the text and data mining and its link with copyright through EU Member State law and Directive 2019/790	3
Chapter 1: Technical Analysis and TDM Use Cases	3
Section 1: Definition	3
Section 2: Origin and Functioning	5
Section 3: Use Cases	8
§1. Lessons from the Covid-19 pandemic.....	8
§2. Correlations between stock market news sentiments and real stock market evolution	10
§3. Mapping crime through newspaper reports.....	11
§4. Conclusion.....	12
Section 4: Clashing with the EU copyright framework	12
§1. The Input	13
§2. The Output.....	14
Chapter 2: TDM before and after Directive 2019/790, or when the European Union gets overtaken by its (ex-)Member States	15
Section 1: Copyright framework exceptions used by national legislations	15
Section 2: Pre-existing TDM frameworks in Member States	17
§1. France	18
§2. Germany	20
§3. The United Kingdom.....	23
Section 3: The new TDM exceptions from Directive 2019/790	25
§1. The genesis of a directive to reform copyright in the digital single market.....	26
§2. Article 3: TDM for the purpose of scientific research	29
§3. Article 4: TDM for everyone	30
Section 4: Conclusive comparison table	32

Part II: What works and what doesn't – A critical analysis of the Directive's TDM exceptions and Member States' transposition progress	33
Chapter 1: Two exceptions in troubled waters	33
Section 1: Global scholarship position overview	33
Section 2: Scope of application	35
§1. Problems of research and related organisations	36
§2. Article 4's opt-out	38
§3. Lawful Access	38
§4. Forgotten other exclusive rights?	39
§5. Inconsistencies	41
Section 3: Limitations	41
§1. Technical Protection Measures (TPMs)	42
§2. Contractual overwriting of copyright exceptions	43
Section 4: Conclusive thoughts	44
Chapter 2: Diverse national transpositions	46
Section 1: Overview	46
Section 2: Post Directive 2019/790 national TDM frameworks	47
§1. France	47
A. Modifications of the general copyright regime	48
B. Modifications of computer program, neighbouring and database rights	51
C. Comparison with the 2016 regime	51
D. Release of the application decree for clarification	51
§2. Germany	52
A. §44b - The general exception	53
B. §60d - The exception for scientific research	55
C. Conclusion	58
§3. Other Member States	59
A. The Netherlands	59

B. Spain	60
C. Ireland	61
§4. Conclusive comparison table	63
Conclusion.....	64
Bibliography.....	65
Legislation.....	65
Official documents	68
Case law	70
Literature	70
Annexes.....	77
Annex I: Interview conducted with Felix Reda on the 9 th of March 2022.....	77
Annex II: Interview conducted with Axel Voss on the 6 th of April 2022	85
Annex III: Member States-EU TDM provisions comparison table (2016-2019)	91
Annex IV: Member States-EU TDM provisions comparison table (2019-2022): CDSM art. 3	92
Annex V: Member States-EU TDM provisions comparison table (2019-2022): CDSM art. 4	94

Acknowledgements

The writing of this thesis has not always been a straight line. Between finding and fixing the correct subject, internships and exams, it has not been completed when it was originally supposed to be, namely in June 2022. It's instead around two months later that the last letter will find its way onto this (digital) paper. Luckily, I was able to multiply my dedication to this topic because of the infinite support I received from my girlfriend. We motivated each other every single day of July, where we had the time to continue working on our own theses. I would like to thank her very much for that.

I would also like to express my gratitude towards a select few people who either provided their own input to this thesis, gave me explanations or guided down the correct paths.

The first two people would be FELIX REDA and AXEL VOSS, with who I was allowed to hold an interview in March and April respectively. Their views have substantially contributed to shaping my understanding and critical view of Directive 2019/790, its procedure and outcome. They were both very much accessible and friendly, and as such I would like to thank them a lot for the great discussions we had together.

Next in line would be PIETER STEVENS, from the Belgian Federal Public Service of the Economy, and a jurist from the German Federal Ministry of Justice, whose name I sadly never got to hear. They both answered my questions about their respective domestic text and data mining framework. Without the German jurist, I would have had a very hard time finding literature about TDM in Germany, since although German is my native language, I never had to consult such literature in any available database before. As such, I would like to thank them both for their guidance and expertise.

I also shouldn't forget to explicitly mention ALAIN STROWEL, ROSSANA DUCATO and LUC DESAUNETTES, who I owe the interest in the subject and even the possibility to write my thesis on it. Mr. STROWEL guided me both actively and passively in the making of this thesis, through the various calls and email exchanges we had, but also through the works he himself released on the subject together with Ms. DUCATO. He also was the one who referred me to Mr. DESAUNETTES, who allowed me to grasp the importance of a research work on the subject I finally chose.

Lastly, I must thank SAMUEL-JAY PULMAN for formally correcting this thesis.

Thank you all very much.

Introduction

In 2020, 2,5 million terabytes of data were generated every day¹. Included in this incredibly high number are the millions of research papers published every year by the scientific community². Accordingly, for individuals or for groups of scientists, consulting every publication on a topic they would like to research is getting more and more difficult, if not « [a]bsolutely impossible. »^{3 4}.

But what if the same task would not have to be accomplished by a human, but by a computer? It is, of course, much easier to feed all the information to a computational device, than to read through hundreds of thousands of abstracts manually to find the proper research publication to have a deeper look at. This is the exact premise which text and data mining (« TDM ») is built on: finding new information, trends, differences and correlations in a very large number of pre-existing works.

Of course, who says « works », says « copyright », and although such a perspective seems amazing for researchers and scientists, it does not appear very attractive to rightholders. They attempted to cling themselves to everything contract law and copyright had to offer, to secure their economic returns in light of a potentially invasive practice. This resulted in voices arising, becoming louder, and calling for legislative intervention to regulate the long legally uncertain methods. Some Member States, including France and Germany, ceded to the demands of the research sector, but many others did not. They patiently waited for what was going to come at the European level, with Directive 2019/790 and the two new copyright exceptions for text and data mining. This time around, every single Member State is enacting the new exceptions into their domestic legal frameworks. Some of them more successful than others...

In the course of this thesis, the first part will technically analyse what text and data mining is, notably through practical use cases to illustrate the theory, before covering its interplay with copyright and the efforts of certain Member States to regulate the phenomenon before the EU did. Comparisons between the different regimes and then the directive will often be made. The second part will subsequently be dedicated to the critically received TDM additions from Directive 2019/790 in the first chapter. It will be completed by a detailed review of their

¹ J. BULAO, « How Much Data Is Created Every Day in 2022? », 3rd June 2022, available at www.techjury.net, accessed on 7th August 2022.

² S. BOON, « 21st Century Science Overload », 7th January 2017, available at www.blog.cdnsiencepub.com, accessed on 7th August 2022.

³ E. BOTKIN-KOWACKI, « Millions of research papers are published in a year. How do scientists keep up? », available at www.phys.org, 27th April 2022, accessed on 7th August 2022.

⁴ *Ibid.*

transposition in the French, German and in other selected legislative frameworks. Here again, the research will lead up to a comparison of these results.

Between fragmentation and unity, how did the legislative text and data mining framework evolve in the European Union? And will the long, troubled transposition process lead to the desired results? We will attempt to find an answer to these questions.

Part I: The technicity of the text and data mining and its link with copyright through EU Member State law and Directive 2019/790⁵

This part is dedicated to a profound analysis of the more technical part of text and data mining (hereafter TDM) and its link with copyright through an overview of the legal prescriptions in different EU Member States, whose legal framework is subject to change because of the relevant articles contained in Directive 2019/790 (hereafter DSM Directive).

The first chapter will cover the definition and the technicity of TDM, including its functioning in the digital world and a handful of use cases in a variety of domains. The second chapter will then focus on French, German and British TDM copyright law, before the adoption of the two new European TDM exceptions. These latter exceptions from Directive 2019/790 will be exposed in the same chapter afterwards.

Chapter 1: Technical Analysis and TDM Use Cases

In this chapter, we will introduce the concept of text and data mining through an overview of the various existing definitions and a more focused examination of its two building blocks: text mining and data mining. The second section will then deal with the origin of text and data mining in the digital world before explaining the functioning of this technology, meaning the different steps a TDM software undergoes to complete its task. The same section will also cover some of the different methodologies that exist. Lastly, the third section will expose three specific use cases of text and data mining, which could be incorporated into real world tools.

Section 1: Definition

In light of the considerable mass of texts published each day by the international scientific community, it continuously becomes more and more difficult, if not even impossible, for a single researcher or a group of scientists to keep track of the works necessary for their research. The technical process of text and data mining promises a solution to this problem, by proposing a complete and ready-to-use analysis of a large number of textual documents for example. The goal is to deliver to the reader an insight on which works he should focus on to find the relevant information he's looking for. TDM searches for correlations, differences, and new information

⁵ 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, *O.J.E.U.*, L 130, 17th May 2019.

that only a computer would be able to discover. Such a process requires that all documents be available in digital form and readable by a machine⁶.

The definition of text and data mining varies. Certain researchers, such as BERGMAN, HUNTER, and RZHETSKY define it as « [...] an interdisciplinary field combining techniques from linguistics, computer science and statistics to build tools that can efficiently retrieve and extract information from digital text. »⁷. Other like ANAWIS, describe it as « the analysis of semi-structured or unstructured text data. The goal is to turn text information into numbers so that data mining algorithms can be applied. »⁸. After having reviewed many different definitions stemming from Japanese, British and Irish legislation as well as from nine sources of legal and scientific literature, the authors of the ‘Study on the legal framework of text and data mining in the EU’⁹ defined it as « the automated processing of digital materials, which may include texts, data, sounds, images or other elements, or a combination of these, in order to uncover new knowledge or insights. »¹⁰.

However, it is important to distinguish between text mining and data mining. The first concerns the analysis of unstructured data, meaning data without a pre-defined format. These types of elements include a wide variety of textual information such as articles, books, social media posts but also, despite its name, images, and audio-visual sources¹¹. From the computer’s point of view such information is, by its nature, more complex to analyse¹² and requires the usage of machine learning, combined with statistics, and linguistic techniques to transform such unstructured data into structured data. Artificial intelligence is applied through natural language processing (NLP), allowing a computer to process human language more efficiently¹³.

On the other hand, data mining concentrates only on structured data, such as numbers, phone numbers, names¹⁴ or other data that can usually be found inside larger databases and searched

⁶ See for example D. McDONALD and U. KELLY, « The Value and Benefits of Text Mining », 14th March 2012, available at www.jisc.ac.uk, accessed on 2nd February 2022, p. 13; or S. DYAS-CORREIA and M. ALEXOPOULOS, « Text and Data Mining: Searching for Buried Treasures », *Serials Review*, 2014, vol. 40, n° 3, p. 211.

⁷ C. BERGMAN, L. HUNTER and A. RZHETSKY, « Announcing the PLOS Text Mining Collection », available at www.everyone.plos.org, 17th April 2013, accessed on 1st February 2022.

⁸ M. ANAWIS, « Text Mining: The Next Data Frontier », available at www.rdworldonline.com, 6th January 2014, accessed on 1st February 2022.

⁹ A comprehensive study funded by the European Commission in 2014, in the context of an adaptation of the UK copyright rules, to gain insights on the current legal status of text and data mining in the European Union. See J-P. TRIAILLE, J. DE MEEÛS D’ARGENTEUIL and A. DE FRANQUEN, *Study on the legal framework of text and data mining*, EU Commission Publications Office, Brussels, 2014, pp. 6-7 & 10-17.

¹⁰ *Ibid.*, p. 17.

¹¹ C. TAYLOR, « Structured vs. Unstructured Data », available at www.datamotion.com, 21st May 2021, accessed on 11th July 2022.

¹² X, « Text mining vs. data mining », available at www.expert.ai, 8th November 2016, accessed on 1st February 2022.

¹³ X, « Text Mining », available at www.ibm.com, 16th November 2020, accessed on 1st February 2022.

¹⁴ *Ibid.*

by humans or algorithms¹⁵. Seeing as it must not be transformed, this type of data is much easier to handle for a machine¹⁶. Here again, artificial intelligence and machine learning are applied, though rather with the goal of finding trends, classifying results, or even predicting numeral values based on a set of given numbers¹⁷. As such, it is here where impressive TDM results are generated.

Text mining can thus be considered as a subcategory of data mining, considering that the first allows transforming unstructured data into structured data, which will then subsequently be more systematically dealt with by the second¹⁸.

Section 2: Origin and Functioning

Text and data mining finds its origin in a collection of multiple individually used techniques in the 1950s. Text classification and clustering technologies, based on the content and the topics of texts, emerged in the field of information recognition. They were followed in 1958 by the automatic summarisation technology, made possible for the first time on a personal computer¹⁹. The actual idea of summarization and classification is much older though. An example can be traced back as early as the 17th century to the work of THOMAS HYDE at the University of Oxford²⁰.

It is only in the 1980s and the 1990s with the rapid development of the internet, an ever-increasing rise in computational power, and specific research projects dedicated to text and data mining (notably by the American DARPA²¹), that information extraction became an important field of research²². Today, the demand for such software has exploded. Be it for the analysis of public opinion through social media, the prediction of the development of the financial markets through the mining of reports, opinions, and press articles, or even the measuring of consumer

¹⁵ C. TAYLOR, *op. cit.*

¹⁶ X, « Structured vs. Unstructured Data: A Complete Guide », available at www.talend.com, *s.d.*, accessed on 11th July 2022.

¹⁷ X, « What's the difference between data mining and text mining? », available at www.blogs.opentext.com, 13th March 2019, accessed on 1st February 2022.

¹⁸ X, « Text & Data Mining: What is TDM? », available at www.libguides.cam.ac.uk, *s.d.*, accessed on 1st February 2022.

¹⁹ C. ZONG, R. XIA and J. ZHANG, *Text Data Mining*, Pekin, Tsinghua University Press, 2021, p. 2.

²⁰ G. MINER *et al.*, *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*, Oxford, Elsevier, 2012, p. 6.

²¹ Short for 'Defense Advanced Research Projects Agency'. The 'Message Understand Conferences' (MUC) were organised to research the « automated analysis of military messages containing textual information. ». See R. GRISHMAN and B. SUNDHEIM, « Message Understanding Conference – 6: A Brief History », *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, 1996, vol. 1, p. 466.

²² G. MINER *et al.*, *op. cit.*, p. 10-11.

reactions after launching a certain product on the market, a lot of private and public actors nowadays are making use of such tools²³.

A very important preliminary step before the actual text and data mining process can start, is the compilation of a database of sources which the program will have access to. This could for example, be a collection of scientific articles gathered by an automatic tool or an already pre-compiled database put freely at the disposal of anyone to experiment on it with TDM tools²⁴.

Once such a database has either been established or acquired and its content has been optimised ('pre-processed'²⁵) to be machine-readable, a TDM software can begin its analysis process. The latter can be divided into four steps: information search, natural language processing, information extraction and data mining²⁶. If we go along the example of a database of collected scientific research articles, the computer first searches inside the documents submitted to it, using, for instance, a keyword entered by a user and selects all texts which could be relevant for later deconstruction. Secondly, using sophisticated language recognition software, the TDM program analyses all words contained in a document in order to transform them into structured data²⁷. For this, natural language processing (hereafter NLP), an artificial intelligence technology, has proven very efficient to classify words grammatically, reveal their true meaning in the context of a document, or even to grammatically represent the structure of an entire phrase, not just of a single word²⁸.

Thirdly, all useful information is extracted. Here, the document is first tokenised, meaning broken down in its constitutive elements, before synonyms are identified and the text is transformed so to accurately represent the words used in it. Today, this extraction is often based on statistical models, as opposed to again using NLP²⁹.

The computer now has all necessary data at hand to move on to the fourth and last step: the data mining itself. Using a variety of algorithms, for predictive data mining even machine learning³⁰, the TDM software begins searching for new information and trends throughout the entirety of

²³ C. ZONG, R. XIA and J. ZHANG, *op. cit.*, p. 2.

²⁴ D. TAYLOR, « Data Mining Tutorial: What is Data Mining? Techniques, Process », available at www.guru99.com, 4th June 2022, accessed on 11th July 2022.

²⁵ « Cleaning and parsing are required because text is created and stored in such a way that, while understandable to humans, it may not be in a form amenable to having the encoded meaning easily extracted by a computer », see E. ROSATI, *The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Technical Aspects*, EUP Policy Department for Citizens' Rights and Constitutional Affairs, Brussels, 2018, p. 6.

²⁶ N. ZANINI and V. DHAWAN, « Text Mining: An introduction to theory and some applications », *Research Matters*, 2015, vol. 19, p. 38.

²⁷ D. McDONALD and U. KELLY, *op. cit.*, p. 13.

²⁸ N. ZANINI and V. DHAWAN, *op. cit.*, p. 39.

²⁹ E. ROSATI, *op. cit.*, p. 7.

³⁰ *Ibid.*

the documents and their related structured data sorted to this effect, and then finally reveals the results to the user³¹.

These four fundamental steps can always be found in the variety of different TDM methods that exist. The following exemplary list reveals that each follows a certain goal:

The text classification method allows, as the name says, to classify a certain text based on its title and its content, into multiple predefined categories³². An example of such a method can be found on press websites, which classify news articles into their corresponding category (like politics, local news, sports, etc.). Text clustering is also a classification method, but the categories into which different texts can be classified are not predefined by the user. Here, the different elements are automatically grouped together based on their similarities and differences, according to certain standards and evaluation criteria³³.

The topic model is an algorithmic probability method used to classify a large number of documents according to their general and/or specific themes. It allows a reader not only to collect many texts concerning a certain topic, but also to analyse the links between those texts, or the evolution of a certain topic throughout time. This method is particularly efficient to organise and summarise digital archives, whose sheer size would make the same task simply impossible for a human being³⁴.

Another method is the sentiment analysis, which is also called opinion mining. This method focuses on how the author of a text expressed himself and verifies whenever the latter had a positive or a negative point of view. Such a method is very useful when researching opinions about a product launched on the market, or to analyse the reaction of a given public to a certain event³⁵.

The last method worth mentioning is automatic text summarisation. After having been practiced on one of the first ever commercial computers³⁶ in 1958³⁷, it evolved drastically and now allows condensing one single or multiple documents into a certain number of important phrases, taken directly from the content of these documents in question. It is clearly in line with the general

³¹ D. McDONALD and U. KELLY, *op. cit.*, p. 14.

³² G. MINER *et al.*, *op. cit.*, p. 32.

³³ C. ZONG, R. XIA and J. ZHANG, *op. cit.*, p. 3.

³⁴ D. BLEI, « Surveying a suite of algorithms that offer a solution to managing large document archives », *IEEE Signal Process. Mag.*, April 2012, vol. 55, n° 4, pp. 77 and 78.

³⁵ C. ZONG, R. XIA and J. ZHANG, *op. cit.*, p. 4.

³⁶ F. DA CRUZ., « The IBM 704 », available at www.columbia.edu, February 2001, accessed on 11th July 2022.

³⁷ H. P. LUHN, « The Automatic Creation of Literature Abstracts », *IBM Journal of Research and Development*, April 1958, vol. 2, n° 2, p. 159.

objective of TDM, which is to considerably limit the mass of information needed to be treated by the human brain³⁸.

These different methods stem from their respective ‘practice areas’, which themselves come from the different interrelated fields of research that text mining revolves around. There are seven of these areas: document classification, document clustering, information extraction, natural language processing, concept extraction, web mining and information retrieval. They originate out of the fields of data mining, AI and machine learning, databases and computational linguistics. This shows how large the whole text mining concept is, and how difficult it is to characterise³⁹.

Section 3: Use Cases

This section covers three different use cases for text and data mining. We will first examine a project from an American researcher who made use of TDM to explore the masses of research articles published about the Covid-19 pandemic with the goal of facilitating the informational search task for his scientific colleagues. The second use case is about analysing the sentiment attached to financial news articles and their correlation with real financial market evolution. The third and last use case we’ll cover, explores how news reports about criminal activities like thefts can, through the usage of TDM, be automatically retrieved and classified before being visually represented on a map.

§1. Lessons from the Covid-19 pandemic

In March 2021, the American researcher BILLIE ANDERSON published a study relating to the analysis of scientific research articles about Covid-19. The objective of this study was to allow researchers to find the information they need, among the thousands of articles published since the start of the pandemic, and whose number appears to be exponentially growing every month⁴⁰.

The author used a small part of the Covid-19 Open Research Dataset. This collaboration effort from the Allen Institute for AI, the White House and various other leading research groups, was first published on the 12th of March 2020. Today, the resource contains over 500.000 scholarly

³⁸ M. GAMBHIR et V. GUPTA, « Recent automatic text summarisation techniques: a survey », *Artificial Intelligence Review*, 2017, vol. 47, p. 2.

³⁹ G. MINER *et al.*, *op. cit.*, p. 30-31.

⁴⁰ B. ANDERSON, « Using text mining to glean insights from COVID-19 literature », *Journal of Information Science*, 2021, n° 9, p. 1.

articles, including over 200.000 with full text⁴¹. In particular, the author used 83.264 abstracts, with an average word count of 75. All documents included in the dataset had already been converted into a machine-readable format⁴².

Before feeding the text into a clustering algorithm, ANDERSON had to pre-process the data by specifying certain terms which would be excluded from the later analysis (so-called stop words), in addition to the exclusion of common speech words. He also had to define synonyms, so different terms could be processed equivalently, before moving on the second step of transposing the terms from the dataset into structured data objects. After applying different structuring algorithms, the author was able to use the clustering method to classify the different abstracts into data-driven groups⁴³.

The result is a table containing 25 clusters, labelled by the author, classified in order of frequency, for example (1) Risk factors (9589) [...], (3) Problems in China (6022) [...], (6) Testing (4462), or (15) Impacts of COVID-19 related to cancer (2476) [...], the first number representing the cluster number, the second number representing the number of articles inside the cluster. The descriptive terms for the first cluster contain for example outcome, risk, increase, clinical, severe, or treatment. Importantly, some of these terms utilised to describe one cluster, repeat themselves among other clusters⁴⁴.

While the author concludes that the text clustering technique offers valuable insight into emerging research trends and offers researchers a possibility to more rapidly find articles corresponding to their needs⁴⁵, he appears to ignore that certain articles could fall within multiple clusters. Indeed, the number of articles contained in each cluster adds up to the total number of 83.264 article abstracts used, meaning that every abstract can only be found inside one single cluster, the one it primarily relates to. This therefore means that, when searching relevant information in one of the categories established by this model, researchers could fail to retrieve the information they are looking for, simply because ANDERSON's model considered them as secondary. It thus appears crucial to mention that informational search must be based directly on the descriptive terms and not on the cluster groups themselves.

⁴¹ X, « COVID-19 Open Research Dataset Challenge (CORD-19) », available at www.kaggle.com, 12th March 2020, accessed on 9th February 2022.

⁴² B. ANDERSON, *op. cit.*, p. 3.

⁴³ *Ibid.*, p. 3-5.

⁴⁴ *Ibid.*, p. 7.

⁴⁵ *Ibid.*, p. 6.

§2. Correlations between stock market news sentiments and real stock market evolution

The second use case focuses on sentiment analysis. The two researchers ANURAG NAGAR and MICHAEL HAHSLER used text and data mining on published news articles to find and determine the emotion and sentiment attached to them and compare these with the evolution of the stock prices of certain companies, like Apple. As they point out, the stock price of a company can be affected by positive or negative news articles, announcements from the company itself, or even public opinion delivered through Twitter. Being able to identify the sentiment of such information could prove useful in predicting the evolution of financial markets⁴⁶.

The authors make use of a specific programming language and environment called 'R', specifically developed for statistical research⁴⁷, to create a programming engine which gathers and processes the latest financial information from various sources in conjunction with Google Finance News Feeds. The research pair took the example of scanning articles for the stock symbol of Apple, AAPL, to find relevant articles. Inside these articles, multiple different natural language processing tools have been used to filter out important sentences from the body text, thereby ignoring the rest of an article, dealing with, for example, a different stock. A similar approach could have been taken with headlines⁴⁸.

The next step was to use a combination of pre-existing databases of positive and negative sentiment terms to identify the sentiment of a phrase, paragraph or document using such keywords. Then, a simple scoring system was established: a sentence can be considered as positive if it contains more positive than negative words and vice versa. Should the amount be equal, the sentiment is neutral⁴⁹.

The overall result is a graphical representation of the accumulation of positive points for a selected stock symbol in the course of a day, which can be compared with the real market development of a stock over time. Days are represented on the X axis and the number of positive sentiments on the Y axis. Although the correlation isn't always correct, some similarities can be seen. In this regard, the authors note that there is a time discrepancy of around two days between their tool and the actual stock development⁵⁰. Indeed, the research tool lags behind and

⁴⁶ A. NAGAR and M. HAHSLER, « Using Text and Data Mining Techniques to extract Stock Market sentiment from Live News Streams », *2012 International Conference on Computer Technology and Science*, 2012, vol. 47, p. 91-92.

⁴⁷ X, « What is R? », available at www.r-project.org, *s.d.*, accessed on 14th February 2022.

⁴⁸ A. NAGAR and M. HAHSLER, *op. cit.*, p. 92-93.

⁴⁹ *Ibid.*, p. 93-94.

⁵⁰ *Ibid.*, p. 94.

therefore does not seem like a viable tool to provide financial assistance. However, if its accuracy can be improved, it could be a powerful tool to analyse the opinion history of the market in the past and thus potential serve as a basis for a predictive analysis of the market in the future.

§3. Mapping crime through newspaper reports

Italian researchers LAURA PO and FEDERICA ROLLO released a TDM experiment in 2018 about the analysis of newspaper articles relating to criminal offences. According to both researchers, the lack of local or national open data initiatives makes it hard to find and compare recent crime statistics. Their goal is thus to create a tool which extracts this information from newspapers and applies it to a map, so to not to rely on the Italian national statistics agency ISTAT anymore. This ‘theft map’ will, for this research experiment, cover the province of Modena only, and allow local citizens and law enforcement agents to identify high-risk areas, compare crime statistics, but also to generally feel safer. Governments, investors, and tourists could also make use of such a tool⁵¹.

The team used a web crawler on the ModernaToday website to collect relevant newspaper reports. From these articles, the main information needed for generating a map was extracted: the URL, title, description, address, publication date and the stolen object⁵². In a more recent publication from 2020, the precise text and data mining process for a very similar tool is detailed: articles must first be gathered through the web crawler, then collected to build a minable corpus. It’s here where articles are filtered for relevant information (thefts, in our case). Once this database has been established, natural language processing is used during the ‘pre-processing’ stage to identify relevant words or phrases in the texts, remove unnecessary punctuation marks and delete stop words. Then, the frequency of words is analysed, and news are labelled according to their characteristics. Finally, all « entities », meaning the relevant pieces of information we’re searching for (addresses, names, dates, etc.), are identified, extracted, and classified⁵³.

The TDM results are visually shown on a heatmap. Crime coordinates are located using the Google Maps Geocoding API on the basis of the addresses specified in the news articles. The

⁵¹ L. PO and F. ROLLO, « Building an Urban Theft Map by Analyzing Newspaper Crime Reports », *2018 13th International Workshop on Semantic and Social Media Adaptation and Personalization*, 1st September 2018, p. 13-14.

⁵² *Ibid.*, p. 15.

⁵³ M. SALDAÑA *et al.*, « Mapping of the Perception of Theft Crimes from Analysis of Newspaper Articles Online », *2020 15th Iberian Conference on Information Systems and Technologies*, 1st June 2020, p. 3.

more crimes there are in a certain area, the darker the colour becomes: red means that there were a lot of crimes, while green means that there weren't many. Yellow is in-between⁵⁴.

While the results look promising, the accuracy of such a technique is debatable. As the authors recognise themselves, 23% of the mined articles couldn't be located correctly by the API as the address was either absent, unreadable, or incomplete⁵⁵. Such a percentage is high and questions the validity of the achieved results. In addition, no de-duplication has been applied⁵⁶, meaning that certain reports could have been added multiple times, again distorting the results further. While this approach thus seems interesting at first, the limited number of news articles and their unreliability in relation to the address where the theft crimes have taken place, makes this tool not accurate enough to be properly trusted by the envisioned addressees.

§4. Conclusion

The aforementioned text and data mining tools all appear interesting, yet their results might not always be accurate enough to be widely used by the general public. They do however very well fit their role of exemplary visualisations of what could be achieved with TDM: creating new information by feeding a great deal of already acknowledged information to a computer.

Section 4: Clashing with the EU copyright framework

Now that it's clear how text and data mining works both in theory and in practice, another major element must not go unaddressed: how TDM works in the copyright environment, and whenever its use possibly infringes parts of the EU copyright system. For this, one is obliged to break down the different copyright-sensitive activities, which are implied in a text and data mining process.

This section will not yet take into account the evolutions stemming from Directive 2019/790, but will exclusively focus on the pre-existing EU copyright directives and the protection they attribute to materials used in a text and data mining process. The following paragraphs therefore relate to the different stages. The first section of the next chapter will, in turn, cover the copyright exceptions which have been used by Member States to fit in their national TDM exceptions.

⁵⁴ L. PO and F. ROLLO, *op. cit.*, p. 15-16.

⁵⁵ *Ibid.*, p. 16.

⁵⁶ *Ibid.*, p. 17.

§1. The Input

As we know, all the information which shall be mined must first be identified and collected, either individually⁵⁷ or by using databases prefabricated to this effect⁵⁸. This type of activity involves most of the time⁵⁹, in some form or another, copying many relevant elements and, where necessary, pre-processing them to enable the machine to read them. In some cases, it might even be required to upload these materials⁶⁰. Such acts can, rightfully or not, undoubtedly be considered as a reproduction of the original material⁶¹.

What thus matters most here is what is being used: whenever or not the collected elements are protected by copyright. Logically, when there is no copyright, it cannot be infringed.

In this sense, we remember the famous words of the European Court of Justice and the criterion of originality, reserving the exclusive right of reproduction⁶² only to parts of a work « [...] contain[ing] elements which are the expression of the intellectual creation of the author of the work. »⁶³. This stance on the definition of a work and the criterion of originality have been held by the Court for years⁶⁴. Consequently, any reproduction of a work, without the authorisation of the rightholder, is not allowed.

The same is true for databases: under the copyright regime, there are, similar to works, also protected if « [...] by reason of the selection or arrangement of their contents, [they] constitute the author's own intellectual creation [...] »⁶⁵. This criterion is, again, fulfilled through

⁵⁷ The second and the third use cases from section 3 individually collect news articles to build their own corpus.

⁵⁸ The first use case from section 3 uses such a database.

⁵⁹ There are instances where certain sources can be directly accessed, therefore precluding the need of copying them to a separate environment for the purpose of any relevant TDM activity. However, such access is only suitable for more limited and simple projects. See, J-P. TRIAILLE, J. DE MEEÛS D'ARGENTEUIL and A. DE FRANQUEN, *op. cit.*, p. 47.

⁶⁰ C. GEIGER, G. FROSIO and O. BULAYENKO, *The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects*, Policy Department for Citizens' Rights and Constitutional Affairs, Brussels, 2018, p. 5.

⁶¹ J-P. TRIAILLE, J. DE MEEÛS D'ARGENTEUIL and A. DE FRANQUEN, *op. cit.*, p. 46; and A. STROWEL, « Reconstructing the Reproduction and the Communication to the Public Rights: How to Align Copyright with Its Fundamentals », in B. HUGENHOLTZ (dir.), *Copyright Reconstructed, Rethinking Copyright's Economic Rights in a Time of Highly Dynamic Technological and Economic Change*, Alphen aan den Rijn, Wolters Kluwer, 2018, p. 204-205.

⁶² 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, *O.J.E.U.*, L 167, 22nd June 2001, art. 2.

⁶³ C.J.E.C., judgement *Infopaq International A/S v Danske Dagblades Forening*, 16th June 2009, C-5/08, EU:C:2009:465, pt. 39.

⁶⁴ See, most notably, aside from *Infopaq*: C.J., judgement (Grand Chamber) *Football Association Premier League Ltd and others v QC Leisure and others and Karen Murphy v Media Protection Services Ltd.*, 4th October 2011, joined cases C-403/08 and C-429/08, EU:C:2011:631; C.J., judgement *Eva-Maria Painer v Standard VerlagsGmbH e.a.*, 1st December 2011, C-145/10, EU:C:2011:798; and C.J., judgement *Levola Hengelo BV v Smilde Foods BV*, 13th November 2018, C-310/17, EU:C:2018:899.

⁶⁵ Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, *O.J.E.U.*, L 77, 27th March 1996, art. 3, §1.

originality, when expressed by « free and creative choices »⁶⁶, showing a « personal touch »⁶⁷, and not fulfilled when technical rules or constraints hinder any kind of creative freedom⁶⁸. It also gives an exclusive right to the author to allow or disallow the reproduction of his database⁶⁹.

Under the *sui generis* regime, there is a protection for the owner of a database who « [...] qualitatively and/or quantitatively [put] a substantial investment in either the obtaining, verification or presentation of the contents [...] »⁷⁰ against unlawful extraction or re-utilisation of the whole or a substantial part of the database. Such may occur whenever certain elements are transferred from one database to the other in an effort of collection⁷¹.

But the input stage does not only concern the collection of materials, but also potentially their transformation and adaption for the purpose of de-duplication, consistency, or generally to make information machine-readable⁷². This might infringe a rightholders exclusive right of adaption⁷³.

All the aforementioned points entail, that only facts and data, and anything not protected by the legal framework, can undoubtedly be used as sources of a text and data mining software without leading to any form of liability⁷⁴.

§2. The Output

The final result of a text and data mining project does not usually contain any original parts of the sources used in the process. It creates something newly original which is itself going to be protected by copyright⁷⁵.

The result is however different when the source material is being stored and shared with peers to verify the results. Such practice might infringe the communication to the public right⁷⁶.

⁶⁶ C.J., judgement *Football Dataco Ltd and Others v Yahoo! UK Ltd and Others*, 1st March 2012, C-604/10, EU:C:2012:115, § 38.

⁶⁷ *Ibid.*

⁶⁸ R. DUCATO and A. STROWEL, « Limitations to Text and Data Mining and Consumer Empowerment. Making the Case for a Right to “Machine Legibility” », *CRIDES Working Paper Series*, 31st October 2018, p. 6.

⁶⁹ Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, *O.J.E.U.*, L 77, 27th March 1996, art. 5, (a).

⁷⁰ *Ibid.*, art. 7, §1.

⁷¹ C. GEIGER, G. FROSIO and O. BULAYENKO, *op. cit.*, p. 7.

⁷² J-P. TRIAILLE, J. DE MEEÛS D’ARGENTEUIL and A. DE FRANCQUEN, *op. cit.*, p. 47.

⁷³ *Ibid.*, p. 48.

⁷⁴ C. GEIGER, G. FROSIO and O. BULAYENKO, *op. cit.*, p. 6-7.

⁷⁵ J-P. TRIAILLE, J. DE MEEÛS D’ARGENTEUIL and A. DE FRANCQUEN, *op. cit.*, p. 49.

⁷⁶ C. GEIGER, G. FROSIO and O. BULAYENKO, *op. cit.*, p. 7.

Chapter 2: TDM before and after Directive 2019/790, or when the European Union gets overtaken by its (ex-)Member States

France is one of the few countries in the European Union, which took it upon itself to enact a copyright exception for text and data mining, before the EU itself adopted Directive 2019/790 in April 2019⁷⁷. The four other countries are Germany⁷⁸, the United Kingdom⁷⁹, Estonia⁸⁰ and, to a certain degree, Greece⁸¹.

In the following sections, we will start by explaining the copyright exceptions in conformity of which Member States had to establish their new legislation. Then, we'll go over the three aforementioned text and data mining exceptions in France, Germany, and the United Kingdom respectively before, exploring the context and the content of the relevant articles from the DSM Directive.

Section 1: Copyright framework exceptions used by national legislations

It is important to keep in mind, that even though said Member States took matters into their own hands when regulating text and data mining, they did not operate in a legal vacuum. Indeed, they all had to comply with existing community legislation. We've already mentioned the important copyright directives, providing for the legal scope outside of which Member States may not regulate.

As it is validly explained by the German proposal, which introduced text and data mining into their copyright framework, the first and most important limit to any potential new copyright exception is set out by article 5 of Directive 2001/29/EC (hereafter InfoSoc Directive)⁸². As it

⁷⁷ Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique, *J.O.R.F.*, 8th October 2016, art. 38.

⁷⁸ Gesetz zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft - Urheberrechts-Wissensgesellschafts-Gesetz – UrhWissG, *B.G.Bl.*, 7th September 2017, art. 1, §16.

⁷⁹ The Copyright and Rights in Performances (Research, Education, Libraries and Archives) Regulations 2014, 19th May 2014, Section 3, point 2.

⁸⁰ Säilituseksemplari seadus, *R.T.*, 7th July 2017, art. 24.

⁸¹ Ρύθμιση θεμάτων του Κρατικού Πιστοποιητικού Γλωσσομάθειας, της Εθνικής Βιβλιοθήκης της Ελλάδας και άλλες διατάξεις (Regulation on State Language Certificate subject matter, on the National Library of Greece and on other provisions), *ΦΕΚ (Official Journal)*, 15th February 2017, art. 4(4)(b).

The Greek legislation for text and data mining is different than the others, as it does not provide for a general or specific copyright exception relating to text and data mining, but rather a task for the National Library of Greece to organise, oversee and coordinate national TDM activities (of private or other public actors), especially when they relate to data or metadata of Greek culture. Such a rule seems paradoxical, considering the absence of a proper copyright exception in national law on this matter. See M. BOTTIS *et al.*, « Text and Data Mining in the EU 'Acquis Communautaire' Tinkering with TDM & Digital Legal Deposit », *Erasmus Law Review*, November 2019, n° 2, p.198.

⁸² 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, *O.J.E.U.*, L 167, 22nd June 2001.

is recalled by that Directive's recital n°32, the list contained in that article is exhaustive, thus requiring Member States to fit any new exception into one of these elements. Inside the list, there is no explicit mentioning of text and data mining, but §3, (a) contains an exception for scientific research for non-commercial purposes. Member States have made use of this provision to fit in their exceptions⁸³. The French position regarding the interplay with the InfoSoc Directive is, however, a bit odd: The original project of the law does not introduce an article on text and data mining, neither did therefore the explanatory memorandum mention it. The exceptions we can find in the final text were, instead, adopted by parliamentary amendment⁸⁴, even though the responsible commission and the government voted against it, citing notably the exhaustivity of the list of exceptions from Directive 2001/29/EC⁸⁵. This unfortunately leaves us with little to no explanation on how the parliament felt the interplay with the directive in question. It even promoted some authors to conceive the new text as being in violation of the directive⁸⁶.

Another European directive, relevant for a potential national text and data mining framework, is Directive 96/9/EC on the legal protection of databases⁸⁷. As we've learned from Chapter I, both the sources and the results of a TDM process may be stored in databases, explaining the relevancy of this directive. It contains the same exception about research for non-commercial purposes as Directive 2001/29, which can be found in its article 6, §2, (b), and for the *sui generis* regime in article 9, (b). The German proposal also notably cites Directive 2009/24/EC on the legal protection of computer programs⁸⁸ and Directive 2012/28/EU on certain permitted uses of orphan works⁸⁹, containing provisions which should also be respected. At the same time, it criticises the lack of connectivity between all those directives, leading to apparent insecurities,

⁸³ Entwurf eines Gesetzes zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft, Gesetzesentwurf der Bundesregierung, *Bundestag Drucksachen*, Bundestag, 2016-2017, n° 18/12329, p. 24, point b) (for Germany); Projet de loi pour une République numérique, avis présenté au nom de la Commission des affaires culturelles et de l'éducation, *Dossier législatif*, Ass. Nat., 2015-2016, n° 3389, p. 18 (for France); and Draft Copyright And Rights In Performances (Research, Education, Libraries And Archives) Regulations 2014, explanatory memorandum, HM Government, 2013-2014, p. 5, point 7.9.1 (for the United Kingdom).

⁸⁴ Projet de loi pour une République numérique, amendement, *Doc.*, Ass. Nat., 2015-2016, n° 180.

⁸⁵ Projet de loi pour une République numérique, annales parlementaires, *Doc.*, Ass. Nat., 2015-2016, n° 107, p. 448-449.

⁸⁶ M. BOURGEOIS and G. COUSIN, « Les apports de la loi pour une République numérique en matière de propriété intellectuelle », *La Semaine Juridique Entreprise et Affaires*, n° 49, 8th December 2016, p. 36.

⁸⁷ Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, *O.J.E.U.*, L 77, 27th March 1996.

⁸⁸ Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs, *O.J.E.U.*, L 111, 5th May 2009.

⁸⁹ Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on certain permitted uses of orphan works, *O.J.E.U.*, L 299, 27th October 2012.

especially in the context of cross-border usages⁹⁰. As we'll see in a later section, even with the new exceptions, these insecurities remained, as even the EU itself appears confused about the scope of application of its TDM exceptions within the European copyright landscape (see *infra*⁹¹).

Taking into account the possibilities left to Member States in the aforementioned directives, it was possible to add a new copyright exception for text and data mining within the EU framework. However, as pointed out by FELIX REDA⁹², the limitation to non-commercial and scientific purposes was one of the main reasons why calls for new legislation at the European level arose. In addition, publishers actively tried to hinder users with a lawful access from exercising text and data mining, provoking even more demands for protection and legal security⁹³.

Interestingly, none of the three proposals mention the exception for temporary acts of reproduction, enshrined in article 5, §1 of the *InfoSoc* Directive. Without going too much into detail, the Court of Justice set out four different conditions which all must be cumulatively met in order for this exception to apply: (1) the temporary copy must be transient or incidental, (2) it has to be an integral and essential part of a technological process, (3) its sole purpose must be (a) to enable a transmission in a network between third parties by an intermediary (b) or a lawful use of a work or protected subject-matter and (4) the act must have no independent economic significance⁹⁴. This absence of mention might very well be because, as the authors of the « Study on the legal framework of text and data mining » concluded, the many restrictive conditions only very rarely apply to the copies involved in the mining process⁹⁵.

Section 2: Pre-existing TDM frameworks in Member States

As already exposed in this chapter's introduction, the three pioneer countries we're going to analyse, who overtook the EU before it could create its own text and data mining exception, are France, Germany and the United Kingdom. Even though they all had to obey to the same

⁹⁰ Entwurf eines Gesetzes zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft, Gesetzesentwurf der Bundesregierung, *Bundestag Drucksachen*, Bundestag, 2016-2017, n° 18/12329, p. 24-25, points b) and c).

⁹¹ Part II, Chapter 1, Section 2, §5.

⁹² Formerly Julia Reda, former Vice-President of the Greens at the European Parliament and shadow rapporteur on Directive 2019/790. See X, « Felix Reda », *s.d.*, available at www.greens-efa.eu, accessed on 10th March 2022.

⁹³ Interview conducted with Felix Reda on the 9th of March 2022, annex I, p. 78.

⁹⁴ J-P. TRIAILLE, J. DE MEEÛS D'ARGENTEUIL and A. DE FRANCOUEN, *op. cit.*, p. 42.

⁹⁵ *Ibid.*, p. 50.

European legislation, they nevertheless all took other unique approaches while still sharing certain similarities and differences.

§1. France

French efforts for pushing the digital agenda have been in place since the plan for the creation of a high-speed fibreglass network across the country in 2012. From there on, the government launched in 2014 a national consultation process about the economical and societal challenges of the digital revolution. The result of this inspired the digital strategy of the French government in 2015 and lead to the adoption of the *Loi pour une République numérique* (Law for a digital Republic) in October 2016⁹⁶.

The explanatory memorandum of the law in question cites a twofold objective for the French government: first, jumping ahead in the digital competition against other countries and second, reinforcing the rights of individuals in the digital world. The law therefore not only tackles two new text and data mining exceptions, but also data strategies, general principles on copyright protection on the internet (including further regulation of online platforms) and improvements on digital accessibility⁹⁷.

About text and data mining, article 38 of law n°2016-1321 reads as follows:

« [The author of a work/the rights-holder of a database, may not prohibit:]

[...] 10° Digital copies or reproductions made from a lawful source, with a view to exploring texts and data included in or associated with scientific writings for the purposes of public research, excluding any commercial purpose. A decree shall set the conditions under which text and data mining is carried out, as well as the methods for storing and communicating the files produced at the end of the research activities for which they were produced; these files shall constitute research data;

[...] 5° Digital copies or reproductions of the database made by a person who has lawful access to it, with a view to mining texts and data included in or associated with scientific writings in a research context, excluding any commercial purpose. The conservation and communication of technical copies resulting from processing, at the end of the research activities for which they were produced, are ensured by bodies designated by decree. Other copies or reproductions shall be destroyed. »⁹⁸

There are three important elements in these paragraphs: first, a reproduction (of a work or a database), purely digital in nature, must be made by someone with lawful access to its source. Consequently, mining text and/or data without consent from the rightholder to access the source

⁹⁶ X, « Pour une République numérique », *s.d.*, available at www.gouvernement.fr, accessed on 24th February 2022.

⁹⁷ Projet de loi pour une République numérique, projet de loi, *Dossier législatif*, Ass. Nat., 2015-2016, n° 3318, pp. 3 and 4.

⁹⁸ Personal translation of Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique, *J.O.R.F.*, 8th October 2016, art. 38.

material (through subscription, payment, etc.) is forbidden⁹⁹. This stands in sharp contrast with text and data mining in the United States of America, where such a practice falls under the “fair use” copyright exception¹⁰⁰. Second, the mining may only be done on « texts and data included in or associated with scientific writings ». In the absence of a definition, this further appears to limit the scope of application to texts (no images, or audio-visual content), purely scientific in nature, whatever that science may be (social, formal, natural, etc.), as well as their corresponding metadata¹⁰¹. This can be read together with the third requirement: absence of any commercial purpose.

Such conditions essentially limit the possibility for exercising text and data mining to public research organisations only, conducting their research on any scientific writings, or text and data associated with these, and make their results freely available¹⁰².

However, there is a problem. While at first glance it appears exceptional that the French legislature pushed this topic forward, their advances have died before they could even see the light of day. Indeed, both the first paragraph, notably for the conditions under which text and data mining shall be carried out, and the second paragraph, for designating bodies for the conservation and communication of technical copies, are subject to a decree. Unfortunately, such a decree has never been adopted by the French government, despite promises from former Prime Minister MANUEL VALLS. As of November 2017, its creation was suspended in light of the (at that time thought to be) upcoming directive at the European level¹⁰³. A reason for this was difficult to find, but it has been assessed that the application decree received a negative review from the *Conseil d'État*, because of an incompatibility with the InfoSoc Directive. But since these reviews of governmental decrees are not public, we cannot be certain¹⁰⁴. This should nevertheless not preclude our assessment and comparison of these articles, especially in light of the critical analysis of the new exceptions added in Directive 2019/790, since they share a certain amount of resemblance with the pre-existing French conditions (for a detailed analysis of these new EU provisions, see *infra*, sections 3).

⁹⁹ C. CHRISTOPHE, « République numérique rime avec exceptions et limitations au droit d'auteur », *Revue Mensuelle Lexisnexis Jurisclasseur*, vol. 11, n° 89, November 2016, p. 2.

¹⁰⁰ R. DUCATO and A. STROWEL, « Ensuring Text and Data Mining: Remaining Issues with EU Copyright Exceptions and Possible Ways Out », *E.I.P.R.*, 2021, vol 43, n° 5, p. 334.

¹⁰¹ M. BOURGEOIS and G. COUSIN, *op. cit.*, p. 36-37.

¹⁰² R. JOHNSON, O. FERNHOLZ and M. FOSCI, *Text and data mining in higher education and public research*, Association des Directeurs & personnels de direction des Bibliothèques Universitaires et de la Documentation, Paris, 2016, p. 20.

¹⁰³ X. BERNE, « Loi Numérique: l'exception de « text & data mining » en passe de rester lettre morte », 14th November 2017, available at www.nextimpact.com, accessed on 10th March 2022.

¹⁰⁴ A. LUCAS, « La transposition en France de la directive droit d'auteur dans le marché unique numérique 2019/790 », *A&M*, 2021, n°3, p.309.

On top of this failure to actually apply the new exception in practice, it was also substantially criticised. Authors pointed out a couple of flaws, most notably in its wording and inconsistency amongst the two paragraphs. They question why in the first paragraph the exception mentions « exploration » and in the second it suddenly is « mining », even though both of these words are similar¹⁰⁵. The subsequent decree was supposed to specify the details of both paragraphs, which would have been very much needed for precision. The focus on destruction in the second paragraphs also appears unusually detailed compared to the first. In its entirety, this ‘reform’ consequently seems very limited and confusing at points¹⁰⁶. In addition to a clarification via decree, a focus on text and data mining by the government from the start in the original proposition would have possibly prevented this outcome.

§2. *Germany*

The modification of the German copyright environment, with regards to text and data mining, has been the result of a process specifically focused on reforming the currently applicable copyright rules. The objective was to comply with to present demands of the knowledge-based society. Indeed as the new law’s explanatory memorandum points out, new knowledge is most often based on consulting existing works. They all form the basis necessary to create something new. In this regard, digitalisation has not only accelerated, but also fundamentally changed the way this basis is consulted. As such, adapting the copyright framework to new ways of utilisation was necessary¹⁰⁷.

At its core, the German copyright environment in the digital world, has been profoundly marked by 2001’s InfoSoc Directive, transposed into domestic law in 2003 and 2008, making the whole substantial environment around two decades old. Since its adoption many things have changed and the legal framework has not yet followed suit. The goal of this copyright reform is therefore to not only simplify the structure of German copyright law, but also to reform the currently applicable academic and research exceptions¹⁰⁸. The new reform has not been unanimously received though as publishers in particular, criticise the fact that the academic sector may now digitally make available parts of copyrighted works for free to students and pupils¹⁰⁹. Although

¹⁰⁵ C. CHRISTOPHE, *op. cit.*, p. 2.

¹⁰⁶ M. BOURGEOIS and G. COUSIN, *op. cit.*, p. 37.

¹⁰⁷ Entwurf eines Gesetzes zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft, Gesetzesentwurf der Bundesregierung, *Bundestag Drucksachen*, Bundestag, 2016-2017, n° 18/12329, p. 1.

¹⁰⁸ *Ibid.*, p. 2.

¹⁰⁹ D. PACHAL, « Urheberrecht für Wissenschaft und Bildung: Bundestag beschließt Reformen », 30th June 2017, available at www.irights.info, accessed on 17th March 2022.

the law was supposed to be limited in time up to February 2023, this restriction was later removed by the law transposing Directive 2019/790¹¹⁰.

The new provision on text and data mining, introduced in §60d of the German Copyright Act (hereafter UrhG¹¹¹), reads as follows:

«(1) In order to automatically evaluate a large number of works (original material) for scientific research, it is permissible 1. to reproduce the original material, also automatically and systematically, in order to create a corpus to be evaluated from it, in particular by normalisation, structuring and categorisation, and 2. to make the corpus publicly accessible to a specifically delimited circle of persons for joint scientific research as well as to individual third parties for the purpose of checking the quality of scientific research. In doing so, the user may only pursue non-commercial purposes.

(2) If database works are used in accordance with paragraph 1, this shall be deemed as normal use in accordance with § 55a, first sentence. If insubstantial parts of databases are used in accordance with paragraph 1, this shall be deemed as compatible with the normal exploitation of the database as well as with the legitimate interests of the database producer [...].

(3) The corpus and the copies of the original material shall be deleted upon completion of the research work; the making available to the public shall be terminated. However, it shall be permissible to transfer the corpus and the copies of the original material to the institutions referred to in sections 60e and 60f for permanent storage. »¹¹²

The new rule is not only presented as an exception to the rights of reproduction but also of the right of communication to the public. In addition, the transfer of a corpus to the institutions mentioned in the last paragraph is an exception of the right of distribution¹¹³. Its scope includes works as well as databases, which may both be reproduced in their entirety¹¹⁴.

The addressees are not limited to a specific type of person or entity, provided that the text and data mining is done without any commercial purpose and for scientific research. In that sense, one may also have recourse to third parties for all of parts of the process, as long as the commercial purpose remains excluded¹¹⁵. Such a ban has been criticised by scholarship, since

¹¹⁰ X, « Gesetz zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft (UrhWissG) », 15th June 2021, available at www.bmj.de, accessed on 17th March 2022.

¹¹¹ The official shorting for *Urheberrechtsgesetz*, the German name of the Copyright Act.

¹¹² Personal translation of Gesetz zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft - Urheberrechts-Wissensgesellschafts-Gesetz – UrhWissG, *B.G.Bl.*, 7th September 2017, art. 1, §16, §60d.

¹¹³ W. BULLINGER, « UrhG §60d Text und Data Mining », *Praxiskommentar Urheberrecht*, A.-A. WANDTKE and W. BULLINGER (dir.), München, C. H. Beck, 2019, UrhG §60d, III.

¹¹⁴ K. ANTON, « UrhG §60d Text und Data Mining », *Recht der elektronischen Medien*, G. SPINDLER and F. SCHUSTER (dir.), München, C. H. Beck, 2019, UrhG §60d, II, 2.

¹¹⁵ W. BULLINGER, *op. cit.*, III, 1.

it would « [...] rob the barrier of its practical effectiveness. »¹¹⁶. On the matter of scientific research, it is understood as in §60c¹¹⁷, namely « [...] the striving for new knowledge. »¹¹⁸.

The « original material » includes all types of works (texts, images, audio, videos, etc.), to which, even though it is not explicitly said¹¹⁹, researchers must have a lawful access¹²⁰. The new exception, indeed, does not create a ‘right to access’ to any work¹²¹. A lawful access includes all works which are lawfully in the possession of a researcher, but also all those which are freely accessible on the internet and those to which access has been validly given by a third party, whenever in analogue or in digital form¹²². But the absence of an explicit mentioning of this access requirement raised division in legal scholarship, as some authors consider this condition to be non-existent¹²³. Finally, researchers are allowed to technically modify works contained in the corpus, for example to make them machine readable. This however does not preclude a rightholders right to adapt¹²⁴. In any case, waiving the exception by contract is impossible as stated in §60g, paragraph 1 UrhG¹²⁵.

The new exception as well refers to the undefined adjective of a « large number » of works. Scholarship does not agree whenever it is indeed necessary to create a corpus including a « large number » of works, or whenever it is enough to include only one single work¹²⁶.

The second point of the first paragraph further allows a corpus to be publicly accessible by a « specifically delimited circle of persons ». This includes, for example, peer-review or common projects with other researchers. It is a little controversial whenever this actually counts as the « public »¹²⁷. It should however be noted that a corpus may not be shared with other researchers for projects going further than the one for which it was originally conceived. Additionally, in

¹¹⁶ Personal translation of H. SCHACK, « Das neue UrhWissG – Schranken für Unterricht, Wissenschaft und Institutionen », *Zeitschrift für Urheber- und Medienrecht (ZUM)*, n° 11, 2017, p. 806. For further criticism, see B. RAUE, « Text und Data Mining. Die neue Urheberrechtsschranke des § 60d UrhG », *Computer und Recht (CR)*, n°10, 2017, p. 657.

¹¹⁷ This article provides for an exception to the right of reproduction and communication to the public for scientific research. More precisely, 15% of a work may be used when the research is non-commercial and 75% may be used if the research is personal. Other works of a small scale may also be used completely.

¹¹⁸ Personal translation of K. ANTON, « UrhG §60c Wissenschaftliche Forschung », *Recht der elektronischen Medien*, G. SPINDLER and F. SCHUSTER (dir.), München, C. H. Beck, 2019, UrhG §60d, II, 2.

¹¹⁹ B. RAUE, *op. cit.*, p. 658.

¹²⁰ W. BULLINGER, *op. cit.*, III, 2.

¹²¹ Entwurf eines Gesetzes zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft, Gesetzesentwurf der Bundesregierung, *Doc.*, Bundestag, 2016-2017, n° 18/12329, p. 41.

¹²² B. RAUE, *op. cit.*, p. 658.

¹²³ C. GEIGER, G. FROSIO and O. BULAYENKO, « Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU », in C. GARCIA and R. LLORCA (dir.), *Propiedad Intelectual Y Mercado Único Digital Europeo*, Tirant lo blanch, Valencia, 2019, p. 46.

¹²⁴ W. BULLINGER, *op. cit.*, III, 2.

¹²⁵ B. RAUE, *op. cit.*, p. 661.

¹²⁶ See, for a neutral position (referencing the opposition in footnote n° 20), K. ANTON, *op. cit.*, II, 1.

¹²⁷ W. BULLINGER, *op. cit.*, III, 4; and B. RAUE, *op. cit.*, p. 660.

relation to the second paragraph, substantial parts of databases protected by the *sui generis* regime may not be made publicly available on this basis, but only reproduced¹²⁸.

The second paragraph creates a legal fiction, in case relevant works are contained in databases, by categorising TDM as normal use and as in the legitimate interest of the database producer for the *sui generis* database regime if they are used in accordance with the first paragraph¹²⁹.

Once the project is over, the third paragraph provides for a deletion of any available reproduction, meaning the corpus and a privatisation of the said corpus, therefore cancelling the communication to the public¹³⁰. The exact time at which a project can be considered as concluded is to be determined in the concrete situation and depends on the opinion of the researcher. Of course, this cannot be abused by selecting a very wide and undefined research topic for the sole purpose of being able to conserve the corpus after the true completion of the mining project¹³¹.

Before being deleted, a researcher has the possibility to transmit the corpus to one of the institutions mentioned in §60e and §60f UrhG. This was included as a means to find a balance between the interests of publishers and researchers. While the first wanted to prevent the existence of multiple parallel corpora, the later wanted to preserve their possibility to refer to their works for the sake of scientific quality¹³².

Despite the limited scope of the new exception and some remaining irregularities, the advances have been positively received. Scholars very much welcomed the initiative from the government, but still recalled that the provision would, sooner or later, have to be adapted to the upcoming Directive 2019/790¹³³.

§3. The United Kingdom

In 2014, the United Kingdom was the first Member State of the European Union to enact a copyright exception for the purpose of allowing text and data mining. The reform process had already started in 2010, when a review of the intellectual property framework was commissioned by Prime Minister DAVID CAMERON. The recommendations stemming from this

¹²⁸ K. ANTON, *op. cit.*, II, 2.

¹²⁹ W. BULLINGER, *op. cit.*, III, 6.

¹³⁰ *Ibid.*, III, 7.

¹³¹ B. RAUE, *op. cit.*, p. 659.

¹³² W. BULLINGER, *op. cit.*, III, 7.

¹³³ K. DE LA DURANTAYE, « Neues Urheberrecht für Bildung und Wissenschaft – eine kritische Würdigung des Gesetzentwurfs », *Gewerblicher Rechtsschutz und Urheberrecht*, n°6, 2017, p. 561-562; and H. SCHACK, *op. cit.*, p. 806.

report were accepted by the government in 2011 and provided for extending the copyright exceptions as far as possible within the legal framework of Directive 2001/29/EC. Afterwards, the consultation process took over¹³⁴.

The explanatory memorandum describes how the consultation process led the government to realise how publishers were using contracts to limit the extent to which text and data mining could be carried out. This meant that, without a copyright exception, it would remain necessary to always ask each publisher for permission. The new copyright exception thus seeks to limit the latter's powers, by granting researchers the right to carry out text and data mining for non-commercial purposes¹³⁵.

The new regulation was introduced in Part I, Chapter III, Section 29A of the Copyright, Designs and Patents Act 1988, and is the only provision analysed in this chapter, which has not been subsequently modified through the transposition of the DSM Directive. Indeed, the United Kingdom officially made it clear that it would not transpose the latter. The reason for this being, that with the transitional Brexit period ending on the 31st of December 2020, the transposition deadline for the Directive would fall outside this time frame (July 2021)¹³⁶. In any case, Section 29A reads as follows:

« [...] (1) The making of a copy of a work by a person who has lawful access to the work does not infringe copyright in the work provided that— (a) the copy is made in order that a person who has lawful access to the work may carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non-commercial purpose, and (b) the copy is accompanied by a sufficient acknowledgement (unless this would be impossible for reasons of practicality or otherwise).
(2) Where a copy of a work has been made under this section, copyright in the work is infringed if— (a) the copy is transferred to any other person, except where the transfer is authorised by the copyright owner, or (b) the copy is used for any purpose other than that mentioned in subsection (1)(a), except where the use is authorised by the copyright owner.
(3) If a copy made under this section is subsequently dealt with— (a) it is to be treated as an infringing copy for the purposes of that dealing, and (b) if that dealing infringes copyright, it is to be treated as an infringing copy for all subsequent purposes.
(4) In subsection (3) “dealt with” means sold or let for hire, or offered or exposed for sale or hire.
(5) To the extent that a term of a contract purports to prevent or restrict the making of a copy which, by virtue of this section, would not infringe copyright, that term is unenforceable. »¹³⁷

First, and even though it is called a « copy », the legislation refers to the act of reproduction of a work. The exception here indeed creates an exception to the right of reproduction only. Second, a lawful access to a work is required in order not to infringe the rightholders copyright.

¹³⁴ Draft Copyright And Rights In Performances (Research, Education, Libraries And Archives) Regulations 2014, explanatory memorandum, HM Government, 2013-2014, p. 3.

¹³⁵ *Ibid.*, p. 5.

¹³⁶ Answer to the question asked by J. STEVENS, House of Commons, 2020-2021, 16th January 2020, n°4371.

¹³⁷ Copyright, Designs and Patents Act 1988, c. 48, 15th November 1988, Section 29A.

Third, the copy is made for « computational analysis » (a synonym for text and data mining) « of anything recorded in the work », for non-commercial research purposes.

In contrast to the German provision, transfer of a copy (the UK exception indeed makes no reference whatsoever to the existence of a « corpus ») is not allowed if it is not authorised by the copyright owner. The third paragraph confirms this, by emphasising how such a copy may not be offered, sold, let, or else. An explicit reference to lawful access is, however, made in the first paragraph, bringing this exception a little closer to the French one, and therefore gaining the legal uncertainty that Germany is missing in this regard. Though most importantly, this provision does not cover the reproduction of databases.

Frontrunner at times, the United Kingdom has today fallen behind on its limited TDM exception. Member States of the European Union have, since the transposition of Directive 2019/790 (as we'll see *infra* in the next section) a wider regime than the UK¹³⁸. The Intellectual Property Office therefore started a consultation in 2020 on possible changes to this framework with a view of taking back the lead in the development of a legal framework for AI¹³⁹. The result is a division in two camps: on the one hand are rightholders, apparently understanding the situation of the research sector, but nevertheless longing to keep their economic gains through licences. As such, they stand unfavourable to any extension of the exception. On the other hand, the whole technology and research sector positions itself, of course, in favour of enlargement for the sake of competitiveness, notably with the European Union¹⁴⁰. While the government is trying to find a balance between the two, the impact assessment appears to go towards a collective licencing framework. Whenever or not such allows the UK to take back its place at the front remains to be seen¹⁴¹.

Section 3: The new TDM exceptions from Directive 2019/790

The first sub-section will cover the legislative history of the DSM Directive: where it came from, how it evolved and how it very much attracted the attention of various stakeholders. The

¹³⁸ T. BOND, « UKIPO Consultation on AI and IP: What are the implications for UK copyright law? », 30th November 2021, available at www.twobirds.com, accessed on 15th July 2022.

¹³⁹ X, « Artificial intelligence and intellectual property: call for views », 7th September 2020, available at www.gov.uk, accessed on 15th July 2022.

¹⁴⁰ X, « Government response to call for views on artificial intelligence and intellectual property », 23rd March 2021, available at www.gov.uk, accessed on 15th July 2022.

¹⁴¹ A. TRAPOVA and N. GERVASSIS, « UKIPO's public consultation on AI and IP – text and data mining (Part 2) », 16th March 2022, available at www.copyrightblog.kluweriplaw.com, accessed on 15th July 2022.

second and the third sub-section will subsequently give a short overview of articles 3 and 4 of the Directive respectively. They will be further analysed in the next chapter (see *infra*).

§1. The genesis of a directive to reform copyright in the digital single market

Taking into account the previous sections, it was clear that Member States had divergent approaches to the issue of text and data mining. At that point in time, the legal framework was fractured between different definitions, conditions, and a general lack of uniformity¹⁴². Here is where Directive 2019/790 comes into play, attempting to remediate this issue by creating a harmonised, general framework, applicable to all text and data mining operations.

The idea for a directive to reform the current digital copyright landscape stems from the Commission's « Digital Single Market Strategy for Europe » from May 2015¹⁴³. Indeed, it was part of JUNKER's ten political priorities during his mandate and made up of three pillars: improving access to digital goods and services, creating an environment where digital networks and services can prosper and using digital technologies as a driver from maximising the growth of the EU's economy¹⁴⁴.

Under point 2.4. of the Strategy, the European Commission underlines how necessary a stable and adapted copyright framework is. Especially within an ever-growing digital sector, containing cross-border access to digital services, online purchase of copyright protected content, and research innovations based on the use of text and data mining. The remuneration of creators is also central to the reform process. The Commission therefore pledged to create a new harmonised regime, that allows users to profit from all available offers, while at the same time protecting content creators. A legislative proposal (to notably harmonise copyright exceptions) was due until the end of 2015¹⁴⁵.

The Commission specifically addressed the issue of copyright in the digital single market in a subsequent communication from December 2015¹⁴⁶. It recalled the importance of modern

¹⁴² C. GEIGER, G. FROSIO and O. BULAYENKO, « Text and Data Mining... », *op. cit.*, p. 47.

¹⁴³ Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A Digital Single Market Strategy for Europe, COM(2015) 192 final, 6th May 2015.

¹⁴⁴ X, « What is the Digital Single Market about ? », *s.d.*, available at www.ec.europa.eu, accessed on 22nd March 2022.

¹⁴⁵ Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A Digital Single Market Strategy for Europe, COM(2015) 192 final, 6th May 2015, p. 6-8.

¹⁴⁶ Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Towards a modern, more European copyright framework, COM(2015) 626 final, 9th December 2015.

harmonised rules in the digital environment, to avoid fracture and friction inside the market, and to offer a high level of protection for the competitiveness of Europe's creative industries¹⁴⁷. Point 3 of this document explicitly targets the issue of legal uncertainty regarding text and data mining. The lack of a clear provision on TDM for research and scientific purposes, would harm competitiveness of the EU in the research and innovation sector, as increasing cross-border collaboration is subject to ambiguity¹⁴⁸. Therefore, the Commission undertook to propose, now by Spring of 2016, a legal instrument allowing « [...] public interest research organisations to carry out text and data mining of content they have lawful access to, with full legal certainty, for scientific research purposes [...] »¹⁴⁹.

The proposal for a directive was finally adopted on the 14th of September 2016, with its article 3 on text and data mining reflecting the idea set out in the communication: research organisations may « [...] carry out text and data mining of works or other subject-matter to which they have lawful access for the purposes of scientific research. »¹⁵⁰. Such may not be prevented by contractual terms or technical protection measures (hereafter TPM) except for those necessary to ensure the security and integrity of the networks or databases¹⁵¹.

What follows is a one of the most lobbied EU adoption processes of all time. As it is stated in a press release from the European Parliament: « [s]ome statistics inside the European Parliament show that MEPs have rarely or never been subject to a similar degree of lobbying before [...] »¹⁵², while at the same time classifying it as « excessive »¹⁵³. According to BINCTIN and PRÈS, most lobbying was sparked off in 2018, when the parliamentary rapporteur AXEL VOSS's report was published and adopted by the Committee on Legal Affairs. It introduced a responsibility for content sharing platforms in favour of journalists and authors, and therefore enjoyed major support from collective societies and publishers but, was targeted by aggressive lobbying campaigns from the other side of the spectrum: big companies like Google and Facebook strongly advocated against such a responsibility, which could put their business

¹⁴⁷ Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Towards a modern, more European copyright framework, COM(2015) 626 final, 9th December 2015, p. 1.

¹⁴⁸ *Ibid.*, p. 6-7.

¹⁴⁹ *Ibid.*, p. 8.

¹⁵⁰ Proposal for a Directive of the European Parliament and the Council on copyright in the Digital Single Market, COM(2016) 593 final, 14th September 2016, art. 3, §1.

¹⁵¹ *Ibid.*, §2-3.

¹⁵² J. SCHRANZ, « Questions and Answers on issues about the digital copyright directive », 27th March 2019, available at www.europarl.europa.eu, accessed on 22nd March 2022.

¹⁵³ *Ibid.*

models at risk. Similarly, the driving forces of a free internet, like the Mozilla or Wikimedia foundations, stood beside the tech giants and against the draft¹⁵⁴.

Considering how text and data mining had industry implications on both sides, FELIX REDA explains how companies like Microsoft lobbied in his favour, while notably publishers opposed REDA's position in favour of a more general obligation to allow TDM for research articles. It was however much less lobbied than other elements of the Directive¹⁵⁵.

The final version of the DSM Directive was signed off and adopted on the 17th of April 2019, with its transposition due for the 7th of June 2021 at last¹⁵⁶. This, however, did not mean that the work was over. In hindsight, most Member States (23 from 27) missed this deadline, prompting the Commission to start infringement procedures against them¹⁵⁷. As of August 2022, only 17 Member States have (at least partially) implemented the provisions from the Directive, with many others having draft laws available¹⁵⁸.

The very controversial article 17 has already been debated in front of the Court of Justice¹⁵⁹. An action from Poland against the article in question was brought in May 2019: it allegedly violated the freedom of expression as guaranteed by article 11 of the Charter on fundamental rights. A judgement upholding the article was handed down in April 2022¹⁶⁰. Besides the contentious side, the legislative process on the Digital Services Act will very probably also be relevant for further interpretation of article 17¹⁶¹.

It thus appears that, all things considered, the process following the adoption of the DSM Directive might mirror the difficulties and controversies which it faced while it was in the hands of the European legislator. The second part of this thesis will be dedicated to this subsequent part of the story. As E. ROSATI puts it: « [...] no, things will not get any easier. Hopefully, however, they will become clearer ... in the end. »¹⁶².

¹⁵⁴ N. BINCTIN and X. PRÈS, *Directives 2019/790 et 2019/789 sur le droit d'auteur dans le marché unique numérique*, Bruylant, Bruxelles, 2021, p. 8.

¹⁵⁵ Interview conducted with Felix Reda on the 9th of March 2022, annex I, p. 80.

¹⁵⁶ 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, *O.J.E.U.*, L 130, 17th May 2019, art. 29.

¹⁵⁷ X, « Copyright: Commission calls on Member States to comply with EU rules on copyright in the Digital Single Market », 26th July 2021, available at www.digital-strategy.ec.europa.eu, accessed on 23rd March 2022.

¹⁵⁸ X, « Copyright in the Digital Single Market Directive – Implementation. An EU Copyright Reform Resource », *s.d.*, available at www.create.ac.uk, accessed on 6th August 2022.

¹⁵⁹ *Ibid.*

¹⁶⁰ C.J., judgement *Republic of Poland v European Parliament and Council of the European Union*, 26th April 2022, C-401/19, EU:C:2022:297.

¹⁶¹ E. ROSATI, « The DSM Directive Two Years On: Do Things Ever Get Easier? », *IIC International Review of Industrial Property and Copyright Law*, 2021, vol. 52, pp. 1140 and 1141.

¹⁶² *Ibid.*, p. 1142.

§2. Article 3: TDM for the purpose of scientific research

Before going into the article itself, it should be noted that in article 2, §1, a research organisation is defined as « [...] a university, including its libraries, a research institute or any other entity, the primary goal of which is to conduct scientific research or to carry out educational activities involving also the conduct of scientific research: (a) on a not-for-profit basis or by reinvesting all the profits in its scientific research; or (b) pursuant to a public interest mission recognised by a Member State; in such a way that the access to the results generated by such scientific research cannot be enjoyed on a preferential basis by an undertaking that exercises a decisive influence upon such organisation [...] » and a cultural heritage institution as « [...] a publicly accessible library or museum, an archive or a film or audio heritage institution [...] » (in §3).

The same article also defines text and data mining in its second paragraph, as « any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations ».

The Directive's third article, labelled « Text and data mining for the purposes of scientific research », reads as follows:

« 1. Member States shall provide for an exception to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, and Article 15(1) of this Directive for reproductions and extractions made by research organisations and cultural heritage institutions in order to carry out, for the purposes of scientific research, text and data mining of works or other subject matter to which they have lawful access.

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

3. rightholders shall be allowed to apply measures to ensure the security and integrity of the networks and databases where the works or other subject matter are hosted. Such measures shall not go beyond what is necessary to achieve that objective.

4. Member States shall encourage rightholders, research organisations and cultural heritage institutions to define commonly agreed best practices concerning the application of the obligation and of the measures referred to in paragraphs 2 and 3 respectively. »

First of all, this provision provides an exception to a couple of rights established in the community copyright framework:

- The right of reproduction of the database copyright holder and the rights of reuse and extraction of the producer of a database¹⁶³;
- The rights of reproduction of the copyright holder and the holders of related rights¹⁶⁴;

¹⁶³ Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, *O.J.E.U.*, L 77, 27th March 1996, art. 5, (a) and 7, §1 respectively.

¹⁶⁴ 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, *O.J.E.U.*, L 167, 22nd June 2001, art. 2.

- And the right of reproduction of the press publisher¹⁶⁵.

It therefore covers all usual rights of reproduction and extraction, and goes the furthest in this regard compared to Member States, but does not cover the communication to the public or the distribution rights.

In contrast to the other exceptions adopted prior to its release, not everyone may benefit from this provision. While in Member States, anybody with a lawful access could exercise text and data mining for a scientific purpose, the Directive limits this capacity to research organisation and cultural heritage institutions, as defined in article 2. The beneficiaries must have an intrinsically linked scientific research purpose (covering both natural and human sciences as explained in recital 12), and they may practice TDM solely under that purpose. Consequently, single, independent individuals like lone researchers outside an institution or journalists, may not benefit from this provision. As it stems from recital 11 however, private partners who only assist these institutions (without having access to the results) and persons linked to these institutions, shall also benefit from this exception¹⁶⁶.

When it comes to the conditions under which these addressees may exercise their TDM activities, the Directive is more similar to what we've explored in the legislation of Member States: First, a non-commercial purpose is implied in the definitions of a research organisation and a cultural heritage institution. Second, the scientific research purpose is recalled and emphasised. Third, only works and other subject matter to which they have a lawful access may be used. Regarding this condition, recital 14 details how it covers content available under and open access policy, under contractual arrangements (such as subscriptions) or even any content freely available online. Fourth, all copies shall be kept in a secure environment and may be retained for scientific and verification purposes. Regarding these conditions, article 7, §1 provides that contractual terms opposing this exception shall not be enforceable.

§3. Article 4: TDM for everyone

Compared to article 3, this second exception was nowhere to be found in the initial proposal of the Commission¹⁶⁷. It was later added during the negotiation process, as, most notably, MEP FELIX REDA pushed for a second approach, which would not hinder the practice of commercial

¹⁶⁵ 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, *O.J.E.U.*, L 130, 17th May 2019, art. 15, §1.

¹⁶⁶ N. BINCTIN and X. PRÉS, *op. cit.*, pp. 28 and 29.

¹⁶⁷ Proposal for a Directive of the European Parliament and the Council on copyright in the Digital Single Market, COM(2016) 593 final, 14th September 2016.

TDM. He in fact argues, that the EU should have never created an exception for TDM in the first place, as such acknowledges the practice as a copyright infringing act. It would have been wiser to rather explain that the process of text and data mining just does not give rise to something which could be qualified as a reproduction of the work. But since this option was off the table, a second exception had to be created to make sure that the commercial sector wasn't hindered by the creation of a single exception for non-commercial research purposes only¹⁶⁸.

The fourth article of Directive 2019/790, titled « Exception or limitation for text and data mining » reads as follows:

- « 1. Member States shall provide for an exception or limitation to the rights provided for in Article 5(a) and Article 7(1) of Directive 96/9/EC, Article 2 of Directive 2001/29/EC, Article 4(1)(a) and (b) of Directive 2009/24/EC and Article 15(1) of this Directive for reproductions and extractions of lawfully accessible works and other subject matter for the purposes of text and data mining.
- 2. Reproductions and extractions made pursuant to paragraph 1 may be retained for as long as is necessary for the purposes of text and data mining.
- 3. The exception or limitation provided for in paragraph 1 shall apply on condition that the use of works and other subject matter referred to in that paragraph has not been expressly reserved by their rightholders in an appropriate manner, such as machine-readable means in the case of content made publicly available online.
- 4. This Article shall not affect the application of Article 3 of this Directive. »

Compared to article 3, the list of rights to which an exception may be established is longer. Here, paragraph 1 also mentions the reproduction, translation, adaptation, arrangement, or any other transformation of a computer program (Directive 2009/24/EC). It seems unclear why article 3 did not cover this directive¹⁶⁹ (see *infra*¹⁷⁰).

This time around, however, everyone may benefit from this exception. The scope of beneficiaries is indeed not mentioned, making this available to both private and public legal persons which may want to undertake TDM operations. The purpose of scientific research is also nowhere to be found. The absence of specifically defined beneficiaries also allows the use of text and data mining for commercial purposes, as confirmed by recital 18.

But this does not mean that no restrictions exist. As also stated in article 3, a lawful access to the works used is required, and, on top of that, the rightholder may expressly reserve the possibility of mining his intellectual property. While such a reserve could be machine-readable (and thus requires a minimum amount of investment from the rightholder¹⁷¹), it puts a

¹⁶⁸ Interview conducted with Felix Reda on the 9th of March 2022, annex I, p. 78.

¹⁶⁹ N. BINCTIN and X. PRÉS, *op. cit.*, p. 33.

¹⁷⁰ Part II, Chapter 1, Section 2, §5.

¹⁷¹ Interview conducted with Felix Reda on the 9th of March 2022, annex I, p. 81.

considerable burden upon the TDM executioner. Combined with the fact that article 7 does not preclude the application of contractual limitations, article 4 is broader, but a lot weaker than article 3. It could even be argued that rightholders have a *de facto* opt-out of this provision¹⁷².

Section 4: Conclusive comparison table

In order to fully grasp the small, sometimes subtle, differences between what has been adopted by Member States prior to the European effort and the final version of the DSM Directive, we created a comparison table which summarises the rights to which an exception was made, the elements covered, the beneficiaries, the different conditions which apply, etc. The table can be found in annex III (page 91).

The comparison also concludes this chapter, as the fruit of all research done in previous sections. It is interesting to see, that there are only very few points where all legislators took an identical approach. Logically, everyone targets at least the right of reproduction from the InfoSoc framework, seeing as it is the most general EU copyright framework, but there is already disagreement on what elements are covered: do scientific *writings* (emphasis added) cover images or audio-visual content? The legal doctrine does not agree¹⁷³. Other points of divergence include the difference between research and scientific research, additional conditions, lawful access, conservation and sharing. Even if these disparities may seem minimal, they always have a negative impact on the behaviour of cross-border users. Harmonisation on the EU level was thus necessary, but the manner in which the EU struck a balance between copyright advocates and tech giants, appears nevertheless debatable.

¹⁷² N. BINCTIN and X. PRÈS, *op. cit.*, p. 33-34.

¹⁷³ For a position against this, see *Ibid.*, p. 29. For a position in favour, see V.-L. BENABOU, « La loi pour une République numérique et la propriété intellectuelle », *Dalloz IP/IT*, 2016, n°11, p. 531 (cited by the first source).

Part II: What works and what doesn't – A critical analysis of the Directive's TDM exceptions and Member States' transposition progress

The second part of this thesis is dedicated only to the TDM articles from Directive 2019/790 and their aftermath. We will break down the two new articles and the criticism there are facing (Chapter 1), before moving on to the final step of the legislative process: the transposition by Member States (Chapter 2). Here again, issues are seeing the light of day, not only inside the limits of certain domestic frameworks, but also when comparing the individual results of a select few Member States between each other.

Chapter 1: Two exceptions in troubled waters

This chapter will be dedicated to an in-depth analysis of the two new exceptions, and more specifically to their potential practical issues. Mirroring the presentation of articles 3 and 4, the chapter is structured as follows: the first section will be dedicated to a global overview of the position of different doctrinal authors. It will serve as an introduction to the topics detailed in this chapter. The second section will then be dedicated to the scope of application of the articles, including the definition of research and research organisations, elements tilting the balance in favour of the rightholder (opt-out and lawful access), the absence of other exclusive rights and inconsistencies. In the third section, technical protection measures and contractual limitations will be thematised, before the last section will expose some of our thoughts on the Directive and conclude this chapter.

The purpose of this first chapter is however not only to criticise the approach taken by the European Union. Positive aspects shall be revealed as well. As it was put by AXEL VOSS, EUP Rapporteur for Directive 2019/790, the role of the EU is always to find a balance between the different interests at stake and actors involved¹⁷⁴. As a result of this, as we'll see in the coming sections, not all parties can be satisfied.

Section 1: Global scholarship position overview

In many cases, conclusions from scientific authors who wrote about the new exceptions for text and data mining, are, at least partially, negative:

¹⁷⁴ Interview conducted with Axel Voss on the 6th of April 2022, annex II, p. 88.

B. HUGENHOLTZ argues that their scope of application is too narrow. In particular article 4 would heavily suffer from limitations put up by content holders, leaving software developers, journalists or other intended TDM users with a disadvantage, notably compared to the copyright framework of the United States. He says that the EU may even « [...] have been better served without any of the TDM exceptions in the new Directive. »¹⁷⁵. Similarly, T. MARGONI and M. KRETSCHMER negatively view article 3 as it was proposed by the Commission, their reason again being the narrow scope of application. Indeed, the exception only covers the right of reproduction (not of distribution or communication to the public), and the legislation does not provide enough safeguards against technical protection measures. In conclusion, while some elements, such as the mandatory nature or the inability to waive it by contract are well received, should the current article as it was proposed by the Commission see the light of day, without further changes and adaptation to critics, « [...] perhaps it should be abandoned altogether. »¹⁷⁶. This, of course, didn't happen.

The two research teams R. DUCATO and A. STROWEL, and C. GEIGER, G. FROSIO and O. BULAYENKO both analysed the two new exceptions in detail and argue that the creation of an exception for text and data mining was beneficial to reduce legal uncertainty and drive innovation but may have failed to reach a fair balance between scientific research and commercial applications, notably through an underwhelming scope of article 4. The first team however also proposed potential solutions: interpreting a work as a work subject to the recognition by a human and relying on Member States to do their best during the transposition process¹⁷⁷.

In support of these rather critical view stands a collective of important research institutions, who, in an open letter to the European Parliament in March 2018 criticised article 3 on multiple levels: the scope should be broadened to include any legal or natural person, contractual terms shall remain unenforceable, technical protection measures should clearly be excluded if used in an unreasonable fashion and datasets shall be stored on secure servers¹⁷⁸. Eventually, the scope remained narrow and TPMs also remained somewhat ambiguous (see *infra*, Section 3, §1), but the other critics were addressed.

¹⁷⁵ B. HUGENHOLTZ, « The New Copyright Directive: Text and Data Mining (Articles 3 and 4) », 24th July 2019, available at www.copyrightblog.kluweriplaw.com, accessed on 24th April 2021.

¹⁷⁶ T. MARGONI and M. KRETSCHMER, « The Text and Data Mining exception in the Proposal for a Directive on Copyright in the Digital Single Market: Why it is not what EU copyright law needs », 25th April 2018, available at www.create.ac.uk, accessed on 6th April 2022.

¹⁷⁷ R. DUCATO and A. STROWEL, « Ensuring Text and Data... », *op. cit.*, p. 337; C. GEIGER, G. FROSIO and O. BULAYENKO, « Text and Data Mining... », *op. cit.*, p. 56.

¹⁷⁸ X, « The European Parliament must improve the Text and Data Mining (TDM) exception to benefit European research and innovation », 26th March 2018, available at www.eare.eu, accessed on 24th April 2021.

According to former MEP FELIX REDA, the fact that TDM has now been legally permitted is a great advancement. In the same way, the absence of remuneration or compensation for the usage of the exception is also positive. As for the negative parts, he pinpoints the lack of clarity regarding which technical uses are actually copyright-relevant and the existence of an opt-out in article 4. But at least the machine-readability condition remedies this problem a little, seeing that an active participation from the rightholder is necessary to enforce it. But more generally, the approach should have been simplified, by merging both articles and creating a regime much more favourable to TDM actors¹⁷⁹.

Similarly, although he endorses the compromise found by the EU, MEP AXEL VOSS also envisioned a simpler approach to the issue by making increased use of machine-readable content and promoting a more active rightsholder persona¹⁸⁰.

Section 2: Scope of application

If we take a step back and look at copyright how it was fundamentally intended, we can acknowledge that it has always been about inter-human relationships. It is about human beings, sharing their creations with others, and in that way, promoting one's creativity through the possibility of securing an economic return. These fundamentals are being neglected in the digital world, where it is no longer possible to exploit every single reproduction of work, and where many copies are not being made to follow this tradition of communication between human beings¹⁸¹. As A. STROWEL puts it: « [w]hen the copy of a work (as in the case of purely technical copies) is just a tool in a digital machine, there is no use of the work as a work, meaning as a nexus for a communication between an author and a public. »¹⁸².

In the context of text and data mining, this vision has prompted several scholars to dismiss the idea of a copyright exception altogether, for the sake of it not being necessary. In short, « [...] TDM techniques do not use copyright works as works, they access the information stored in them. »^{183 184}. As we already pointed out previously, FELIX REDA shared the same view, arguing

¹⁷⁹ Interview conducted with Felix Reda on the 9th of March 2022, annex I, pp. 79 and 81.

¹⁸⁰ Interview conducted with Axel Voss on the 6th of April 2022, annex II, p. 87.

¹⁸¹ A. STROWEL, *op. cit.*, p. 207.

¹⁸² *Ibid.*, p. 212-213.

¹⁸³ M. KRETSCHMER and T. MARGONI, « Data mining: why the EU's proposed copyright measures get it wrong », 24th May 2018, available at www.theconversation.com, accessed on 16th July 2022.

¹⁸⁴ See also T. VOLLMER, « The Limits of Copyright: Text and Data Mining », 21st January 2015, available at www.creativecommons.org, accessed on 16th July 2022; P. MURRAY-RUST, « The Right to Read Is the Right to Mine », 1st June 2012, available at www.blog.okfn.org, accessed on 16th July 2022; and E. ROSATI, « An EU text and data mining exception for the few: would it make sense? », *Journal of Intellectual Property Law & Practice*, 2018, vol. 13, n°6, p. 430.

that the EU « [...] should [have] rather clarif[ied] that the act of reproduction does not apply to these kinds of mechanical extractions, because there is no enjoyment of the work. »¹⁸⁵.

But in light of the extensive interpretation of the exclusive rights by ECJ, it appears that an exception for TDM was nevertheless necessary¹⁸⁶. But was it also well drafted? And does it amount to recognising the principle of ‘the right to read is the right to mine’?

This section will cover all problematic elements relating to the scope of application of the two new articles. In this regard, we will first go over the definitions of research and of the beneficiaries article 3 refers to. The second sub-section will then explain how rightholders may opt-out of the exception provided for in article 4. In the third sub-section, we’ll thematise the issue of lawful access and how it tilts the balance in favour of rightholders, before questioning ourselves about the absence of the communication to the public and the distribution rights in sub-section four. The final element of this section will be dedicated to the inconsistencies between, on the one hand, both articles themselves and, on the other hand, with Directive 2009/24/EC.

§1. Problems of research and related organisations

As we know, article 3 refers to « research organisations » and « cultural heritage institutions », who may use text and data mining for the « purpose of scientific research ».

Concerning first the beneficiaries, critics argued that the scope is too narrow¹⁸⁷. It is true that such criticism could be supplemented by the larger article 4, but here the issue is that the opt-out possibility for rightholders seems to cripple any possible advances¹⁸⁸. Indeed, it would have theoretically been possible to enact a broader exception, without giving rightholders such a strong option to hinder text and data mining. It was even considered in the impact assessment and strongly pushed for by the research sector (so-called ‘option 4’)¹⁸⁹. But publishers in particular, feared that a large part of their revenue from licencing would fade away, if a mandatory exception for everybody with lawful access would have been enacted, for both

¹⁸⁵ Interview conducted with Felix Reda on the 9th of March 2022, annex I, p. 79.

¹⁸⁶ T. MARGONI and M. KRETSCHMER, *op. cit.*

¹⁸⁷ *Ibid.*

¹⁸⁸ C. GEIGER, G. FROSIO and O. BULAYENKO, « Text and Data Mining... », *op. cit.*, p. 51.

¹⁸⁹ T. MARGONI and M. KRETSCHMER, *op. cit.* The authors refer to the fourth option proposed in the Impact Assessment, namely a mandatory TDM exception for anybody with lawful access and for any scientific purpose. See Commission Staff Working Document. Impact Assessment on the modernisation of EU copyright rules. Part 1/3, SWD(2016) 301 final, 14th September 2016, p. 109.

commercial and non-commercial purposes¹⁹⁰. These licencing issues were one of the reasons why a TDM exception was adopted in the United Kingdom¹⁹¹. The impact assessment also recognised that text and data mining was most used in Europe, not only by specific research organisations but by the non-business sector as a whole¹⁹².

The vagueness of the definition of « research » also doesn't play in favour of article 3. We've learned that Recital 12 does not give a precise definition but restricts itself to settings out that it relates to natural and human sciences. But what are these, and what specific fields fall under this category? Statistics? Mathematics? Health? Computer science? The lack of concrete examples and references makes it difficult to grasp which fields fall under the exception. As such, the danger lies within the fact that it gets interpreted more narrowly than it actually is. DUCATO and STROWEL therefore highly suggest a clarification by Member States in their transposition¹⁹³. It would have been better defined as « [...] any form of activity carried out in a methodological and systematic way that aims at exploring a certain subject-matter to discover new data or information or to generate new knowledge to advance the state-of-the-art in a certain field »¹⁹⁴.

Together with the non-commercial requirement, article 3 ultimately fails to recognise today's reality of scientific research¹⁹⁵. In fact, many AI research organisations who are using text and data mining are at least partially privately funded¹⁹⁶, like OpenAI which is heavily funded by Microsoft¹⁹⁷. Similarly, lone researchers or journalists who might work closely together with non-profit research institutions will also not be able to profit from this article, even if they do the exact same task as the mentioned research organisations or cultural heritage institutions¹⁹⁸. This might lead to very high licencing costs for those who are not part of these beneficiaries and as such discourage many people to even engage in TDM activities. Some might even resort

¹⁹⁰ Commission Staff Working Document. Impact Assessment on the modernisation of EU copyright rules. Part 1/3, SWD(2016) 301 final, 14th September 2016, p. 107.

¹⁹¹ E. ROSATI, « An EU text and data... », *op. cit.*, p. 429.

¹⁹² *Ibid.*

¹⁹³ R. DUCATO and A. STROWEL, « Ensuring Text and Data... », *op. cit.*, p. 326-327.

¹⁹⁴ K. CHRISTENSEN, « A European solution for Text and Data Mining in the development of creative Artificial Intelligence. With a specific focus on articles 3 and 4 of the Digital Single Market Directive », *Stockholm Intellectual Property Law Review*, 2021, vol. 4, n°2, p. 30-31.

¹⁹⁵ R. DUCATO and A. STROWEL, « Ensuring Text and Data... », *op. cit.*, p. 327.

¹⁹⁶ K. CHRISTENSEN, *op. cit.*, p. 30.

¹⁹⁷ D. ETHERINGTON, « Microsoft invests \$1 billion in OpenAI in new multiyear partnership », 22nd July 2019, available under www.techcrunch.com, accessed on 16th July 2022.

¹⁹⁸ R. DUCATO and A. STROWEL, « Ensuring Text and Data... », *op. cit.*, p. 327.

to illegal, copyright infringing activities outside the exception system, for the simple fact that they might also be precluded from using article 4 because of an opt-out of the rightholder¹⁹⁹.

§2. Article 4's opt-out

The third paragraph of article 4 grants rightholders the option to expressly reserve the possibility to conduct text and data mining on their works. Recital 18 explains this further, setting out that such could appropriately be done by machine-readable means like metadata and terms of service, but also contractually and by unilateral declaration.

This limitation undoubtedly tilts the balance in favour of these rightholders and leaves TDM enthusiasts on the sidelines. In particular, it undermines the principle of 'the right to read is the right to mine'²⁰⁰. Why would somebody with a lawful access be allowed to read as many works as he wishes, but not use a computer to do the same task (better)?

FELIX REDA calls the opt-out a « strange beast »²⁰¹ and that it would « kinda defeat[...] the purpose of having an exception »²⁰². At the same time, the machine-readability requirement would save it. In his opinion, the active participation of the rightholder to setup a working machine-readable system, would be favourable to the exception in situations where a rightholder cannot be bothered with setting it up²⁰³. Of course, such a situation can only arise when there is no contract and no unilateral declaration from the rightholder.

§3. Lawful Access

As we already mentioned about Germany's 2017 TDM exception, the condition of lawful access is not a right to access or to exercise text and data mining on any possible source. It therefore makes a lot of sense to include this requirement. Anything else would have largely negative repercussions on rightholders, which runs contrary to any balanced approach. Recital 14 of the Directive defines it similarly to what B. RAUE had already explained²⁰⁴: there is a lawful access whenever content is accessed under an open access policy, under contractual arrangements or through other lawful means. It also covers content freely available online.

¹⁹⁹ K. CHRISTENSEN, *op.cit.*, p. 30.

²⁰⁰ C. GEIGER, G. FROSIO and O. BULAYENKO, « Text and Data Mining... », *op. cit.*, p. 50.

²⁰¹ Interview conducted with Felix Reda on the 9th of March 2022, annex I, p. 81.

²⁰² *Ibid.*

²⁰³ *Ibid.*

²⁰⁴ B. RAUE, *op. cit.*, p. 658.

The issue nevertheless remains that giving out a lawful access is, again, something only rightholders may do²⁰⁵. In this regard, Recitals 10 and 17 provide for a limited warning against creating separate licences for the beneficiaries of article 3, which would be required in addition to those which research organisations and cultural heritage institutions already possess for their lawful access. But DUCATO and STROWEL warn about the bindingness of the recitals and the apparently optional character of Recital 17. They raise the possibility that publishers might try to circumvent this prohibition of additional licences by simply raising the cost for existing ones²⁰⁶. Other scholars share this view, citing that it might even lead to discrimination among research institutions, between those who receive a lot of money and the others who only have a more limited access to funding. This kind of problem would be somewhat avoided if research institutions backed by private, commercial investors would also be able to make use of article 3²⁰⁷.

Lawful access is also required in the context of article 4, as confirmed by Recital 18. Identically to article 3, a rightholder may restrict the use of text and data mining through machine-readable and other means. But if licencing conditions in terms of service cannot be crawled by a computer, and no other agreements exist, such may mean that a user would have to individually verify each source and its terms of service for potential TDM restrictions. This scenario seems highly impractical²⁰⁸.

A solution to the endless gathering of licences would be the promotion of centralised data pools containing many works on a specific subject matter. These would be specifically created for the purpose of text and data mining by market actors. The licencing possibilities of such ready-made corpora might incentivise rightholders to open up their works to TDM, while at the same time satisfying those who wish for a simpler, and maybe cheaper alternative to individual licencing operations²⁰⁹.

§4. Forgotten other exclusive rights?

Articles 3 and 4 from the Directive only cover the various rights of reproduction and the right of extraction. The two big absent rights are the rights of communication to the public and the right of distribution, which, although present Germany's exception, were excluded right from

²⁰⁵ C. GEIGER, G. FROSIO and O. BULAYENKO, « Text and Data Mining... », *op. cit.*, p. 54.

²⁰⁶ R. DUCATO and A. STROWEL, « Ensuring Text and Data... », *op. cit.*, p. 330.

²⁰⁷ C. GEIGER, G. FROSIO and O. BULAYENKO, « The Exception for text... », *op. cit.*, p. 22.

²⁰⁸ R. DUCATO and A. STROWEL, « Ensuring Text and Data... », *op. cit.*, p. 330.

²⁰⁹ K. CHRISTENSEN, *op. cit.*, p. 32.

the get-go in the impact assessment²¹⁰. It does nevertheless appear in the recommendations from the EU Parliament's experts, who called for an exception « [...] enable[ing] storing and communication of research files created for TDM. »²¹¹.

This raises questions on whenever or not such an exclusion is justified. According to some authors, neither the communication to the public, nor the distribution rights are relevant for text and data mining. They argue, first, that the TDM process does not involve a distribution of the source material to a public, since only those involved in the analysis will have access to the data. Second, and most importantly, the results should not amount to a distribution or a communication to the public of the underlying elements, either because they are only statistics, new knowledge, or, because the parts reproduced from the analysed texts are insufficient to amount to an infringement of either exclusive right²¹².

Others however share the opposing view, arguing that it would have been much better to have a larger ground for this exception, covering also the two mentioned rights. As a reason, they cite, that certain areas like natural language processing could involve the sharing of corpora including more than what is necessary to trigger a copyright infringement (citing more than 11 consecutive words²¹³). Including the two exclusive rights in the exception would put it more in line with what has already been set out in the exception for parody and quotation, while at the same time being more innovation friendly²¹⁴.

Whenever an exception to both rights is truly necessary remains doubtful. It should be noted and recalled that the German exception is very limiting in this regard, since only the corpus may be shared, not the original works. In addition, it does not, in fact, seem like the limited circle of persons, to which access may be given, fits the definition of a 'public', namely an « [...] indeterminate number of potential viewers and implice[ing], moreover, a fairly large number of people [...] »²¹⁵. The inclusion was nevertheless welcomed for the sake of legal security²¹⁶.

²¹⁰ The communication to the public right was already excluded in the second option present in the assessment, without any further explanation. The distribution right is not mentioned in the text and data mining section. See Commission Staff Working Document. Impact Assessment on the modernisation of EU copyright rules. Part 1/3, SWD(2016) 301 final, 14th September 2016, p. 108.

²¹¹ C. GEIGER, G. FROSIO and O. BULAYENKO, « The Exception for Text... », *op. cit.*, p. 25.

²¹² J-P. TRIAILLE, J. DE MEEÛS D'ARGENTEUIL and A. DE FRANQUEN, *op. cit.*, p. 33. This view appears to be shared by DUCATO and STROWEL, see R. DUCATO and A. STROWEL, « Ensuring Text and Data... », *op. cit.*, p. 323, footnote 13.

²¹³ Referring to the European Court of Justice's fundamental Infopaq judgement.

²¹⁴ T. MARGONI and M. KRETSCHMER, *op. cit.*

²¹⁵ C.J., judgement *Stichting Brein v Jack Frederik Wullems*, 26th April 2017, C-527/15, EU:C:2017:300, §32

²¹⁶ B. RAUE, *op. cit.*, p. 660.

§5. Inconsistencies

DUCATO and STROWEL further pointed out a set of inconsistencies which we will briefly mention in this sub-section. The first concerns the DSM Directive itself, and the second Directive 2009/24/EC²¹⁷ (hereafter Computer Program Directive).

They first explain how, both articles 3 and 4 contain an odd exception to the press publisher's right of reproduction, from the newly established article 15, §1 of Directive 2019/790. Such an exception would only make sense, if it covered the services used by the TDM exception's beneficiaries for research purposes, such as web crawling. That is because article 15 neither grants a press publisher any right against a third party, nor do research organisations and cultural heritage institutions easily qualify as information society service providers²¹⁸.

Second, they find it inconsistent how only article 4 covers the Computer Program Directive. To answer why, the team suggests that research organisations or public heritage institutions can already benefit from the observance and studying exception from article 5, §3 of said Directive. A 'research exception' could therefore already exist, but for anything going further, article 4 DSM Directive would subject this to the rightholder's possible reservation²¹⁹.

But this would still be problematic as both exceptions do not cover exactly the same grounds which would be necessary for text and data mining. In this case, the option of resorting to article 4 would also not be satisfactory, seeing how easily it can be restricted. If we were to apply both article 4 DSM Directive and article 5 Computer Program Directive, it could potentially lead to a paradoxical situation where all activities under the first article may be restricted by contract, but not the activities under the second, even though they would be similar²²⁰.

Section 3: Limitations

This third section will cover potential technological and contractual limitations to which the DSM Directive does not offer a fully satisfactory solution. Technical protection measures, especially, are practical burdens which may pose important obstacles to TDM enthusiasts, particularly when they are deeply embedded into the system, hosting the works they would like

²¹⁷ Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs, *O.J.E.U.*, L 111, 5th May 2009.

²¹⁸ R. DUCATO and A. STROWEL, « Ensuring Text and Data... », *op. cit.*, p. 328.

²¹⁹ *Ibid.*, p. 328.

²²⁰ *Ibid.*, p. 328-329.

to have access to. In such cases, even a lawful access as such may not be sufficient to mine anything.

§1. Technical Protection Measures (TPMs)

Copyright has always been envisaged as a monopoly for an author to exploit his work and gain an economic advantage through it. This exploitation right is nevertheless not absolute. Legislation has more and more provided for exceptions, to effectively allow those with legitimate claims to access and take advantage of author's creative output. In this context, our digitalised world has made access and transfer of protected works, in high quality, much simpler than it was before. Consequently, technical protection measures were created to limit and prevent unauthorised access to digital works. While they had a positive, legitimate intention, they can just as well prejudice access for those which do have a right and a legitimate claim²²¹.

With InfoSoc Directive 2001/29, the European Union attempted to regulate this phenomenon in article 6, §4. It established a system in which rightholders shall take appropriate, voluntary measures in order to make available, to the beneficiaries of the exceptions, the works covered. If such measures do not see the light of day, there is a subsidiary obligation for Member States to ensure that rightholders make their works available²²².

Neither « voluntary measures », nor the length Member States must go in their 'enforcement' had been defined by the Directive, prompting criticism about the effectiveness of such provision when it was adopted²²³. Over two decades later, it turns out that the criticism was justified. Indeed, Member States have not consistently adopted these measures to protect the beneficiaries of the exceptions, which do not dispose of a legal basis to circumvent an illegally established technical protection measure²²⁴. As such, it may hinder or prevent their access to works²²⁵. Article 7 of Directive 2019/790, nevertheless, makes this same ineffective system also applicable to the new TDM exceptions. It is true, that the system appears suitable on paper, but the practical reality of the subject matter questions, whenever it was a well thought-out decision

²²¹ A. ASHOK, « Technological Protection Measures and Copyright Exceptions: The Rights of the Common Man in the Developed World », *Cochin University Law Review*, 15th June 2011, p. 100-101.

²²² S. DUSOLLIER, « Exceptions and Technological Measures in the European Copyright Directive of 2001 - An Empty Promise », *International Review of Intellectual Property and Competition Law*, 2003, vol. 34, n°1, p. 71.

²²³ *Ibid.*, p. 75.

²²⁴ T. MARGONI and M. KRETSCHMER, *op. cit.*

²²⁵ C. GEIGER, G. FROSIO and O. BULAYENKO, « Text and Data Mining... », *op. cit.*, p. 55.

to refer to the old framework, instead of establishing a mandatory prohibition of TPMs, blocking the usage of copyright exceptions²²⁶.

A last point in this regard could be made on « measures to ensure the security and integrity of the networks and databases » from article 3, §3. These measures are exclusively concerned with the technical aspects of cybersecurity of the system, where works and databases are hosted²²⁷. According to Recital 16, these may include IP address validation and user authentication and involve a proportionality test between the security and integrity measures necessary and the risk involved. C. GEIGER and his team compare it to similar safeguards set out by telecom operators for the purpose of network security and in case of network congestion. Recital 16 appears to refer to the latter, citing « a potentially high number of access requests »²²⁸. These measures should (normally) not hinder access to works.

§2. Contractual overwriting of copyright exceptions

One positive element of the Directive which should be highlighted, is the impossibility to overwrite article 3 by contractual means. This inability to waive the exception by contract is central to its effective application. The EU here took the example of Belgium, where all copyright exceptions are already ‘imperative’²²⁹. Unfortunately, as already explained earlier, this does not cover article 4.

The importance of this should nonetheless be stressed. Some might argue that it removes all possibility for parties to negotiate, and that users may find it more difficult to be granted licences for the exploitation of one or multiple works. However, an extensive interpretation of the reproduction right has, already in the past, lead to restrictions concerning computer programs. The exceptions for decompilation, and for observance and study, were therefore added in the Computer Program Directive. Notably, their mandatory character stemmed from the reasoning, that copyright should not be used as an obstacle to the ideas and the information underlying the creative expression. Such also appears very much relevant in the context of text and data mining, where a similar access-reasoning is present²³⁰ (see also the introduction to Section 2, *supra*).

²²⁶ R. DUCATO and A. STROWEL, « Ensuring Text and Data... », *op. cit.*, p. 329.

²²⁷ *Ibid.*, p. 330.

²²⁸ C. GEIGER, G. FROSIO and O. BULAYENKO, « Text and Data Mining... », *op. cit.*, p. 55.

²²⁹ J-P. TRIAILLE, J. DE MEEÛS D’ARGENTEUIL and A. DE FRANQUEN, *op. cit.*, p. 108, referring to Code de droit économique, *M.B.*, 28th February 2013, art. XI.193.

²³⁰ *Ibid.*, p. 109.

Section 4: Conclusive thoughts

There are a lot of negative elements to point out about these European copyright changes. Whenever it is the narrow, sometimes inconsistent, scope of application, or practical limitations because of TPMs and contracts, there are many points which can be criticised.

At the same time, it is not because this evolution is not perfect, that necessarily the EU got it wrong²³¹. When the European legislator is acting, it is always about finding a balanced approach between all opinions stemming from the market. In certain situations, these opinions may be rather far away from each other, and every effort to reconcile them will be difficult. The outcome of EU legislation is therefore, in many cases, a trade-off, usually resulting in criticism from at least one side of the spectrum²³². In this regard, scholarship welcomes the changes but finds them insufficient²³³, failing to address the AI revolution²³⁴, and doubtful of their efficiency²³⁵.

This, in turn, does not mean that no positive elements could be found. There are indeed elements which are good, and which should not be neglected underneath the stack of negative impressions perceived:

- The sole existence of these exceptions and their harmonisation, brings legal security and certainty to market actors²³⁶;
- The absence of remuneration/compensation makes the exceptions much more accessible to those who might lack proper funding²³⁷;
- Both exceptions are mandatory and article 3 may not be waived by contract²³⁸;
- And article 4 is available to everyone²³⁹.

In our opinion, a necessary step had to be taken to ensure legal certainty. In the context of the European copyright framework, it was hardly conceivable that the system would suddenly completely turn itself on its back and, for example, allow text and data mining for everyone without a lawful access. It would have been an opportunity to go in this direction, but the strong protection of the rightholder could not just disappear overnight, especially in light of the

²³¹ *A contrario*, M. KRETSCHMER and T. MARGONI, *op. cit.* The article's title is quite literally « [...] why the EU's proposed copyright measures get it wrong ».

²³² Interview conducted with Axel Voss on the 6th of April 2022, annex II, p. 89.

²³³ C. GEIGER, G. FROSIO and O. BULAYENKO, « Text and Data Mining... », *op. cit.*, p. 56-57.

²³⁴ K. CHRISTENSEN, *op. cit.*, p. 32-33.

²³⁵ R. DUCATO and A. STROWEL, « Ensuring Text and Data... », *op. cit.*, p. 337.

²³⁶ Interview conducted with Felix Reda on the 9th of March 2022, annex I, p. 79.

²³⁷ *Ibid.*

²³⁸ R. DUCATO and A. STROWEL, « Ensuring Text and Data... », *op. cit.*, p. 331.

²³⁹ *Ibid.*, p. 332.

bombardment of lobbying campaigns which MEPs faced. We do lay faith in the words of FELIX REDA and AXEL VOSS, who clearly explained us, how they tried their best to find the most suitable compromise solution. At the end of the day, for the system to really change, one would have to abandon the old, ‘analogue’ vision of copyright, and start interpreting the exclusive economic rights more in line with our digital world of the 21st Century: less focused on every single copy, but in turn concentrating on what truly matters: communicating with other human beings.

All things considered, the DSM Directive wasn’t the end of the line. What followed was the period of national transposition, and hopes that Member States potentially correct some inaccuracies and oversights from the European legislator. This is especially true since they were entitled to go further than the Directive, where the EU copyright framework allowed them to²⁴⁰.

²⁴⁰ R. DUCATO and A. STROWEL, « Ensuring Text and Data... », *op. cit.*, p. 332.

Chapter 2: Diverse national transpositions

In this second Chapter, we will thematise the results from the efforts made by a certain amount of Member States, to transpose the TDM articles from the DSM Directive. Out of the all 27, most have by now fully implemented the changes, others are still lacking behind.

The first section will provide an overview of the situation, the second section will then go into detail about the new provisions of a select few Member States before concluding the chapter with two comparison tables. As we will see, many similar regimes came out of the transposition process. But (sometimes significant) differences nevertheless remain.

Section 1: Overview

When the DSM Directive was adopted, the deadline for its transposition was set out to be the 7th of June 2021²⁴¹. As we've already covered (see *supra*²⁴²), only four Member States transposed the Directive before this date, leading the Commission to start infringement procedures against the others²⁴³. As of the 19th of May 2022, out of all 'bad students', 13 still did not yet notify the Commission of any transposition measures regarding Directive 2019/790, prompting the latter to continue the infringement procedures by sending out reasoned opinions. Starting from there, all Member States not in compliance had two months until the Commission may refer the matter to the Court of Justice. France and Belgium are notably part of that list²⁴⁴.

But this information is a little out of date, since both France and Belgium have now adopted legislative instruments to introduce the (remaining) articles of Directive 2019/790 into their domestic legal framework, on the 23rd of June²⁴⁵ and the 16th of June 2022²⁴⁶ respectively. In

²⁴¹ 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, *O.J.E.U.*, L 130, 17th May 2019, art. 29.

²⁴² Part I, Chapter 2, Section 3, §2.

²⁴³ X, « Copyright: Commission calls on... », *op. cit.*

²⁴⁴ X, « Copyright: Commission urges Member States to fully transpose EU copyright rules into national law », 19th May 2022, available at www.ec.europa.eu, accessed on 20th July 2022.

²⁴⁵ Décret n° 2022-928 du 23 juin 2022 portant modification du code de la propriété intellectuelle et complétant la transposition de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE, *J.O.R.F.*, 24th June 2022.

²⁴⁶ Projet de loi transposant la directive (UE) 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE, projet de loi, *Doc.*, Chambre, 2022-2023, n° 2608/009, 16th June 2022. The project has been voted by the parliamentary assembly on the 16th of June 2022, but as of the 31st of July 2022, it has not yet been published to the *Moniteur belge*.

contrast, Germany had already fully transposed the Directive on the 31st of May 2021²⁴⁷, ranking in third place amongst all Member States²⁴⁸.

E. ROSATI argues that the causes of failure to meet the deadline might have been different elements, such as Brexit or the global disruption caused by the COVID-19 pandemic, which significantly stalled, not only national legislative processes, but also the EU itself²⁴⁹. The ongoing war in Ukraine and the energy crisis stemming from it will probably also add their fair share of disruption to countries still wrestling with the transposition.

Section 2: Post Directive 2019/790 national TDM frameworks

As with the analysis in the second Chapter of Part I, we will first start with France in sub-section 1 and then move on to Germany in sub-section 2. As the United Kingdom has not transposed the DSM Directive, we will, instead, focus on the Netherlands, Spain and Ireland.

The analysis will cover the important elements, and potential changes compared to the Directive and the old system, that had been put in place before the legislative intervention at the EU level, if there ever was such a system, of course. This, all the while referring to the elements we might have already analysed in a previous section.

§1. France

The context of Directive 2019/790 is one of the rare occasions where the French Republic has anticipated the elements of European legislation, instead of, as usual, being rather slow and often missing deadlines²⁵⁰. From a more methodological point of view, the French legislature has chosen to create a rather central provision in the general copyright framework, and additionally, new provisions individually for database copyright, the *sui generis* regime, and neighbouring rights. They also chose to fragmentally implement most articles by way of ordonnances and not through one single law, which has the advantage of speeding up the legislative process, but arguably limits the explanations accompanying the instruments and therefore the possibilities of interpretation²⁵¹.

²⁴⁷ Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, *BGBl*, 4th June 2021.

²⁴⁸ X, « DSM Directive Implementation... », *op. cit.*

²⁴⁹ E. ROSATI, « The DSM Directive Two... », p. 1140.

²⁵⁰ A. LUCAS, *op. cit.*, p.307.

²⁵¹ *Ibid.*, p. 308.

The authors of the *Conseil Supérieur de la Propriété Littéraire et Artistique* (CSPLA – Superior Council for Literary and Artistic Property²⁵²) furnished valuable advice to the French government for correctly transposing the two text and data mining exceptions. As such, in the third part, the report from the CSPLA first recalls how the government does not have much leeway when transposing directives about the single market, and how the new exceptions from the DSM Directive are different than those from the InfoSoc Directive, since they are both mandatory²⁵³.

A. Modifications of the general copyright regime

In any case, on the 3rd of December 2020, the French legislature adopted a law on the adaption of the domestic legal framework to the law of the EU in economic and financial matters. This law is the first DSM Directive transposition measure, as it allows the government, through its article 34, to modify the *Code de la Propriété intellectuelle* (Intellectual Property Code, hereafter CPI), with the aim of implementing Directive 2019/790. The aforementioned article nevertheless splits the transposition in two parts: one relates to articles 2, (6) and 17 to 23 of the Directive, one relates to all the other articles (except for article 15, which the law itself transposes)²⁵⁴. Apart from speeding up the process, the reason for this split is not entirely clear. It can however be noted that the very controversial article 17 was also supposed to be part of a law, as it was the case here for article 15, but because of the pandemic the law was abandoned²⁵⁵. Article 34 also sets out two different time limits for the two implementations: the first ordinance had to be enacted no longer than six months after the release of the law, the other no longer than twelve months after it. Both had to be ratified by the parliament²⁵⁶.

This led to the two ordinances n° 2021-580²⁵⁷ and n° 2021-1518²⁵⁸, the latter including the articles on text and data mining. In the *Livre I* on copyright, the ordinance first completely

²⁵² A consulting instance which gives advice to the Ministry of Culture on literary and artistic property matters. Its work very much relates to the newest questions of copyright in the digital world and gives a possibility to all copyright-relevant actors to make their voice heard. See X, « Conseil supérieur de la propriété littéraire et artistique », *s.d.*, available at www.culture.gouv.fr, accessed on 22nd July 2022.

²⁵³ A. BENSAMOUN and Y. BOUQUEREL, *Rapport de mission. Transposition des exceptions de fouille de textes et de données : enjeux et propositions*, Paris, CSPLA, 2020, p. 46-47.

²⁵⁴ Loi n° 2020-1508 du 3 décembre 2020 portant diverses dispositions d'adaptation au droit de l'Union européenne en matière économique et financière, *J.O.R.F.*, 4th December 2020, art. 34, I.

²⁵⁵ A. LUCAS, *op. cit.*, p. 308.

²⁵⁶ Loi n° 2020-1508 du 3 décembre 2020 portant diverses dispositions d'adaptation au droit de l'Union européenne en matière économique et financière, *J.O.R.F.*, 4th December 2020, art. 34, II and III.

²⁵⁷ Ordonnance n° 2021-580 du 12 mai 2021 portant transposition du 6 de l'article 2 et des articles 17 à 23 de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE, *J.O.R.F.*, 13th May 2021.

²⁵⁸ Ordonnance n° 2021-1518 du 24 novembre 2021 complétant la transposition de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE, *J.O.R.F.*, 25th November 2021.

replaces the old regime by simply adding copies and reproductions for the purpose of text and data mining to the list of exceptions, saying that such will be detailed in article L122-5-3²⁵⁹. It does appear odd to mention both copies and reproductions, since they are perfect synonyms in these circumstances²⁶⁰. Regardless, there are a few new articles which were created for TDM.

The first, and most important one, L122-5-3 CPI, reads as follows:

« I - Text and data mining, within the meaning of Article L. 122-5, 10°, is defined as the implementation of any automated analytical technique aimed at analysing text and data in digital form in order to generate information, notably patterns, trends and correlations.

II - Digital copies or reproductions of works that have been lawfully accessed may be made without the authors' authorisation, with a view to text and data mining carried out solely for the purposes of scientific research, by research organisations, libraries accessible to the public, museums, archive services or institutions that are depositories of the cinematographic, audio-visual or sound heritage, or on their behalf and at their request by other persons, including within the framework of a non-profit partnership with private actors.

The provisions of the previous paragraph are not applicable when a company, shareholder or associate of the organisation or institution carrying out the mining, has privileged access to their results.

Digital copies and reproductions made during text and data mining shall be stored with an appropriate level of security and may be retained for the sole purpose of scientific research, including for the verification of research results.

Copyright owners may implement proportionate and necessary measures to ensure the security and integrity of networks and databases in which works are hosted.

An agreement concluded between the organisations representing the copyright holders and the organisations and institutions mentioned in the first paragraph of this II may define the best practices relating to the implementation of its provisions.

III - Without prejudice to the provisions of II, digital copies or reproductions of lawfully accessed works may be made for the purpose of text and data mining by any person, regardless of the purpose of the mining, unless the author has objected in an appropriate manner, notably by machine-readable processes for content made available to the public online.

Copies and reproductions shall be stored with an appropriate level of security and destroyed after the text and data mining has been completed. »²⁶¹

The French transposition appears relatively faithful at first, transposing the definition in I, article 3 in II and article 4 DSM Directive in III. There are nevertheless a few elements, which we should highlight. First, we can note that the definition has been slightly altered from the Directive. TDM is said to be the « *implementation of any automated analytical technique...* » (we emphasised), the two first words having been added. Similarly, « which includes but is not limited to » from the definition in article 2, (2) of the Directive have been replaced by the word « notably ». These changes appear to make the definition slightly easier to read, and do not seem to change in any way its scope of application.

²⁵⁹ Code de la Propriété intellectuelle, *J.O.R.F.*, 3rd July 1992, art. L122-5, 10°.

²⁶⁰ A. LUCAS, *op. cit.*, p. 310.

²⁶¹ Personal translations of Code de la Propriété intellectuelle, *J.O.R.F.*, 3rd July 1992, art. L122-5-3.

What might slightly affect the scope though is that cultural heritage institutions are not mentioned in the list in II. Instead, the French legislature chose to add the list of institutions from article 2, (3) of the Directive. It might seem clearer that way, but it is still missing the generic, encompassing notion.

The CSPLA and author A. LUCAS further seem to suggest that research organisations and cultural heritage institutions were not on equal footing in the Directive, more precisely in Recital 11 on the matter of public-private partnerships, as well as in Recital 12 on their non-commercial nature and collaboration with commercial partners. Indeed, they argue that the Directive only reserves these elements to research organisations²⁶². But, neither did we, nor did other scholars reveal this inconsistency in the analysis of the Directive²⁶³. We therefore perceive this as a too literal reading of the Directive and approve the French transposition's exclusion of commercial partnerships and privileged access for private actors, for both research organisations and cultural heritage institutions alike.

Second, the inability to waive the exception by contract, provided for in article 7, §1 of the Directive, is missing. It might be implied in the wording of III, which contains the transposition of article 4 and is written as a contrast to II, but it is certainly not explicit and therefore appears as a mistaken implementation²⁶⁴. This might be due to the fact that the CSPLA recommended to only set out the basic principles in the legislative instrument and then later the conditions of application separately²⁶⁵. At the same time, their suggested final text nonetheless reads that « any stipulation contrary to the provisions [...] shall be null and void »²⁶⁶.

Lastly, the French executive slightly twisted the condition of lawful access. Where the Directive cites « have [a] lawful access », L122-5-3 CPI cites « have been lawfully accessed ». This would imply that the first covers all works and other subject matter to which a lawful access theoretically exists, but the latter covers those which have already been lawfully accessed in practice. The Directive's approach would thus be broader, covering even those to which a lawful access exists, but which has not been effectively exercised yet. The concrete implications of this literal reading nevertheless remain undetermined.

²⁶² A. LUCAS, *op. cit.*, p. 309-310; and A. BENSAMOUN and Y. BOUQUEREL, *op. cit.*, p. 61.

²⁶³ See *supra*, Part II, Chapter I.

²⁶⁴ A. LUCAS, *op. cit.*, p. 309.

²⁶⁵ A. BENSAMOUN and Y. BOUQUEREL, *op. cit.*, p. 6.

²⁶⁶ *Ibid.*, p. 89.

B. Modifications of computer program, neighbouring and database rights

As the transposition method chosen was to implement the two articles from the Directive into their respective parts of the copyright framework, there are multiple new articles scattered around the French Intellectual Property Code. As such, for computer programs, the new article is L122-6-1, VI CPI, which simply refers to L122-5-3 for the same activities mentioned in the Directive.

For neighbouring rights, the Ordinance of the 24th of November 2021 added point 8° to article L211-3 CPI, which also refers to article L122-5-3. The rights covered are those also mentioned in article 2 of Directive 2001/29 as well as the new press publishers right introduced by the DSM Directive. Similar to the definition and the TDM article as such, 8° refers to « digital copies *or* reproductions » (we emphasised). We refer to the comment made *supra* on this matter.

For database rights, there is against a reference made back to L122-5-3 in article L342-3, 6° CPI.

C. Comparison with the 2016 regime

As the CSPLA recommended²⁶⁷, what was introduced by article 38 of the *Loi pour une République numérique* was removed by the Ordinance. The French legislature therefore made no effort to reconcile the new provisions from the Directive with the already existing system. Seeing how little was changed from the Directive in the transposition, a comparison between the old and the new regime would almost amount to comparing the 2016 regime to the articles from Directive 2019/790, which is something we already did (see *supra*²⁶⁸ and *infra*, Annex III).

The replacement does not come as a surprise, as it was already established in 2019 that the French exception would have to be re-written to accommodate for the changes made by the Directive²⁶⁹.

D. Release of the application decree for clarification

On the 23rd of June 2022, the long-awaited application decree²⁷⁰ for the Ordinance of November 2021 released. It specifies some elements which were left vague in the former text, such as a

²⁶⁷ A. BENSAMOUN and Y. BOUQUEREL, *op. cit.*, p. 6.

²⁶⁸ Part I, Chapter 2, Section 4.

²⁶⁹ C. ALLEAUME, « Les exceptions en faveur de l'enseignement et de la recherche scientifique (Commentaire des articles 3 à 5 de la directive) », *Communication Commerce Électronique*, 2019, n°10, 3, p. 3.

²⁷⁰ Décret n° 2022-928 du 23 juin 2022 portant modification du code de la propriété intellectuelle et complétant la transposition de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur

right to information for rightholders or how precisely the opt-out from L122-5-3, III should take place²⁷¹.

The Decree is not revolutionary and confirms the approach which had already been taken by the legislature. It does however precise some interesting elements for practitioners. For example, in the new article R122-23 CPI, persons who may benefit for the TDM exception for research are those which are linked to research organisations, or the other institutions mentioned in the article, through internal procedures or contracts. It also goes over certain elements which must be present in contracts with private actors or with institutions whose mission will be to store the corpora. Interestingly, also private actors may serve in this function, specifically those who « [...] provide these institutions with data storage, computing or electronic communications network services and equipment. »²⁷².

As already mentioned, the Decree likewise adds information obligations for both beneficiaries and rightholders. Whereas the first must furnish the latter with the justifications, on the basis of which they exercise text and data mining²⁷³, the latter are under the obligation to inform the first of network security and integrity measures, as well as how text and data mining may be exercised²⁷⁴.

At last, the Decree explains how the opt-out must not be motivated and can be expressed by any means, such as through machine-readable possibilities or a web site's terms of service²⁷⁵.

Contrary to the recommendation from the CSPLA²⁷⁶, the Decree did not precise the notion of lawful access, not even integrating Recital 14 from the Directive. The inability to waive to waive the exception by contract is also not further thematised, leaving this issue very much ambiguous.

§2. Germany

After yet another, comparatively highly mediated legislative procedure, the German *Bundestag* finally adopted on the 20th of May 2021 its new law transposing Directive 2019/790 into the German Copyright Act. As it was foreseeable, many interventions from and discussions

et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE, *J.O.R.F.*, 24th June 2022.

²⁷¹ O. WANG, « Décret complétant la transposition de la directive DAMUN : la gestion collective étendue précisée », 5th July 2022, available at www.dalloz-actualite.fr, accessed on 22nd July 2022.

²⁷² Personal translations of Code de la Propriété intellectuelle, *J.O.R.F.*, 3rd July 1992, art. R122-24.

²⁷³ *Ibid.*, art. R122-25 and R122-27, for the research exception and the exception for everyone respectively.

²⁷⁴ *Ibid.*, art. R122-26, for the research exception only.

²⁷⁵ *Ibid.*, art. R122-28.

²⁷⁶ A. BENSAMOUN and Y. BOUQUEREL, *op. cit.*, p. 87.

with the public were made on the topic of the contentious article 17 DSM Directive, which revealed great potential for conflict about the whole subject matter, and how difficult it was to find a balance between all opposing opinions²⁷⁷. For a new law, that is probably the most substantial reform of the German Copyright Act since its inception in 1966²⁷⁸, the level of public debate appears nonetheless acceptable.

For text and data mining, the new « Law on the Adaption of Copyright to the Requirements of the Digital Single Market »²⁷⁹ modified the pre-existing §60d UrhG, which remains the exception for scientific research (article 3 DSM Directive), and added §44b which concerns the exception for everyone (article 4 DSM Directive). Methodologically, the transposition is similar to the French, as it creates two central exceptions in the main copyright regime, to which other articles from elsewhere refer. It however takes the form of a legislative act, as opposed to the executive ordinances and decree seen in the previous sub-section. The division in two separate articles can be explained by the tradition of the German Copyright Act, which has always split general exceptions (§44a ff.) from research exceptions (§60a ff.)²⁸⁰.

A. §44b - The general exception

The newest addition to the Germany Copyright Act is §44b. It reads as follows:

« (1) Text and data mining is the automated analysis of single or multiple digital or digitised works, in order to extract information from them, in particular about patterns, trends and correlations.
(2) Reproductions of lawfully accessible works for text and data mining are permitted. The reproductions shall be deleted when they are no longer necessary for text and data mining.
(3) Uses according to paragraph 2 sentence 1 are only permissible if the rightholder has not reserved them. A reservation of use in the case of works accessible online shall only be effective if it is made in machine-readable form. »²⁸¹

This new addition first contains the German definition for text and data mining, to which also §60d refers. It is also an adapted version of the definition contained in article 2 of the Directive, covering all types of works, whenever « single or multiple ». With this wording, the definition

²⁷⁷ F. KLEINKOPF and T. PFLÜGER, « Digitale Bildung, Wissenschaft und Kultur – Welcher urheberrechtliche Reformbedarf verbleibt nach Umsetzung der DSM-RL durch das Gesetz zum Urheberrecht im digitalen Binnenmarkt? », *Zeitschrift für Urheber- und Medienrecht (ZUM)*, 2021, n° 8/9, p. 643.

²⁷⁸ J. B. NORDEMANN, J. PUKAS and J. WAIBLINGER, « Transposition of the DSM Copyright Directive 2019/790 in Germany », *A&M*, 2021, n°3, p. 325.

²⁷⁹ Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, *BGBl*, 4th June 2021.

²⁸⁰ W. BULLINGER, « UrhG §60d Text und Data Mining », *Praxiskommentar Urheberrecht*, A.-A. WANDTKE and W. BULLINGER (dir.), München, C. H. Beck, 2022, III, 1.

²⁸¹ Personal translation of Gesetz über Urheberrecht und verwandte Schutzrechte (Urheberrechtsgesetz), *BGBl*, 9th September 1965, §44b.

explicitly responds to the criticism of the 2017 definition, where there was uncertainty about whenever single works were also inside its scope of application (see *supra*)²⁸².

Further than what is contained in the DSM Directive, the definition also mentions « digitalised » works. The legislature here established that text and data mining may not only be conducted on already digitally available works, but also on analogue works which are digitalised for the sole purpose of TDM. Of course, this provision may not be used to collect works with the goal of creating a personal, parallel archive²⁸³.

On its scope of application, the German legislature chose to transpose it faithfully, covering the works, databases and computer programs (§69d, (4) UrhG) copyright, as well as the *sui generis* database regime (§87c, (1), 4 UrhG).

The second paragraph does not substantially differ from what is provided for in article 4 of the Directive: it is not linked to a particular purpose and as such, commercial uses are permitted if there is a lawful access. The latter notion was not defined, thus its understanding remains identical to what was set out before the Directive (see *supra*)²⁸⁴. Furthermore, §44b does not profit from the possibility of sharing the corpora with a limited circle of people (as §60d allows). It also sets out that a corpus must be deleted once it is no longer required. It has been debated whenever such a requirement could cease²⁸⁵, for example by contractually modulating the requirement with rightholders. It is seen as an important element to be aware of, since the creation of substantial corpora often takes a lot of time and work²⁸⁶.

The third paragraph does reserve a positive surprise for the exception's beneficiaries. The second sentence, about a possible opt-out from the rightholder, provides that, for content freely available online, such may only be effectively made « in machine-readable form ». It was included to make sure that text and data mining may indeed be made automatically. As such, it can also be included in the terms of service, but the latter must in that case be machine-readable²⁸⁷. The negative formulation of the first sentence nevertheless places the burden of

²⁸² B. RAUE, « Die Freistellung von Datenanalysen durch die neuen Text und Data Mining-Schranken (§§ 44b, 60d UrhG) », *Zeitschrift für Urheber- und Medienrecht (ZUM)*, 2021, n° 10, p. 795.

²⁸³ Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, Gesetzesentwurf der Bundesregierung, *Bundestag Drucksachen*, Bundestag, 2020-2021, n° 19/27426, p. 88.

²⁸⁴ B. RAUE, « Die Freistellung von Datenanalysen... », *op. cit.*, p. 796.

²⁸⁵ W. BULLINGER, « UrhG §44b Text und Data Mining », *Praxiskommentar Urheberrecht*, A.-A. WANDTKE and W. BULLINGER (dir.), München, C. H. Beck, 2022, III.

²⁸⁶ B. RAUE, « Die Freistellung von Datenanalysen... », *op. cit.*, p. 796.

²⁸⁷ Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, Gesetzesentwurf der Bundesregierung, *Bundestag Drucksachen*, Bundestag, 2020-2021, n° 19/27426, p. 89.

proof on the user²⁸⁸. This ‘addition’ is however only reserved to the circumstance of content freely available online. For all other situations, it must only be « [...] explicitly stated and done in a manner appropriate to the automated processes involved in text and data mining. »²⁸⁹. B. RAUE argues in this regard, that it would have been better to include the explanation from Recital 18 of the DSM Directive in a clearer way, seeing how difficult and time consuming it can be to verify all possible opt-outs from rightholders in extensive corpora²⁹⁰.

It should be noted that the inclusion of a possible machine-readable prohibition of TDM on a website, should not have unjustified, negative effects on the rightsholder for other uses, such as classification on a search engine²⁹¹.

B. §60d - The exception for scientific research

Next in line is the updated version of §60d UrhG. It reads as follows:

« (1) Reproductions for text and data mining (Section 44b(1) and (2), first sentence) shall be permitted for the purposes of scientific research in accordance with the following provisions.

(2) Research organisations shall be entitled to make reproductions. Research organisations are universities, research institutes or other institutions which conduct scientific research, provided they 1. do not pursue commercial purposes, 2. reinvest all profits in scientific research or 3. operate in the public interest within the framework of a state-recognised mandate.

Not entitled under sentence 1 are research organisations which cooperate with a private enterprise which has a determining influence on the research organisation and preferential access to the results of scientific research.

(3) Also entitled to make reproductions are 1. libraries and museums, provided they are open to the public, as well as archives and institutions in the field of film or audio heritage (cultural heritage institutions), 2. individual researchers, provided they do not pursue commercial purposes.

(4) Authorised persons pursuant to subsections (2) and (3) who pursue non-commercial purposes may make reproductions pursuant to subsection (1) publicly available to: 1. a specifically delimited circle of persons for their joint scientific research, and 2. individual third parties for the purpose of reviewing the quality of scientific research.

As soon as the joint scientific research or the review of the quality of scientific research has been completed, the making available to the public shall be terminated.

(5) rightholders under subsections (2) and (3)(1) may retain reproductions under subsection (1) with appropriate safeguards against unauthorised use for as long as they are necessary for the purposes of scientific research or the review of scientific knowledge.

²⁸⁸ M. STIEPER, « Die Schranken des Urheberrechts im Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes », *Zeitschrift für Urheber- und Medienrecht (ZUM)*, 2021, n°10, p. 778.

²⁸⁹ Personal translation of Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, Gesetzesentwurf der Bundesregierung, *Bundestag Drucksachen*, Bundestag, 2020-2021, n° 19/27426, p. 89.

²⁹⁰ B. RAUE, « Die Freistellung von Datenanalysen... », *op. cit.*, p. 797.

²⁹¹ W. BULLINGER, « UrhG §44b Text und... », *op. cit.*, IV.

(6) rightholders shall be authorised to take necessary measures to prevent the security and integrity of their networks and databases from being compromised by reproductions under paragraph 1. »²⁹²

The new article tries, as much as possible, to stick with the pre-existing framework. Consequently, it retains the core aspects of the 2017 exception, all the while clarifying some aspects and updating it to be in conformity with article 3 DSM Directive²⁹³.

In the first paragraph, it is recalled that reproductions are permitted for the purpose of scientific research. The definition is the one which can be found in §44b UrhG, which we already covered (see *supra*, A.). The German legislature chose not to define « scientific research », thereby referring to the distinction between natural and human sciences from Recital 12 DSM Directive, and, as such, not answering the criticism of this distinction (see *supra*²⁹⁴)²⁹⁵. However, the notion of « lawful access » was added from the Directive, refuting previous criticism on the possible absence of such a condition in the old version.

As with §44b UrhG, the scope of this article covers all types of copyright protected works and databases, as well as neighbouring rights and the *sui generis* database regime (§87c, (1), 5 UrhG). Expressly excluded are nevertheless computer programs (§69d, (6) UrhG). Even though it was said, that the absence of the Computer Programs Directive in the list from article 3 DSM Directive was an inconsistent legislative oversight (see notably *supra*²⁹⁶), the German legislature did not want to risk being sanctioned by the ECJ, which might hold a different view²⁹⁷.

The second and the third paragraphs concern the beneficiaries of the exception. We can see that the conditions were transposed in accordance with the Directive, while adding precision, notably on « research institutes and other institutions » and the availability of this exception for « individual researchers »²⁹⁸. The latter is particularly important as the legislature here went further than the Directive, basing himself on article 5, §3 of the InfoSoc Directive. Such additions are expressly allowed by the DSM Directive (art. 24, §2). The limitation to a purely non-commercial purposes for individual researchers shall not be precluded when these beneficiaries receive a professional fee²⁹⁹.

²⁹² Personal translation of Gesetz über Urheberrecht und verwandte Schutzrechte (Urheberrechtsgesetz), *BGBI*, 9th September 1965, §60d.

²⁹³ M. STIEPER, *op.cit.*, p. 776.

²⁹⁴ Part II, Chapter 1, Section 2, §1.

²⁹⁵ W. BULLINGER, « UrhG §60d Text und... », 2022, *op. cit.*, III, 1.

²⁹⁶ Part II, Chapter 1, Section 2, §5.

²⁹⁷ B. RAUE, « Die Freistellung von Datenanalysen... », *op. cit.*, p. 800.

²⁹⁸ W. BULLINGER, « UrhG §60d Text und... », 2022, *op. cit.*, III, 3, a).

²⁹⁹ *Ibid.*, III, 3, c).

This type of addition is unique and answers the criticism faced by the Directive about excluding individual researchers, such as journalists, from the list of possible beneficiaries (see *supra*³⁰⁰). It is also a continuation of the old §60d, from which everyone was able to profit³⁰¹.

The fourth paragraph also retains the idea of continuing what had already been imagined in 2017. The exception to the right of communication to the public, which had already been inserted in the old §60d, is indeed reiterated in this paragraph, enabling researchers to work together in a larger team and share corpora for the purpose of peer review³⁰². As the DSM Directive does not include such an exception, it is, again, based on article 5, §3 of the InfoSoc Directive³⁰³.

But the criticism from the old regime remains applicable to this new provision: it is unclear whenever the delimited circle of people actually represents the « public » (see *supra*³⁰⁴). However, the fact that the legislature kept this exception, even though he clearly stated that even loose groups of scientists may share their materials³⁰⁵, might make it look like a broader circle of persons, similar to a public, would be targeted here. But the wording is clearly in contradiction with the European concept of a public. In consequence, it appears that this exception primarily exists for the sake of clarity and legal security³⁰⁶.

Paragraph five is another evolution for the German copyright framework, which hitherto did not enjoy a possibility to retain corpora past the end of a research project. The legislature did nevertheless slightly sway away from the precise wording of the DSM Directive. In fact, the transposition requires a condition of necessity to be fulfilled in order for retentions to be allowed. This cannot be found in the original text, which exclusively limits it only to the purpose of text and data mining³⁰⁷.

Such a limitation may make it seem as if the retention was limited in time. But the legislature holds a different view, explaining that permanent storage is possible³⁰⁸. The reason for this

³⁰⁰ Part II, Chapter 1, Section 2, §1.

³⁰¹ Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, Gesetzesentwurf der Bundesregierung, *Bundestag Drucksachen*, Bundestag, 2020-2021, n° 19/27426, p. 97.

³⁰² B. RAUE, « Die Freistellung von Datenanalysen... », *op. cit.*, p. 799.

³⁰³ Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, Gesetzesentwurf der Bundesregierung, *Bundestag Drucksachen*, Bundestag, 2020-2021, n° 19/27426, p. 97.

³⁰⁴ Part I, Chapter 2, Section 2, §2.

³⁰⁵ Entwurf eines Gesetzes zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft, Gesetzesentwurf der Bundesregierung, *Bundestag Drucksachen*, Bundestag, 2016-2017, n° 18/12329, p. 39.

³⁰⁶ W. BULLINGER, « UrhG §60d Text und... », 2022, *op. cit.*, III, 4.

³⁰⁷ *Ibid.*, III, 5.

³⁰⁸ Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, Gesetzesentwurf der Bundesregierung, *Bundestag Drucksachen*, Bundestag, 2020-2021, n° 19/27426, p. 97.

ambiguous choice of words is even less clear, if compared to the old §60d, whose third paragraph expressly mentioned « permanent storage »³⁰⁹. It is argued that the new law might have mistakenly mixed up the condition of necessity from article 4 DSM Directive with the transposition of the Directive's article 3, and therefore restricted its scope of application³¹⁰. The basic principles of good scientific practice would require materials to be saved for at least ten years, while also keeping them findable, accessible, interoperable and reusable (FAIR)³¹¹.

On the matter of accessibility, it still remains unclear how permanent storage interplays with paragraph four, and how stored corpora should be made accessible for control to independent third parties, without creating parallel archives. This question is of particular importance, since the whole purpose of permanent store is to ensure later accessibility³¹². The release of new best practices, some of them having to be updated from the 2017 regime, may hopefully clarify the topic³¹³.

The exclusion of individual researchers from the retention of corpora has also been criticised as hindering the verifiability of results stemming from this type of research projects³¹⁴.

The sixth paragraph does not require particular comment. We can refer to what we explained about these security aspects in the DSM Directive (see *supra*³¹⁵).

C. Conclusion

Our review of the new TDM provisions shows how an effort was made to consolidate the DSM update with the old system. The German legislature continued its TDM exception to the communication to the public right, and clarified some criticised elements from the DSM Directive, by imposing machine-readability for freely available websites and allowing the research exception even for individual researchers, not linked to a research organisation or a cultural heritage institution. Some ambiguities nevertheless remain, notably about the delimited circle of persons, permanent storage and accessibility.

³⁰⁹ W. BULLINGER, « UrhG §60d Text und... », 2022, *op. cit.*, III, 4.

³¹⁰ H. HEESSEN and L. JÜNGELS, « Der Regierungsentwurf der Text und Data Mining-Schranken (§§ 44b, 60d UrhG-E) », *Recht und Zugänge (RZ)*, 2021, n°1, p. 50.

³¹¹ F. KLEINKOPF, J. JACKÉ and M. GÄRTNER, « Text- und Data-Mining. Urheberrechtliche Grenzen der Nachnutzung wissenschaftlicher Korpora bei computergestützten Verfahren und digitalen Ressourcen », *Multimedia und Recht (MMR)*, 2021, n°3, p. 197.

³¹² F. KLEINKOPF, J. JACKÉ and M. GÄRTNER, *op. cit.*, p. 198.

³¹³ W. BULLINGER, « UrhG §60d Text und... », 2022, *op. cit.*, III, 4.

³¹⁴ Evaluierungsbericht der Bundesregierung zu den durch das Urheberrechts-Wissengesellschafts-Gesetz reformierten Vorschriften der §§ 60a bis 60h des Urheberrechtsgesetzes, *Bundestag Drucksachen*, Bundestag, 2022-2023, n° 20/1825, p. 36.

³¹⁵ Part II, Chapter 1, Section 3, §1.

Yet, Germany clearly went into the right direction. This is confirmed by the 2017 regime's evaluation report from 2022, where stakeholders were permitted to express their opinion on, notably, the new TDM provisions. They globally express a positive view on the practicability of the reform in concrete situations³¹⁶. And as usual in copyright matters, it is the role of the legislature to find a fair balance between the (sometimes heavily) confronting interests of all involved stakeholders. In any case, no systematic application problems were revealed during the consultation process. The goal of the law, notably to reorganise the permissible uses for education and science, has thus been reached³¹⁷.

§3. Other Member States

After having tackled the most interesting evolution of France and Germany, we are going to have a short look at progress from other Member States, to exemplify other types of transpositions.

The first country to be analysed is also the first to have transposed the Directive: The Netherlands. Second in line is Spain with a, for now provisional, but nevertheless odd transposition. Our third and last candidate will be Ireland, the only Member of the EU still combining civil and common law frameworks.

A. The Netherlands

The pioneer of the DSM Directive adopted its implementation act on the 16th of December 2020³¹⁸. Its entry into force was set for the 7th of June 2021, with exceptions for the transposition of articles 16 and 19 of the Directive³¹⁹. The Dutch legislature reportedly attempted to remain as close as possible to pre-existing legislative provisions³²⁰, something which of course wasn't possible in the context of text and data mining.

Similar to what we've seen in France and Germany, the Netherlands opted for two new centralised articles in the Copyright Act³²¹ (15n and 15o), with one article, respectively from

³¹⁶ Evaluierungsbericht der Bundesregierung zu den durch das Urheberrechts-Wissensgesellschafts-Gesetz reformierten Vorschriften der §§ 60a bis 60h des Urheberrechtsgesetzes, *Bundestag Drucksachen*, Bundestag, 2022-2023, n° 20/1825, p. 34.

³¹⁷ *Ibid.*, p. 54.

³¹⁸ Implementatiewet richtlijn auteursrecht in de digitale eengemaakte markt, *Staatsblad*, 29th December 2020.

³¹⁹ Besluit van 18 december 2020 tot vaststelling van het tijdstip van inwerkingtreding van de Implementatiewet richtlijn auteursrecht in de digitale eengemaakte markt, *Staatsblad*, 29th December 2020, art. 1.

³²⁰ S. VAN GOMPEL, « Omzetting van de DSM Auteursrecht Richtlijn 2019/790 in Nederland », *A&M*, 2021, n°3, p. 338.

³²¹ Auteurswet, *Staatsblad*, 23rd September 1912.

the Database Act³²² (4a, points a and b) and the Neighbouring Rights Act³²³ (10, points p and q), referring to them.

The transposition as a whole is very faithful to the Directive. Certain elements, such as the three main definitions (TDM, research organisations and cultural heritage institutions) are even completely identical to their DSM Directive's counterparts³²⁴. We were only capable of finding one oversight about excepting research TDM from the new press publisher right: article 10, p. from the *Wet op de naburige rechten*, establishes the list of neighbouring rights to which the TDM exception from article 15n applies. Here, the word *perspublicatie* (press publication) is nowhere to be found. It only appears in the subsequent article 10, q. about 15o and text and data mining for everyone.

The Netherlands might have made a very fast transposition, but in matters of text and data mining they did only exactly that: implementing the content of the Directive, nothing more, just a little less. They had the opportunity to go further and slightly adapt the provisions, but they chose not to, apparently favouring a faster procedure. This happened, even though things like the inconsistent scope of application with the Computer Programs Directive were discussed³²⁵.

B. Spain

The Kingdom of Spain chose to transpose Directive 2019/790 through a Royal Decree on the 2nd of November 2021³²⁶. This peculiar instrument of implementation is only used in case of urgency (probably justified by surpassing the 7th of June 2021 transposition deadline) and must subsequently be validated by the parliament³²⁷ to fully enter into effect. During this validation procedure, the parliament rejected the text from the Royal Decree, instead favouring a proper

³²² Datenbankenwet, *Staatsblad*, 20th July 1999.

³²³ Wet op de naburige rechten, *Staatsblad*, 18th March 1993.

³²⁴ S. VAN GOMPEL, *op. cit.*, p. 339.

³²⁵ Voorstel van de implementatiewet richtlijn auteursrecht in de digitale eengemaakte markt, memorie van toelichting, *Kamerstukken II*, Tweede Kamer, 2019-2020, n°35454, p. 18.

³²⁶ Real Decreto-ley 24/2021, de 2 de noviembre, de transposición de directivas de la Unión Europea en las materias de bonos garantizados, distribución transfronteriza de organismos de inversión colectiva, datos abiertos y reutilización de la información del sector público, ejercicio de derechos de autor y derechos afines aplicables a determinadas transmisiones en línea y a las retransmisiones de programas de radio y televisión, exenciones temporales a determinadas importaciones y suministros, de personas consumidoras y para la promoción de vehículos de transporte por carretera limpios y energéticamente eficientes, *Boletín Oficial del Estado*, 3rd November 2021.

³²⁷ R. XALABARDER, « Transposition of the DSM Copyright Directive 2019/790 in Spain », *A&M*, 2021, n°3, p. 357.

democratic consultation process³²⁸. As such, the provisions are only provisional, but nevertheless interesting the analyse.

Apart from the definition taken with only very slight modification from the Directive³²⁹, the sole addition for text and data mining is article 67 of the Decree. Contrary to what we've seen in other Member States, there is only one single, centralised article, transposing both articles 3 and 4 of the DSM Directive, without making a proper distinction between the exception for scientific research and the exception for everyone. The third paragraph therefore defeats the nature of article 3's transposition, as it subjects all beneficiaries to a rightholder's reservation³³⁰.

Additionally, the sixth paragraph seems to create a wider exception to computer programs than what was provided for in the Directive by allowing it for both article 3 and article 4 situations³³¹. This welcomed element is unfortunately overshadowed, by the problematic mixture which the Spanish legislature created.

Taking into account the forgoing, it appears as if Spain failed to correctly transpose the TDM provisions. One can only hope that the parliamentary procedure corrects its issues.

C. Ireland

The Republic of Ireland did not only rank seventh in the 'transposition race', it was also part of the countries which had a pre-existing text and data mining provision³³². We did not cover this provision previously, because of its similarity the UK provision³³³.

In any case, the Irish legislature published its copyright update on the 19th of November 2021, with the provisions on text and data mining included in Sections 3 and 4³³⁴. As it was with Germany, the legislature attempted to stick as closely as possible to the already enacted framework. But it's even more visible here, as certain parts are left completely unchanged.

³²⁸ A. MATAS, « Copyright directive series: progress in the directive transposition », 21st April 2022, available at www.pro.europeana.eu, accessed on 28th July 2022.

³²⁹ Real Decreto-ley 24/2021 de 2 de noviembre, *Boletín Oficial del Estado*, 3rd November 2021, art. 66, point 1.

³³⁰ R. XALABARDER, *op. cit.*, p. 359.

³³¹ *Ibid.*

³³² C. BRADY *et al.*, « Copy That! Ireland Implements EU Copyright Directive », 17th December 2021, available at www.williamfry.com, accessed on 28th July 2022, p. 1.

³³³ It indeed provides for the same scope, beneficiaries, conditions, safeguards, etc., but is only slightly differently structured and worded. For more information, see E. ROSATI, « Irish Government proposes introduction of Irish text and data mining exception », 14th March 2018, available at www.ipkitten.blogspot.com, accessed on 28th July 2022.

³³⁴ European Union (Copyright and Related Rights in the Digital Single Market) Regulations 2021, *Iris Oifigiúil*, 19th November 2021.

Section 53A of the Irish Copyright and Related Rights Act 2000 (hereafter CRRA) remains the exception for non-commercial research in the general copyright framework and is joined by the new Section 53B on text and data mining for commercial purposes. These two are partially reproduced about neighbouring rights in Sections 225A and 225AA, as well as for computer programs in Section 82 CRRA. The new act also includes new provisions concerning TDM and the press publisher rights³³⁵. All in all, it is a very much decentralised framework which contains no cross-references between relevant articles, as it was the case with previously analysed Member States. Combining both the new and the old system is, however, made difficult in the absence of an official consolidated Copyright Act.

Substantially, the provisions go a lot further than the Directive. By keeping parts of the old, open to everyone, system, the legislature included no mention of research organisations or cultural heritage institutions in the text and data mining sections. They are both only mentioned in the new regulations on the press publisher rights (regulations 14 and 15). It is true that cultural heritage institutions also make an appearance in other parts of the 2021 Act, but they relate to other parts of the DSM directive³³⁶. As such, non-commercial text and data mining for research purposes, remains possible for everyone³³⁷.

But there are even more interesting aspects. First, the Irish TDM exception for commercial text and data mining also imposes, like Germany's §44b UrhG, machine-readability for content publicly made available online³³⁸. Second, the legislature introduced an obligation of information, benefiting the author of a work, who may be informed of the making of a copy and the appropriate security measures taken. He may also request additional security measures³³⁹. Third, it is argued that the Irish legislature wanted to anticipate commonly agreed best practices about security measures and wanted to make sure that rightholders were able to protect their networks and databases accordingly. They thus reinforced and detailed the relating provisions³⁴⁰. At last, even the commercial purposes exception cannot be waived by contract³⁴¹.

³³⁵ European Union (Copyright and Related Rights in the Digital Single Market) Regulations 2021, *Iris Oifigiúil*, 19th November 2021, regulations 14 and 15.

³³⁶ *Ibid.*, regulations 8-11.

³³⁷ Copyright and Other Intellectual Property Law Provisions Act 2019, *Iris Oifigiúil*, 26th June 2019.

³³⁸ European Union (Copyright and Related Rights in the Digital Single Market) Regulations 2021, *Iris Oifigiúil*, 19th November 2021, regulations 4, (a), Section 53B, (3), (a); (b), Section 82, (5), (a); (c), Section 225AA, (3), (a); and regulation 15, (3), (a).

³³⁹ *Ibid.*, regulation 3, (1), Section 53A, (3C); and (2), Section 225A, (3C).

³⁴⁰ C. BRADY *et al.*, *op. cit.*, p. 2.

³⁴¹ European Union (Copyright and Related Rights in the Digital Single Market) Regulations 2021, *Iris Oifigiúil*, 19th November 2021, regulations 4, (a), Section 53B, (6).

We, consequently, discovered a legislature who remained faithful to their original provisions, while at the same time adapting to the new additions. When it comes to the research exception, it can hardly be overlooked how Ireland created the broadest scope of application of all Member States we previously analysed.

§4. Conclusive comparison table

As we covered many different national provisions in detail, this chapter will again be concluded by a comparison. This time it is organised in two separate tables, the first covering a comparison between article 3 DSM Directive and corresponding national provisions, the second covering article 4 DSM Directive and relevant provisions from the domestic frameworks we've previously analysed. The two tables make it much easier to notice the differences between the European instructions and the national outcomes. They can both be found in annexes IV (p. 92) and V (p. 94) respectively.

These two large centralised comparisons also conclude the current chapter, as it was already the case with the second chapter of Part I. One might even say, « it's like poetry, sort of. It rhymes. Each stanza kind of rhymes with the last one. »³⁴². The structure of the second chapters of Part I and II is indeed very much similar, reflecting the evolution from legally uncertain and fragmented domestic frameworks, to newly created, closer, yet still different national frameworks after significant development at the European level.

The tables clearly reflect the impossibility for Member States to ignore their domestic system in favour of a unified European approach. This is not necessarily a bad thing as such. The European Union lives from its divergences. But if the objective was to create a digital single market, where stakeholders could benefit from the same rules in each Member State, it has not been reached. Certain approaches are simply too divergent to be reconciled. One will have to await further changes and precision, most likely through litigation at the Court of Justice.

What comes now remains to be seen, but as we know from the European copyright framework as a whole: it is not because the final words of a legislative texts have found their way into the official journal, that everything has been set in stone... Far from it, in fact.

³⁴² Quote from GEORGE LUCAS, 1999.

Conclusion

While the legislative text and data mining framework has already come a long way from where it was seven years ago, it is still very much in its infancy. As such, it might still be very early to judge on both its background and its future. Compared to the framework from the InfoSoc Directive, where the *Infopaq* judgement from the ECJ came eight years after it was adopted, we are, very probably, only just at the beginning of the journey.

The absence of a judgement on the matter from the Court of Justice nevertheless did not prevent us from analysing in detail what exists to date. The result is that the European copyright framework was, indeed, very much fragmented at the beginning. The initiatives from different Member States were laudable, but they could not, on their own, remove all uncertainty from national and cross-border market actors. Action from the EU was slow, but the product of many debates eventually materialised. It was heavily criticised, but undoubtedly needed. The balance of interests from many, many stakeholders crystallised in the DSM Directive, and opened the door for Member States to take example, and maybe go beyond some of its larger shortcomings. From the slow transposition process, arguably affected by worldwide issues, rose something which could only be described as lacking. Out of the five analysed countries, most stuck closely, maybe even too closely, to the Directive. Only Ireland and Germany managed to go a step further into the right direction (unlike Spain). But their doings might have been more motivated by the necessity to stick close to what they had already enacted in the past, than to actually correct some of the Directive's underlying issues.

Fragmentation was there, and fragmentation is still there, though in a lesser form. If unity is the actual goal, which is what the name « digital single market » implies, then it might very well never be archived as long as the European legislator continues to give leeway to Member States. It is true that copyright has traditionally been regulated through Directives, and Members of the EU have always had strong ties to their domestic copyright frameworks. But the current approach raises questions about whenever the EU's text and data mining legislation is truly fit for the digital revolution, where conventional borders will eventually disappear...

Bibliography

Legislation

Belgium

Code de droit économique, *M.B.*, 28th February 2013.

Projet de loi transposant la directive (UE) 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE, projet de loi, *Doc.*, Chambre, 2022-2023, n° 2608/009, 16th June 2022.

Estonia

Säilituseksemplari seadus, *R.T.*, 7th July 2017.

European Union

Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, *O.J.E.U.*, L 77, 27th March 1996.

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, *O.J.E.U.*, L 167, 22nd June 2001

Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs, *O.J.E.U.*, L 111, 5th May 2009.

Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on certain permitted uses of orphan works, *O.J.E.U.*, L 299, 27th October 2012.

Proposal for a Directive of the European Parliament and the Council on copyright in the Digital Single Market, COM(2016) 593 final, 14th September 2016.

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, *O.J.E.U.*, L 130, 17th May 2019.

France

Code de la Propriété intellectuelle, *J.O.R.F.*, 3rd July 1992.

Projet de loi pour une République numérique, projet de loi, *Dossier législatif*, Ass. Nat., 2015-2016, n° 3318.

Projet de loi pour une République numérique, avis présenté au nom de la Commission des affaires culturelles et de l'éducation, *Dossier législatif*, Ass. Nat., 2015-2016, n° 3389.

Projet de loi pour une République numérique, amendement, *Doc.*, Ass. Nat., 2015-2016, n° 180.

Projet de loi pour une République numérique, annales parlementaires, *Doc.*, Ass. Nat., 2015-2016, n° 107.

Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique, *J.O.R.F.*, 8th October 2016.

Loi n° 2020-1508 du 3 décembre 2020 portant diverses dispositions d'adaptation au droit de l'Union européenne en matière économique et financière, *J.O.R.F.*, 4th December 2020.

Ordonnance n° 2021-580 du 12 mai 2021 portant transposition du 6 de l'article 2 et des articles 17 à 23 de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE, *J.O.R.F.*, 13th May 2021.

Ordonnance n° 2021-1518 du 24 novembre 2021 complétant la transposition de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE, *J.O.R.F.*, 25th November 2021.

Décret n° 2022-928 du 23 juin 2022 portant modification du code de la propriété intellectuelle et complétant la transposition de la directive 2019/790 du Parlement européen et du Conseil du 17 avril 2019 sur le droit d'auteur et les droits voisins dans le marché unique numérique et modifiant les directives 96/9/CE et 2001/29/CE, *J.O.R.F.*, 24th June 2022.

Germany

Gesetz über Urheberrecht und verwandte Schutzrechte (Urheberrechtsgesetz), *BGBI*, 9th September 1965.

Entwurf eines Gesetzes zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft, Gesetzesentwurf der Bundesregierung, *Bundestag Drucksachen*, Bundestag, 2016-2017, n° 18/12329.

Gesetz zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft - Urheberrechts-Wissensgesellschafts-Gesetz – UrhWissG, *B.G.Bl.*, 7th September 2017.

Entwurf eines Gesetzes zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, Gesetzesentwurf der Bundesregierung, *Bundestag Drucksachen*, Bundestag, 2020-2021, n° 19/27426.

Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes, *BGBI*, 4th June 2021.

Greece

Ρύθμιση θεμάτων του Κρατικού Πιστοποιητικού Γλωσσομάθειας, της Εθνικής Βιβλιοθήκης της Ελλάδας και άλλες διατάξεις (Regulation on State Language Certificate subject matter, on the National Library of Greece and on other provisions), *ΦΕΚ (Official Journal)*, 15th February 2017.

Ireland

Copyright and Other Intellectual Property Law Provisions Act 2019, *Iris Oifigiúil*, 26th June 2019.

European Union (Copyright and Related Rights in the Digital Single Market) Regulations 2021, *Iris Oifigiúil*, 19th November 2021.

Netherlands

Auteurswet, *Staatsblad*, 23rd September 1912.

Wet op de naburige rechten, *Staatsblad*, 18th March 1993.

Datenbankenwet, *Staatsblad*, 20th July 1999.

Voorstel van de implementatiewet richtlijn auteursrecht in de digitale eengemaakte markt, memorie van toelichting, *Kamerstukken II*, Tweede Kamer, 2019-2020, n°35454.

Implementatiewet richtlijn auteursrecht in de digitale eengemaakte markt, *Staatsblad*, 29th December 2020.

Besluit van 18 december 2020 tot vaststelling van het tijdstip van inwerkingtreding van de Implementatiewet richtlijn auteursrecht in de digitale eengemaakte markt, *Staatsblad*, 29th December 2020.

Spain

Real Decreto-ley 24/2021, de 2 de noviembre, de transposición de directivas de la Unión Europea en las materias de bonos garantizados, distribución transfronteriza de organismos de inversión colectiva, datos abiertos y reutilización de la información del sector público, ejercicio de derechos de autor y derechos afines aplicables a determinadas transmisiones en línea y a las retransmisiones de programas de radio y televisión, exenciones temporales a determinadas importaciones y suministros, de personas consumidoras y para la promoción de vehículos de transporte por carretera limpios y energéticamente eficientes, *Boletín Oficial del Estado*, 3rd November 2021.

United Kingdom

Copyright, Designs and Patents Act 1988, c. 48, 15th November 1988.

Draft Copyright And Rights In Performances (Research, Education, Libraries And Archives) Regulations 2014, explanatory memorandum, HM Government, 2013-2014.

The Copyright and Rights in Performances (Research, Education, Libraries and Archives) Regulations 2014, 19th May 2014.

Official documents

European Union

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A Digital Single Market Strategy for Europe, COM(2015) 192 final, 6th May 2015.

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Towards a modern, more European copyright framework, COM(2015) 626 final, 9th December 2015.

Commission Staff Working Document. Impact Assessment on the modernisation of EU copyright rules. Part 1/3, SWD(2016) 301 final, 14th September 2016.

GEIGER C., FROSIO G. and BULAYENKO O., *The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects*, EUP Policy Department for Citizens' Rights and Constitutional Affairs, Brussels, 2018.

ROSATI E., *The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Technical Aspects*, EUP Policy Department for Citizens' Rights and Constitutional Affairs, Brussels, 2018.

J. SCHRANZ, « Questions and Answers on issues about the digital copyright directive », 27th March 2019, available at www.europarl.europa.eu, accessed on 22nd March 2022.

X, « What is the Digital Single Market about ? », *s.d.*, available at www.ec.europa.eu, accessed on 22nd March 2022.

X, « Copyright: Commission calls on Member States to comply with EU rules on copyright in the Digital Single Market », 26th July 2021, available at www.digital-strategy.ec.europa.eu, accessed on 23rd March 2022.

X, « Copyright: Commission urges Member States to fully transpose EU copyright rules into national law », 19th May 2022, available at www.ec.europa.eu, accessed on 20th July 2022.

France

BENSAMOUN A. and BOUQUEREL Y., *Rapport de mission. Transposition des exceptions de fouille de textes et de données : enjeux et propositions*, Paris, CSPLA, 2020.

X, « Pour une République numérique », *s.d.*, available at www.gouvernement.fr, accessed on 24th February 2022.

X, « Conseil supérieur de la propriété littéraire et artistique », *s.d.*, available at www.culture.gouv.fr, accessed on 22nd July 2022.

Germany

X, « Gesetz zur Angleichung des Urheberrechts an die aktuellen Erfordernisse der Wissensgesellschaft (UrhWissG) », 15th June 2021, available at www.bmj.de, accessed on 17th March 2022.

Evaluierungsbericht der Bundesregierung zu den durch das Urheberrechts-Wissensgesellschafts-Gesetz reformierten Vorschriften der §§ 60a bis 60h des Urheberrechtsgesetzes, *Bundestag Drucksachen*, Bundestag, 2022-2023, n° 20/1825.

United Kingdom

Answer to the question asked by J. STEVENS, House of Commons, 2020-2021, 16th January 2020, n°4371.

X, « Artificial intelligence and intellectual property: call for views », 7th September 2020, available at www.gov.uk, accessed on 15th July 2022.

X, « Government response to call for views on artificial intelligence and intellectual property », 23rd March 2021, available at www.gov.uk, accessed on 15th July 2022.

Case law

C.J.E.C., judgement *Infopaq International A/S v Danske Dagblades Forening*, 16th June 2009, C-5/08, EU:C:2009:465.

C.J., judgement (Grand Chamber) *Football Association Premier League Ltd and others v QC Leisure and others and Karen Murphy v Media Protection Services Ltd.*, 4th October 2011, joined cases C-403/08 and C-429/08, EU:C:2011:631.

C.J., judgement *Eva-Maria Painer v Standard VerlagsGmbH e.a.*, 1st December 2011, C-145/10, EU:C:2011:798.

C.J., judgement *Football Dataco Ltd and Others v Yahoo! UK Ltd and Others*, 1st March 2012, C-604/10, EU:C:2012:115.

C.J., judgement *Levola Hengelo BV v Smilde Foods BV*, 13th November 2018, C-310/17, EU:C:2018:899.

C.J., judgement *Stichting Brein v Jack Frederik Wullems*, 26th April 2017, C-527/15, EU:C:2017:300.

C.J., judgement *Republic of Poland v European Parliament and Council of the European Union*, 26th April 2022, C-401/19, EU:C:2022:297.

Literature

ALLEAUME C., « Les exceptions en faveur de l'enseignement et de la recherche scientifique (Commentaire des articles 3 à 5 de la directive) », *Communication Commerce Électronique*, 2019, n°10, 3, p. 1-5.

ANAWIS M., « Text Mining: The Next Data Frontier », available at www.rdworltonline.com, 6th January 2014, accessed on 1st February 2022.

ANDERSON B., « Using text mining to glean insights from COVID-19 literature », *Journal of Information Science*, 2021, n° 9, p. 1-9.

ANTON K., « UrhG §60c Wissenschaftliche Forschung », *Recht der elektronischen Medien*, SPINDLER G. and SCHUSTER F. (dir.), München, C. H. Beck, 2019.

ANTON K., « UrhG §60d Text und Data Mining », *Recht der elektronischen Medien*, SPINDLER G. and SCHUSTER F. (dir.), München, C. H. Beck, 2019.

ASHOK A., « Technological Protection Measures and Copyright Exceptions: The Rights of the Common Man in the Developed World », *Cochin University Law Review*, 15th June 2011, p. 100-135.

BERGMAN C., HUNTER L. and RZHETSKY A., « Announcing the PLOS Text Mining Collection », available at www.everyone.plos.org, 17th April 2013, accessed on 1st February 2022.

BERNE X., « Loi Numérique: l'exception de « text & data mining » en passe de rester lettre morte », 14th November 2017, available at www.nextimpact.com, accessed on 10th March 2022.

BINCTIN N. and PRÈS X., *Directives 2019/790 et 2019/789 sur le droit d'auteur dans le marché unique numérique*, Bruylant, Bruxelles, 2021.

BLEI D., « Surveying a suite of algorithms that offer a solution to managing large document archives », *IEEE Signal Process. Mag.*, April 2012, vol. 55, n° 4, p. 77-84.

BOND T., « UKIPO Consultation on AI and IP: What are the implications for UK copyright law? », 30th November 2021, available at www.twobirds.com, accessed on 15th July 2022.

BOTTIS M., PAPADOPOULOS M., ZAMPAKOLAS C. and GANATS P., « Text and Data Mining in the EU 'Acquis Communautaire' Tinkering with TDM & Digital Legal Deposit », *Erasmus Law Review*, November 2019, n° 2, p. 190-208.

BOURGEOIS M. and COUSIN G., « Les apports de la loi pour une République numérique en matière de propriété intellectuelle », *La Semaine Juridique Entreprise et Affaires*, n° 49, 8th December 2016, p. 32-37.

- BRADY C., KIRTON D., SUGRUE J. and CULLEN S., « Copy That! Ireland Implements EU Copyright Directive », 17th December 2021, available at www.williamfry.com, accessed on 28th July 2022.
- BULLINGER W., « UrhG §60d Text und Data Mining », *Praxiskommentar Urheberrecht*, WANDTKE A.-A. and BULLINGER W. (dir.), München, C. H. Beck, 2019.
- BULLINGER W., « UrhG §44b Text und Data Mining », *Praxiskommentar Urheberrecht*, WANDTKE A.-A. and BULLINGER W. (dir.), München, C. H. Beck, 2022.
- BULLINGER W., « UrhG §60d Text und Data Mining », *Praxiskommentar Urheberrecht*, WANDTKE A.-A. and BULLINGER W. (dir.), München, C. H. Beck, 2022.
- CHRISTENSEN K., « A European solution for Text and Data Mining in the development of creative Artificial Intelligence. With a specific focus on articles 3 and 4 of the Digital Single Market Directive », *Stockholm Intellectual Property Law Review*, 2021, vol. 4, n°2, p. 18-33.
- CHRISTOPHE C., « République numérique rime avec exceptions et limitations au droit d'auteur », *Revue Mensuelle Lexisnexis Jurisclasseur*, vol. 11, n° 89, November 2016, p. 1-4.
- DA CRUZ. F., « The IBM 704 », available at www.columbia.edu, February 2001, accessed on 11th July 2022.
- DE LA DURANTAYE K., « Neues Urheberrecht für Bildung und Wissenschaft – eine kritische Würdigung des Gesetzentwurfs », *Gewerblicher Rechtsschutz und Urheberrecht*, n°6, 2017, p. 558-567.
- DUCATO R. and STROWEL A., « Ensuring Text and Data Mining: Remaining Issues with EU Copyright Exceptions and Possible Ways Out », *E.I.P.R.*, 2021, vol 43, n° 5, p. 322-337.
- DUCATO R. and STROWEL A., *Limitations to Text and Data Mining and Consumer Empowerment. Making the Case for a Right to “Machine Legibility”*, CRIDES Working Paper Series, Louvain-La-Neuve, 2018
- DUSOLLIER S., « Exceptions and Technological Measures in the European Copyright Directive of 2001 - An Empty Promise », *International Review of Intellectual Property and Competition Law*, 2003, vol. 34, n°1, p. 62-75.
- DYAS-CORREIA S. and ALEXOPOULOS M., « Text and Data Mining: Searching for Buried Treasures », *Serials Review*, 2014, vol. 40, n° 3, p. 210-216.

ETHERINGTON D., « Microsoft invests \$1 billion in OpenAI in new multiyear partnership », 22nd July 2019, available under www.techcrunch.com, accessed on 16th July 2022.

GAMBHIR M. et GUPTA V., « Recent automatic text summarisation techniques: a survey », *Artificial Intelligence Review*, 2017, vol. 47, p. 1-66.

GEIGER C., FROSIO G. and BULAYENKO O., « Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU », in C. GARCIA and R. LLORCA (dir.), *Propiedad Intelectual Y Mercado Único Digital Europeo*, Tirant lo blanch, Valencia, 2019, p. 24-56.

GRISHMAN R. and SUNDHEIM B., « Message Understanding Conference – 6: A Brief History », *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, 1996, vol. 1, p. 466-473.

HEESEN H. and JÜNGELS L., « Der Regierungsentwurf der Text und Data Mining-Schranken (§§ 44b, 60d UrhG-E) », *Recht und Zugänge (RZ)*, 2021, n°1, p. 45-55.

HUGENHOLTZ B., « The New Copyright Directive: Text and Data Mining (Articles 3 and 4) », 24th July 2019, available at www.copyrightblog.kluweriplaw.com, accessed on 24th April 2021.

JOHNSON R., FERNHOLZ O. and FOSCI M., *Text and data mining in higher education and public research*, Association des Directeurs & personnels de direction des Bibliothèques Universitaires et de la Documentation, Paris, 2016.

KLEINKOPF F. and PFLÜGER T., « Digitale Bildung, Wissenschaft und Kultur – Welcher urheberrechtliche Reformbedarf verbleibt nach Umsetzung der DSM-RL durch das Gesetz zum Urheberrecht im digitalen Binnenmarkt? », *Zeitschrift für Urheber- und Medienrecht (ZUM)*, 2021, n° 8/9, p. 643-655.

KLEINKOPF F., JACKE J. and GÄRTNER M., « Text- und Data-Mining. Urheberrechtliche Grenzen der Nachnutzung wissenschaftlicher Korpora bei computergestützten Verfahren und digitalen Ressourcen », *Multimedia und Recht (MMR)*, 2021, n°3, p. 196-200.

KRETSCHMER M. and MARGONI T., « Data mining: why the EU's proposed copyright measures get it wrong », 24th May 2018, available at www.theconversation.com, accessed on 16th July 2022.

LUCAS A., « La transposition en France de la directive droit d'auteur dans le marché unique numérique 2019/790 », *A&M*, 2021, n°3, p. 307-324.

- LUHN H. P., « The Automatic Creation of Literature Abstracts », *IBM Journal of Research and Development*, April 1958, vol. 2, n° 2, p. 159-165.
- MARGONI T. and KRETSCHMER M., « The Text and Data Mining exception in the Proposal for a Directive on Copyright in the Digital Single Market: Why it is not what EU copyright law needs », 25th April 2018, available at www.create.ac.uk, accessed on 6th April 2022.
- MATAS A., « Copyright directive series: progress in the directive transposition », 21st April 2022, available at www.pro.europeana.eu, accessed on 28th July 2022.
- MCDONALD D. and KELLY U., « The Value and Benefits of Text Mining », 14th March 2012, available at www.jisc.ac.uk, accessed on 2nd February 2022.
- MINER G., DELEN D., ELDER J., FAST A., HILL T. and NISBET R. A., *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*, Oxford, Elsevier, 2012.
- MURRAY-RUST P., « The Right to Read Is the Right to Mine », 1st June 2012, available at www.blog.okfn.org, accessed on 16th July 2022.
- NAGAR A. and HAHLER M., « Using Text and Data Mining Techniques to extract Stock Market sentiment from Live News Streams », *International Proceedings of Computer Science and Information Technology*, 2012, vol. 47, p. 91-95.
- NORDEMANN J. B., PUKAS J. and WAIBLINGER J., « Transposition of the DSM Copyright Directive 2019/790 in Germany », *A&M*, 2021, n°3, p. 325-337.
- PACHALI D., « Urheberrecht für Wissenschaft und Bildung: Bundestag beschließt Reformen », 30th June 2017, available at www.irights.info, accessed on 17th March 2022.
- PO L. and ROLLO F., « Building an Urban Theft Map by Analyzing Newspaper Crime Reports », *2018 13th International Workshop on Semantic and Social Media Adaptation and Personalization*, 1st September 2018, p. 13-18.
- RAUE B., « Die Freistellung von Datenanalysen durch die neuen Text und Data Mining-Schranken (§§ 44b, 60d UrhG) », *Zeitschrift für Urheber- und Medienrecht (ZUM)*, 2021, n° 10, p. 793-802.
- RAUE B., « Text und Data Mining. Die neue Urheberrechtsschranke des § 60d UrhG », *Computer und Recht (CR)*, n°10, 2017, p. 656-661.
- ROSATI E., « An EU text and data mining exception for the few: would it make sense? », *Journal of Intellectual Property Law & Practice*, 2018, vol. 13, n°6, p. 429-430.

ROSATI E., « From the early days of harmonization to the DSM Copyright Directive 2019/790: continuity and complexity within the EU copyright framework », *A&M*, 2021, n°3, p. 289-298.

ROSATI E., « Irish Government proposes introduction of Irish text and data mining exception », 14th March 2018, available at www.ipkitten.blogspot.com, accessed on 28th July 2022.

ROSATI E., « The DSM Directive Two Years On: Do Things Ever Get Easier? », *IIC International Review of Industrial Property and Copyright Law*, 2021, vol. 52, p. 1139-1142.

SALDAÑA M., ESCOBAR C., GALVEZ E., TORRES D. and TORO N., « Mapping of the Perception of Theft Crimes from Analysis of Newspaper Articles Online », *2020 15th Iberian Conference on Information Systems and Technologies*, 15th July 2020, p. 1-7.

SCHACK H., « Das neue UrhWissG – Schranken für Unterricht, Wissenschaft und Institutionen », *Zeitschrift für Urheber- und Medienrecht (ZUM)*, n° 11, 2017, p. 802-808.

STIEPER M., « Die Schranken des Urheberrechts im Gesetz zur Anpassung des Urheberrechts an die Erfordernisse des digitalen Binnenmarktes », *Zeitschrift für Urheber- und Medienrecht (ZUM)*, 2021, n°10, p. 776-785.

STROWEL A., « Reconstructing the Reproduction and the Communication to the Public Rights: How to Align Copyright with Its Fundamentals », in B. HUGENHOLTZ (dir.), *Copyright Reconstructed, Rethinking Copyright's Economic Rights in a Time of Highly Dynamic Technological and Economic Change*, Alphen aan den Rijn, Wolters Kluwer, 2018, p. 203-240.

TAYLOR C., « Structured vs. Unstructured Data », available at www.datamotion.com, 21st May 2021, accessed on 11th July 2022.

TAYLOR D., « Data Mining Tutorial: What is Data Mining? Techniques, Process », available at www.guru99.com, 4th June 2022, accessed on 11th July 2022.

TRAPOVA A. and GERVASSIS N., « UKIPO's public consultation on AI and IP – text and data mining (Part 2) », 16th March 2022, available at www.copyrightblog.kluweriplaw.com, accessed on 15th July 2022.

TRIAILLE J-P., DE MEEÛS D'ARGENTEUIL J. and DE FRANQUEEN A., *Study on the legal framework of text and data mining*, EU Commission Publications Office, Brussels, 2014.

- VAN GOMPEL S., « Omzetting van de DSM Auteursrecht Richtlijn 2019/790 in Nederland », *A&M*, 2021, n°3, p. 338-356
- VOLLMER T., « The Limits of Copyright: Text and Data Mining », 21st January 2015, available at www.creativecommons.org, accessed on 16th July 2022.
- WANG O., « Décret complétant la transposition de la directive DAMUN : la gestion collective étendue précisée », 5th July 2022, available at www.dalloz-actualite.fr, accessed on 22nd July 2022.
- X, « Copyright in the Digital Single Market Directive – Implementation. An EU Copyright Reform Resource », *s.d.*, available at www.create.ac.uk, accessed on 6th August 2022.
- X, « Structured vs. Unstructured Data: A Complete Guide », available at www.talend.com, *s.d.*, accessed on 11th July 2022.
- X, « The European Parliament must improve the Text and Data Mining (TDM) exception to benefit European research and innovation », 26th March 2018, available at www.eare.eu, accessed on 24th April 2021.
- X, « COVID-19 Open Research Dataset Challenge (CORD-19) », available at www.kaggle.com, 12th March 2020, accessed on 9th February 2022.
- X, « Felix Reda », *s.d.*, available at www.greens-efa.eu, accessed on 10th March 2022.
- X, « Text & Data Mining: What is TDM? », available at www.libguides.cam.ac.uk, *s.d.*, accessed on 1st February 2022.
- X, « Text mining vs. data mining », available at www.expert.ai, 8th November 2016, accessed on 1st February 2022.
- X, « Text Mining », available at www.ibm.com, 16th November 2020, accessed on 1st February 2022.
- X, « What is R? », available at www.r-project.org, *s.d.*, accessed on 14th February 2022.
- XALABARDER R., « Transposition of the DSM Copyright Directive 2019/790 in Spain », *A&M*, 2021, n°3, p. 357-373
- ZANINI N. and DHAWAN V., « Text Mining: An introduction to theory and some applications », *Research Matters*, 2015, vol. 19, p. 38-44.
- ZONG C., XIA R. and ZHANG J., *Text Data Mining*, Pekin, Tsinghua University Press, 2021.

Annexes

Annex I: Interview conducted with Felix Reda on the 9th of March 2022

Stephan: Okay, I started the recording, awesome. So, yeah, hi, this will all be GDPR compliant, as you know, as it says, everything will be deleted afterwards if it's not needed anymore, and if you would like to access your recording or the transcript afterwards, you can just ask me and I'm going to send it to you, to make sure I didn't tamper with it or something like that.

Felix: Okay.

S: Cool, so, uhm... Let's start with the first question directly: How would you define text and data mining, and what implications do you think it has on the research community?

F: So, I would say text and data mining is any sort of computational analysis of either text and data for the purpose of extracting information from the text or data. And in terms of the impact on research in general, it's just a research methodology that is used for all kinds of questions and applications. So, everything from having a search function on a website is a form of text and data mining, to trying to answer a particular research question with quantitative methods, you would have to do text and data mining to do so. So, it's actually much, much broader than perhaps what was discussed in the popular concepts.

S: Alright, you just gave me some examples as well, so that's good. So, we can I mean, if you would like to add some examples to that, we can do that.

F: Well, I would say for example, in the research context, if you are trying to find all the publications on the particular subject, you might define some search queries and so that would just be the searching process, but it could also be something like, in computational linguistics, if you want to find out which context, in particular expression is used, for example, you could use text and data mining to analyse the structure of the language in different contexts.

S: OK, awesome. Thank you. So, let's go bit more concretely about the directive itself, so. Of course, there was already a long time before the directive of the European Union was adopted. So, do you know, maybe, if there were Member States who already regulated text and data mining directly before the actual directive was adopted?

F: Yeah. So, the first EU Member State that had a text and data mining exception was the UK and if I remember correctly, Germany also adopted its first text and data mining exception before the directive was adopted. And at the time, I think, it was mainly UK researchers also

who had raised the problem that text and data mining was being prevented by scientific publishers. That was sort of the reason why the legislature became active in the first place. So, the UK government tried to fix the issue through national copyright law. But they had the problem that the InfoSoc directive only allows Member States to make some exceptions in the research context for non-commercial purposes, and so both the UK and then later the German exception were limited to non-commercial users and that was one of the reasons why there was a demand for the EU to introduce a mandatory text and data mining exception

S: Yeah. So, that was also its objective then, and adopting these articles to answer this legal uncertainty.

F: Yes, but to be honest, I don't think it was so much a legal uncertainty. It was mostly a factual problem. You know, nobody was wondering, oh, I'm doing text and data mining, I wonder if this is legal or not, but rather people were trying to do text and data mining and academic publishers weren't preventing people from doing so, for example by blocking access to user accounts like if you try to download more research articles than a certain number, or if you try to use a script to extract information from research articles, your account would get blocked. So, it wasn't really that people were wondering, you know, I'm uncertain if this is allowed or not, it's just that people weren't able to do it and that's why there was a demand for the legislator to come in and sort of answer this question.

S: OK, OK. And now, what was your specific role in the adoption process of Directive 2019/790? Is there something that you specifically can say: it's because of me that this changed, or that I had an influence specifically on that part?

F: Yes, I mean, it's always difficult to say what would have happened if I had not been there. But probably, article 4 would not be part of the directive if I had not participated in the negotiations. Because the original Commission proposal only included article 3 on the text and data mining process for scientific research, and basically there was no problem with text and data mining in the economic sphere. Because if you wanted to use anything other than the research articles of scientific publishers, for text and data mining, there was no problem, because like the only reason why this text and data mining exception was proposed in the first place, was because academic publishers were technologically preventing researchers from doing text and data mining. But up until this point, text and data mining was not considered a copyright-relevant act. So normally, extracting factual information from data, whether or not this data is a copyright protected work or not, is not a copyright-relevant act.

So uhm, by using this method of fixing the problem, which to European Commission shows, which was to say, OK, we say text and data mining can be a copyright relevant factor. And now we create an exception through it. This caused problems for everybody else, basically, especially for economic operators. And so very early on in the discussion, I was sort of criticising this approach, I was saying we should rather clarify that the act of reproduction does not apply to these kinds of mechanical extractions, because there is no enjoyment of the work. And eventually there was an understanding also on the side of the Commission, but especially on the side of the Council, and to some extent from the rapporteur in the Parliament, Mr. Voss. Actually, something needs to be done for commercial text and data mining, so I don't know if this would have happened anyway, but it was definitely something I pushed for. If you are going to take this approach of an exception, then at least you have to make sure that it's also clarified that commercial text and data mining is allowed.

And then, of course, you know many small things like the specific wording of article 3, for example, because I was worried that the exception for research text and data mining wouldn't change anything. If you just say it's allowed, then you have got rid of the legal uncertainty, but then you still have the problem that the academic publishers might technologically prevent you from doing TDM.

So I'm not very happy with the wording of Article 3, paragraph 3, but at least it says that, you know, any measures that the rights holders, which in this case are usually academic publishers, the measures they take should not go beyond what is necessary to protect their systems, and I think in the recitals it also says somewhere that it shouldn't prevent people from actually doing the text and data mining, which was the real problem to begin with.

S: OK then we can jump straight into the next question about this. You did a really critical analysis on these two articles, and so I just wanted to ask what the results of that analysis were.

F: So, my preferred approach was to limit the scope of the right of reproduction. So, to say, extracting information from copyright protected data is not a reproduction in the first place, because it doesn't involve an enjoyment of the work as a work, it's a purely technical process. That was my preferred approach. Most people didn't understand what I was talking about because they didn't really understand the difference between the scope of an exclusive right and the exception to an exclusive right. But that was my preference, and I realised very quickly that it was not going to happen. So, then I tried instead to improve the exception to make it as workable as possible.

But there is a little bit of, sort of, results from this discussion in the text. It's in one of the recitals where it talks about normalising datasets. Where you can still see, I think it's like recital 8, it

says something like 'not all text data mining processes even include copyright relevant acts' and some stuff like that and it says 'all facts and data are not protected by copyright' and so on, so that some of the language that came from that discussion. And I think the only person who really understood it was Therese Comodini. She was the rapporteur from the Parliament side in the beginning and then later she got elected to the Maltese Parliament and was replaced by Axel Voss. Yeah, but that was kind of my, my initial position of how it should be done, but yeah, it wasn't successful.

S: Was there also intensive lobbying against you?

F: Yes. But also, intensive lobbying for my position. It was very interesting because the text and data mining had strong industry involvement on both sides. So, in favour of basically legalising text and data mining or protecting the ability to have text and data mining, you had companies such as Microsoft, for example, because they have been doing this for many, many years and all of a sudden, it was called into question whether this required permission. And then on the other side, you had the academic publishers who are very strongly lobbying against any sort of obligation to allow text and data mining of their research articles. So it was, well, I would say it was controversial and there was strong lobbying on both sides, but nowhere near as much as other elements of the directive. So, it was still considered kind of a specialist and technical issue and a lot of the discussion took place between advisors and it was not as much as the upload filters and so on.

S: OK. OK, thank you. So, if you had to list the positive elements and the negative elements of both articles, what would you say? Like, that's something that's good. And on the other hand, that's rather bad.

F: So, I would say what is good is that there are at least two exceptions, so that whether or not we're talking about research organisations, generally speaking, the directive allows text and data mining. So, I think that's good.

I think what's also good, is that it said that it should be without remuneration, because especially in my country, in Germany, whenever a new copyright exception is adopted, it's always with remuneration and in this case, it really doesn't make sense because the transaction costs would be astronomical. You know, if you're doing text and data mining, you probably want to combine information from thousands or even millions of sources, and so you can't really do fair compensation for that. Everybody would get one cent, or, you know, the costs of administering the system would be much higher. So, that I think is better, pretty good. There was also intense lobbying against that.

What's also, well what's perhaps bad is first of all, that there is this sort of lack of clarity about which uses are copyright-relevant in the first place. Because we are talking about a purely technical process and I'm still not convinced that copyright should regulate these questions in the first place.

What's also perhaps a little bit problematic from a dogmatic perspective is that you now have an exception with an opt out. So, I mean, it's not necessarily bad, it was a compromise, but it's sort of a strange beast from a dogmatic perspective. You have an exception which is supposed to create some kind of user right, and then you say this exception doesn't apply if the rights holder doesn't want it to apply, which kinda defeats the purpose of having an exception. But at least it says that this opt out has to be machine-readable. I think that's kind of what saves it, because that at least this means that you have to actively think about the fact I don't want text and data mining to take place on this and you have to put it on the website, in the form that a machine can understand. So, I think that's quite good. But yeah, it's a little bit of a strange construction to make an exception with an opt out.

S: Yeah. Yeah, on that part: it does mean that if you don't do anything well, it's allowed, and you must keep it in mind and do some actual work if you want it to be forbidden. It's not just a small label, it's implementing all this system that for it to be machine-readable. So, it's actually pretty good. Some people might say I don't want it, but I don't want to go all that way in implementing it.

F: So that yeah, I mean, I think that's the only way it can work because there is no person who is looking at the individual pieces of data and checking whether somebody has put an opt out somewhere. Because it, I mean, that's the whole point that it's a machine that is scraping this information. And so, if you don't have a standard for how to do the opt out, then you have huge transaction costs, and it would defeat the purpose of having an exception in the first place. But yeah, it has created some interesting new questions for lawyers, and I see I still get asked about this sometimes, like how it works and what is the legal nature of this opt out and so on.

S: OK, so if you could change these articles in the end, what would you change?

F: I would probably make it a lot simpler. I would just have one text and data mining exception that would basically say for like under copyright law text and data mining is allowed. Of course, it could still be forbidden for other reasons, like data protection or you know whatever reason there may be, but at least it's not a copyright infringement to automatically extract information from copyright protected works. Full stop. It would make it a lot simpler. I don't think it necessarily needs to have these two different articles for research and non-research. I don't think

you need to have the ability for right holders to, you know, apply security measures and all that stuff. I don't think you need to have the stakeholder dialogues in article 3, paragraph 4, because, like, let's face it, the researchers and the academic publishers are not going to sit around the table and agree amicably on some solution. So, I would just make it a lot simpler and just say that some data mining is not a copyright infringement.

S: OK, I see. They do think that might change in the future or do you think it's going to stay like this for some time?

F: I think it's probably here to stay for quite a while. There's going to be a lot of discussion around what exactly it means, so I expect there to be some litigation around the details of it and at the end of the day, probably the European Court of Justice will lay down some rulings to interpret these provisions. And usually, the European Court of Justice is relatively good at making such provisions workable. But of course, it would have been better if the legislature had done that work. But yeah, I don't think they will change very soon.

S: OK. What's going on currently is, as you know, Member States that transporting this directive. Do you do think that Member States could counter some of that criticism during their transposition process? That they might have the ability to change wordings, to flip stuff around so it's clearer, so it's, you know?

F: I think they're going to have a hard time making the provisions better than they already are. I think there is a lot of room to make it worse. So, I've seen lobbying at national level to introduce compensation even though the recital says that there should be no compensation. I know in Germany there is a problem with the implementation of Article 3, which says that individual researchers do not have the possibility to store the research copies, but only research organisations which I think is plainly in violation of the directive. I haven't seen anybody doing a better job than the directive does.

I think what could be done, what would maybe improve it would be to say like to reverse the burden of proof on the security measures by rights holders. So basically, to say rights holders, are not allowed to apply measures to restrict text and data mining unless they are specifically necessary to ensure security and integrity of the network or something like that. I think that would be helpful, but mostly I would say, they must do what the directive says, so I don't think they can fix the fundamental problems with it.

S: And staying in Germany, were there any issues with transposing that directive and knowing that there was already an existing exception for text and data mining before?

F: No. Uh, so basically the German government was in favour of text and data mining to begin with and before the directive, they just did what they could do under the DSM directive, except maybe for the compensation. Like with compensation, it's a German specialty. But they always give compensation and everything. So, in that case, I think it was helpful to have the directive and specify that there should be no compensation on this.

There is a problem with the German implementation because they say something like text and data mining, you're allowed to make reproductions and you're allowed to make the copies available to other people in your research team. But then when it comes to the compensation question, basically the structure of the German law is to say all research exceptions are subject to compensation. And then there is a list of exceptions through that rule, and it lists the reproductions for text and data mining. And now the German collecting societies are saying, haha, you forgot the 'making available for' text and data mining. So now we ask compensation for that. So yeah, it's more that kind of problem. But I don't think that switching to the new exception as such was met with a lot of opposition in Germany.

S: OK, all right. So as a conclusive question, do you think the European Union made a mistake in regulating it the way they did?

F: I mean, it's not perfect, but I feel like this is actually pretty good for copyright provision, because usually the European Union is always trying to find, you know, a balanced position. And so, if one side is saying there is a problem, namely the researchers, and the other side, the academic publishers, is saying there is no problem, we just don't want them to do text and data mining, usually they try to meet in the middle, which is a bad idea in this case. So, I feel like they have gone at least 80% of the way to actually solve the problem. It's not perfect, but it could have been a lot worse.

S: OK. Well, that's all I wanted to know then right there was interesting. So, would you like me to send you the transcription afterwards?

F: I mean, I probably won't have time to check it. I would be very interested to receive a copy of your thesis when it's finished. I'm always happy to see what's out there. And yeah, good luck with your research.

S: Thank you very much. And you know, have a great day and all that.

F: Thank you.

S: You too.

F: Thank you. Bye-bye.

Annex II: Interview conducted with Axel Voss on the 6th of April 2022

Stephan: Hello, nice to meet you. It's with great pleasure that I'm holding this interview with you today.

Axel: You're welcome, I hope I will be able to help you.

S: So, this interview will be recorded, I'll make a transcript, annex it to my thesis, etc.

A: The recording will be used for personal purposes?

S: Yes, for personal use only. I will delete the recording afterwards.

A: Yeah, okay.

S: Everything is GDPR compliant.

A: Yeah, alright.

S: So, I sent you the list of questions, I think you have received them.

A: Oh, no I don't have them.

S: It's fine, they are pretty straight forward, so it should work like this. Let's start with some more factual questions: how would you define text and data mining? Or do you have any form of person definition for it?

A: I do not think that we have some kind of definition. At least I wouldn't know this by heart. But I'm not sure if we have it in this copyright reform, whenever we put a definition for text and data mining...

S: There is one. But it was just to know whenever you had some personal one.

A: Yeah, yeah, yeah. No. But for me, text and data mining is when you have a kind of a specific, intentional goal, when you are then crawling through all the possible big data collections or whatever, and especially, as it was referred to in the copyright reform, on the internet and then that we're not infringing the copyright protected works. This is how I'm seeing text and data mining. Like if you have a lot of journals and you are, looking for evaluations for the devices you are producing or whatever. So that they are coming forward and finding out what might be the problem.

S: Thank you. Right. So, you already gave me ideas, some examples, that's great... *Clears throat* Sorry, I'm a bit little bit sick *laughs*.

A: Yes, it's the same for me *laughs*.

S: Do you know if before Directive 2019/790 came out, if they were already Member States who were legislating on the matter of text and data mining?

A: Yeah. I would say yes, but I can't give you an example. Because someone probably already had done something. Our intention with the European text and data mining issue has been that there is a need for text and data mining. But if you're doing this on the Internet with the complex system for copyright protected works, we found a balance between text and data mining and the copyright protected work and how we approach these, and this has been our intention to deal with it. It also makes more sense to have a European wide or full harmonised approach in the digital single market, not fragmented, not everywhere in a different way, yeah.

So, this has been our reason for regulating it, but I can't remember really. But I think there was already some regulations in place and in some Member States, but not on the level that we have now.

S: Okay, thank you for that. So, you've already talked about the objective as well, so that's great. Now, what was your specific role in the adoption of Directive 2019/790 and is there may be anything that you would say: "It's because I was there that this is now in the directive?"

A: *laughs* Well, I have been the rapporteur for the European Parliament for this copyright reform, where we had this text and data mining, the so-called value gap, the fair remuneration of creatives and so on. This has been part of it, but also video on demand and so there have been some issues. So, I have been the rapporteur for the Parliament, not only for my group, Then, here on this European level, as a rapporteur for the Parliament, you have somehow certain influence on this.

I insisted, at that time, that text and data mining should be possible, accessible, for everyone and so that copyright protected works should be indicated machine-readable and that this can happen without having the kind of legal uncertainty at the end in place, which had been a kind of hurdles so far. Especially the publishers had a lot of problems with it. They're proud of their archives regarding the day-by-day articles they are producing, and this is their product and I understand. But I told them: We can't because of your business model decrease the possibility of text and data mining.

So that's why I would still say that this approach was not, I would say, extremely influenced by me. I would have gone further, but then I was stopped by some political groups and other stakeholders. Cause every article in this copyright reform creates enemies everywhere and because you have to look for majorities, you have to be very careful how far you are going and

that's why I restrained myself. But still, I would say this is a good way forward for having text and data mining operable. So that's why I think this has been kind of an influence.

But you have been asking something that was not in it at the end? I guess hindered or whatever... This is what I can't remember really. But I think I'm trying to approach all these problems in a kind of a practical way, and that's why sometimes if you have ideological thinking parties with you, then sometimes you won't have a majority. But in this direction, I would say, it was quite good then that we have thought about what we can do and how we can do these and so on.

S: Okay, so you said you wanted to go a bit further, in the matters of text and data mining, what specifically did you have in mind?

A: As far as I know, I had to come back to a situation where text and data mining should or could be used only in a way where it's not that easy. Because of these copyright protected works and the publishers here in this case.

I would have had a more uncomplicated approach, so easier, simpler, that we're just referring to machine readable and that the publishers in this case, let me take this as a kind of an example, are indicating what is not allowed for text the data mining. And this was not accepted. Then it was a more, in a way, complicated solution. But we found a solution. I would still say we need to think in competition dimensions. And that's why we need a kind of very easy, simple approach on this that you can easily use text and data mining for specific purpose. And then, of course, we had to decide between private and research. The issues were that research normally benefits all of us and it's therefore easier to operate than if you're a private company and you're looking for something, then this might have different requirements in place.

S: So how satisfied are you with final result of these two articles on text and data mining, is there anything you would like to change or something?

A: I still would have the intention of simplifying this. And all in a way machine readable, that the machines can communicate and find out what I'm allowed to do and what I'm not allowed to do. And that there is a more active part also for the one who are the rightholders of copyright protected works. Normally the way is different: if you've once created a copyright protected work, then nobody can use it. But in this digital world you have to approach this a bit differently, you have to ask for more, active, more activity from the rightholders, and this is how I would have seen this. That we're saying: so, if you are offering something in the Internet, then also for your own interests you have to make sure what is accessible and what not. And also, if this is machine-readable for text and data mining, indicate this that this is allowed and this is not

allowed. I think this was not too much asking. But this is something what we could not implement in a simpler way. And that's why I would change this.

On the other hand, I'm also, of course, on the side of the press publishers. Because they are producing day by day something that they would like to sell. And this then, of course, shouldn't be used by someone who's not willing to pay. And therefore, I think you need this indication. And now we, as far as I remember, we solved this a little bit more bureaucratically, but probably kind of a slight differentiation, but this is what I would change. To simplify the whole process.

S: So, you said that there were publishers on one side. So, rather, against this text and data mining exception. Who were the other actors on the other side that were rather for it?

A: So, at first, I would say private companies have a big interest in text and data mining and, of course, the big tech is always interested in a lot of data and new knowledge out of this analysing of data. And of course, you have the press publishers, but all the copyright holders have been a little bit nervous because of these text and data mining articles. But I think we achieved somehow access but also safeguards for the rightholders.

So, you have on this one hand all these press publishers and all the others who are creating something and in a kind of a digital version and on the other side you have everyone who might be interested in analysing data. But at the end of the day, I would say it's always a good reason in saying we shouldn't hinder a new technology. But we should safeguard them on the other side of the rightholders.

S: All about finding a balance like it always is.

A: Yes, yes. But in these pragmatic situations in the European Parliament this is really not easy, especially if this it is getting so emotional at the end.

S: I see. So, do you think these two articles are going to stay like this or do you think they might be amended in some future?

A: No, I think they will stay like this because it was only two and a half years ago and now it should have entered into force last summer. So, for the next couple of years, it will stay and also the Member States might also have kind of a slightly different approach to this, because this has been only a directive, so they can add here something of their own interest. Probably there is not really kind of a need and changing these. But so far, I haven't heard any complaints.

S: What do you think Member States are going to do with these articles? Do you think they might go out of their way and try to, you know, modify them as far as they can, of course, within the limitations, or textually transpose them?

A: I did not pay too much attention to these transposing, especially these articles because others like 17 are of courses at different levels, but on these articles I'm not sure how they have transposed it. But I have heard a haven't heard anything negative about it though that's why I think they did it somehow, but probably they did it differently.

S: Alright, so, you think the EU has struck a good balance, you don't think that the articles are in the end too restrictive or something like that?

A: No, I wouldn't say so. Do you have some knowledge here? That this might be too restrictive?

S: Well, I did already have an interview with Julia[/Felix] Reda, so, of course, it's, well not the complete opposite, but it's really on the other side of this spectrum: we should do it differently, and it might be too restrictive, and things should be changed, etc.

A: Yeah, yeah, yeah, but as a pirate she's always really open about that and still not really liking copyright holders or that there is a kind of restriction on analysing. But at the end, this is how it is. This is property of everyone. Even this is intellectual property, but this is also a part of fundamental rights and that's why we have to find a balance. Yeah.

S: Yeah, I agree, I mean, it's not always so easy, you can't just go that one way.

A: No, totally not, especially here in the Parliament and where everything is so fragmented and then so on.

S: Alright, that was all I wanted to ask. If there is anything you'd like to add...

A: No, I don't think so. I still have the opinion that this is kind of a workable practical approach in a way. And I haven't heard anything negative so far, so that I can't see that there's a need to be more open or not. Yes, of course. Sometimes it can be easier, but if you have to balance this, you have to find compromises.

So no, I think we can be satisfied with it. It was not easy to convince the press publishers. But they have probably been happy with this now Article 15 about neighbouring rights. Here they almost opposed this whole issue, but I think we found a way forward.

S: Well, thank you very much for this interview. Thank you for taking your time. It was really interesting. Would you like to receive the transcript? Or maybe once my thesis is done, maybe, the thesis?

A: Yeah, your thesis might be interesting. [...]

S: Okay, I'm going to send you.

A: Yeah, okay. I wish you all the best, and I hope it helps.

S: Thank you, bye!

A: Bye!

Annex III: Member States-EU TDM provisions comparison table (2016-2019)

	France Art. 38	Germany §60d	UK Section 29A	CDSMD Art. 3	CDSMD Art. 4
Rights excepted	Reproduction	Reproduction, communication to the public, distribution	Reproduction	Reproduction, extraction	Reproduction, extraction
Targeted copyright framework	InfoSoc & Databases	InfoSoc & Databases	InfoSoc	InfoSoc, Databases & Press Publishers	InfoSoc, Databases, Press Publishers & Programs
Elements covered	Text and data in or associated with scientific writings	(Database) works	Works	Works or other subject matter	Works or other subject matter
Beneficiaries	Anyone	Anyone	Anyone	Research organisations & cultural heritage institutions	Anyone
Purpose	Scientific research	Scientific research	Research	Scientific research	Any
Commercial use	No	No	No	No	Yes
Lawful access	Yes	Yes/No (debated)	Yes	Yes	Yes
Storage/ Conservation	Only by bodies designated by decree (never adopted)	Only by institutions (§60e & §60f)	Allowed	Allowed (appropriate security & for scientific research)	Allowed (for TDM)
Additional conditions	No	No	Sufficient acknowledgment	No	No reservation by the rightholder
Security	? (Decree never adopted)	No	No	Yes	No
Sharing	? (Decree never adopted)	Yes (limited)	No	No	No
Waivable by contract	? (Decree never adopted)	No	No	No	Yes

Annex IV: Member States-EU TDM provisions comparison table (2019-2022): CDSM art. 3

	CDSMD Art. 3	France Art. L122-5-3, II CPI	Germany §60d UrhG	Netherlands Art. 15n Aw	Spain Art. 67 (+§4) RDL	Ireland Section 53A CRRA
Transposition type	/	Centralised (with references) ³⁴³	Centralised (with references)	Decentralised (with references) ³⁴⁴	Centralised (without references) ³⁴⁵	Decentralised (without references) ³⁴⁶
Rights excepted	Reproduction, Extraction	Reproduction, Extraction	Reproduction, Extraction, Communication to the public	Reproduction, Extraction	Reproduction, Extraction	Reproduction, Extraction
Targeted copyright framework	InfoSoc, Databases, Press Publishers	InfoSoc, Databases, Press Publishers	InfoSoc, Databases, Press Publishers	InfoSoc, Databases	InfoSoc, Databases, Computer programs, Press Publishers	InfoSoc, Databases, Press Publishers
Elements covered	Works or other subject matter	Digital works	Digital or digitalised works	Works of literature, science or art	Works and other legitimately accessible services	Works (or other subject matter)
Beneficiaries	Research organisations and cultural heritage institutions	Research organisations and the list of cultural heritage institutions	Individual researchers, Research organisations and cultural heritage institutions	Research organisations and cultural heritage institutions	Everyone, Research organisations and cultural heritage institutions (§4) ³⁴⁷	Everyone
Purpose	Scientific research	Scientific research	Scientific Research	Scientific Research	TDM/scientific research (§4)	Research
Commercial use	No	No	No	No	No (§4)	No
Lawful access	Yes	Yes (not potential but effective)	Yes	Yes	Yes	Yes
Storage/Conservation	For the purpose of scientific research, including verification	For the sole purpose of scientific research, including verification	As long as necessary for the purpose or	For purposes relating to scientific research	As long as necessary for the purpose of	For the purpose of research, including

³⁴³ One or multiple substantial, central provisions in the general copyright framework to which other provisions simply refer for their application in a different context.

³⁴⁴ One or multiple substantial provisions in the general copyright framework, but also other substantial provisions in different context. On top of being their own standalone provisions, they refer back to the general copyright framework for further details.

³⁴⁵ One single, substantial article, covering every context.

³⁴⁶ Multiple substantial provisions scattered all over the whole copyright framework, without references to other provisions.

³⁴⁷ Paragraph 4 is applicable to research organisations and cultural heritage institutions. It concerns storage, security, network integrity and security, and codes of conduct.

	CDSMD Art. 3	France Art. L122-5-3, II CPI	Germany §60d UrhG	Netherlands Art. 15n Aw	Spain Art. 67 (+§4) RDL	Ireland Section 53A CRRA
			review of scientific research/knowledge		TDM (§2) + for verification (§4)	verification of research results
Rightholder opposition	No	No	No	No	Yes	No
Security	Appropriate level	Appropriate level	Appropriate safeguards	Appropriate level	Appropriate level (§4)	Appropriate level
Network Security & Integrity	Necessary measures	Proportionate and necessary measures	Necessary measures	Necessary measures	Necessary measures (§4)	Only those with a lawful access ³⁴⁸
Waivable by contract	No	No (probably, but not explicit)	No	No	Yes	No
Additional conditions	No	No	No	No	No	Sufficient acknowledgement
Additional obligation of information	No	Yes	No	No	No	Yes

³⁴⁸ Can be archived through IP address validation or user authentication.

Annex V: Member States-EU TDM provisions comparison table (2019-2022): CDSM art. 4

	CDSMD art. 4	France L122-5-3, III CPI	Germany §44b UrhG	Netherlands 15a Aw	Spain 67 RDL	Ireland 53B CRRA
Transposition type	/	Centralised (with references)	Centralised (with references)	Decentralised (with references)	Centralised (without references)	Decentralised (without references)
Rights excepted	Reproduction, Extraction	Reproduction, Extraction	Reproduction, Extraction	Reproduction, Extraction	Reproduction, Extraction	Reproduction, Extraction
Targeted copyright framework	InfoSoc, Databases, Computer Programs, Press Publishers	InfoSoc, Databases, Computer Programs, Press Publishers	InfoSoc, Databases, Computer Programs, Press Publishers	InfoSoc, Databases, Computer Programs, Press Publishers	InfoSoc, Databases, Computer Programs, Press Publishers	InfoSoc, Databases, Computer Programs, Press Publishers
Elements covered	Works or other subject matter	Digital works	Digital or digitalised works	Works of literature, science or art	Works and other legitimately accessible services	Works (or other subject matter)
Beneficiaries	Everyone	Everyone	Everyone	Everyone	Everyone	Everyone
Purpose	Any	Any	Any	Any	TDM	TDM
Commercial use	Yes	Yes	Yes	Yes	Yes	Yes
Lawful access	Yes	Yes (not potential but effective)	Yes	Yes	Yes	Yes
Storage/Conservation	As long as necessary for the purpose of TDM	Destruction at the end of the mining	Deleted when no longer necessary for TDM	As long as necessary for the purpose of TDM	As long as necessary for the purpose of TDM	As long as is necessary for the purpose of TDM
Security	None	Appropriate level	None	None	None	None
Opposition manner	Expressly and appropriately: for example machine-readable means	Expressly and appropriately: for example machine-readable means (not motivated)	For online-accessible works: only by machine-readable means. For others any.	Expressly and appropriately: for example machine-readable means	Expressly and appropriately: for example machine-readable means	For online-accessible works: only by machine-readable means. For others clearly communicated
Waivable by contract	Yes	Yes	Yes	Yes	Yes	No
Additional conditions	No	No	No	No	No	No
Additional obligation of information	No	Yes	No	No	No	No

