



FACULTY OF LAW  
Lund University

Henrik Österström

# Protection of Data in the Context of Big Data

LAGM01 Master Thesis

European Business Law  
30 higher education credits

Supervisor: Anna Tzanaki

Semester: Spring 2022

# Contents

<b>SUMMARY</b>	<b>1</b>
<b>SAMMANFATTNING</b>	<b>3</b>
<b>ABBREVIATIONS</b>	<b>5</b>
<b>1 INTRODUCTION</b>	<b>6</b>
1.1 Purpose and Research Questions	6
1.2 Methodology and Material	6
1.3 Delimitations	7
1.4 Outline	8
<b>2 ANALYSIS</b>	<b>9</b>
2.1 Background	9
2.2 What is Big Data?	11
2.3 Big Data Process and Data Types	13
2.4 Copyright and Related Rights Under the Infosoc Directive	18
2.4.1 <i>Introduction</i>	19
2.4.2 <i>Influencing International Law</i>	19
2.4.3 <i>Protected Subject Matter</i>	20
2.4.4 <i>Authorship</i>	23
2.4.5 <i>Rights</i>	24
2.4.6 <i>Limitations and Exceptions</i>	25
2.4.7 <i>A Second Layer of Protection: Technological Protection Measures</i>	29
2.4.8 <i>Assessment</i>	31
2.4.8.1 Human-Generated Data	31
2.4.8.2 Machine-Generated Data	37
2.5 Protection of Data under the Database Directive	38
2.5.1 <i>Scope of the Database Directive</i>	38
2.5.2 <i>Copyright Protection for Databases</i>	40
2.5.3 <i>The Sui Generis Database Right</i>	41
2.5.4 <i>Assessment</i>	45
2.6 Protection of Big Data under the Trade Secret Directive	50
2.6.1 <i>Assessment</i>	53
2.7 Contractual Protection	54
2.8 Concluding remarks	55
<b>BIBLIOGRAPHY</b>	<b>58</b>



# Summary

Within the framework of Big Data and the use of AI technologies we now possess the ability to automate complex tasks through collecting, sharing, and processing the collected data. Analysis of data can unlock enormous potential within most fields. Meanwhile, legal uncertainty in data ownership rights causes businesses to not fully participate in the data sharing market. In this context this thesis focuses on exploring how data during the big data process, from the collection of data to the analysis of the data, can benefit from protection from copyright as intellectual creations under the Infosoc Directive, copyright protection for databases and the *sui generis* right for databases under the Database Directive, as well as a trade secret under the Trade Secrets Directive.

Copyright protection under the Infosoc Directive is granted for expressions of intellectual creations. Ideas and facts are excluded from the scope of protection. Only natural persons are recognized as authors. In a big data context this will raise issues for protecting collected data since some data may be created on digital platforms and qualify as a work, but the protection would then be vested in the user, rather than the actor collecting data for big data analysis. Furthermore, machine-generated data from sensors readings would for the most part consists of facts that fall outside the scope of protection. The results of data analysis might also consist of facts, but in any case, would struggle to identify a human author in most situations. The most likelihood of protection data would in wrapping the result of data analysis in an expression qualifying as an intellectual creation.

As for copyright protection in databases, it protects the structure and arrangement of a database if it constitutes the author own intellectual creation. This would not protect the data held in the database and in most cases it would hard to copy the structure of a database by extracting the contents thereof. The *sui generis* database right protects databases that have required a

substantial investment. This offers a wider scope of applicability but might require some artificial constructions for a database to fall within the scope, due to excluding investments in creating the data, rather than obtaining the data that constitute the contents of the database.

Finally, regarding protecting data as trade secrets, the main hurdle is that a trade secret cannot be generally known or readily available, which in the case of data often would be the case if the data has been acquired from any publicly available source. For non-publicly available data the Trade Secrets Directive offers scope for protection as long as measures are undertaken to keep the data a secret.

# Sammanfattning

Inom ramen av Big Data och användning av AI-teknik så har vi nu möjlighet att automatisera komplexa uppgifter genom att samla in, dela och behandla inhämtade data. Dataanalys öppnar upp möjligheten att låsa upp enorm potential inom de flesta branscher. Samtidigt råder osäkerhet kring juridiska rättigheter kopplat till ägande av data vilket föranleder att företag väljer att inte delta fullt ut i marknaden för att data-delning. Utifrån denna kontext ämnar denna uppsats undersöka hur data under en Big Data process, från inhämtande av data till analys av data, kan erhålla upphovsrättsskydd under Infosoc-direktivet, upphovsrättsskydd skydd för databaser och som *sui generis* databas under Databasdirektivet, eller som företagshemligheter under företagshemlighetsdirektivet.

Upphovsrättsskydd under Infosoc-direktivet förutsätter ett verk som ger uttryck för upphovsmannens intellektuella skapelse. Idéer och fakta omfattas inte av upphovsrätt. Endast fysiska personer erkänns som upphovsmän. I en big data kontext kan dessa krav medföra svårigheter att skydda inhämtad data eftersom data exempelvis kan ha inhämtas från en digital plattform och utgöra ett upphovsrättsskyddat verk, men skyddet skulle i så fall vara tilldelat användaren snarare än aktören som inhämtar data i syften för dataanalys. Vidare så skulle maskingenererad data som inhämtats genom sensor till största del utgöra fakta och inte omfattas av upphovsrättsskydd. Även dataanalysen kan resultera i mer fakta, och oavsett, skulle en sådan analys endast med svårigheter kunna identifiera en mänsklig upphovsman. Däremot skulle data från dataanalysen kunna analyseras och presenteras i ett uttryck som skulle kunna ses som en intellektuell skapelse.

När det kommer till upphovsrättsskydd i databaser så avser sådant skydd strukturen och arrangemanget av databasen om det anses vara ett uttryck för upphovsmannens egen intellektuella skapelse. Skyddet omfattar således inte data som finns lagrad i databasen och i de flesta fall skulle inte nedladdningar

från en digital databas medföra att strukturen kopieras. *Sui generis* databasskyddet omfattar databaser som har krävt en substantiell investering. Således omfattas ett bredare fång av databaser, men det kan kräva vissa särskilda arrangemang för att falla inom ramen för sådana investeringar eftersom investeringar för att skapa data inte omfattas, utan endast investeringar för att erhålla data.

Slutligen, när det kommer till företagshemlighet, så är det största hindret som måste övervinnas att företagshemligheter inte får vara allmänt kända eller lättillgängliga för kretsen av personer som normalt hanterar sådan typ av information, vilket medför att data som funnits publikt tillgänglig inte kan erhålla skydd som företagshemlighet. Däremot för annan typ av data finns det möjligheter att erhålla skydd så länge åtgärder vidtas för att hålla sådan data hemlig.

# Abbreviations

AI	Artificial Intelligence
CJEU	Court of Justice of the European Union
Database Directive	Directive 96/9/EC on the legal protection of databases
EU	European Union
Infosoc	Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society
IoT	Internet of Things
TDM	Text and Data Mining
TPM	Technological Protection Measure
Trade Secrets Directive	Directive 2016/943 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure



# 1 Introduction

## 1.1 Purpose and Research Questions

The purpose of this thesis is to explore and assess the possibility of protecting data in a Big Data context under EU law. For this purpose, the following questions will be answered:

1. When, if, and how can data in a big data context be protected under the Infosoc Directive, the Database Directive, and the Trade Secrets Directive respectively?

## 1.2 Methodology and Material

I have chosen to employ the legal dogmatic methods for the writing of this thesis. The legal dogmatic method has been used for this essay. This method aims to explore the law through its various sources as to form a coherent system in which it should enable to reach an understanding of *de lege lata*.

This thesis will primarily work with three sources of law, the Infosoc Directive, the Database Directive, and the Trade Secrets Directive. These are sources of secondary law in the hierarchy of EU law, whereas the Treaties and the Charter are primary sources of EU law.

Case law from the CJEU constitute another important source of interpretation of legal provisions. Even though the case law is not formally binding precedents, in practice, the precedents set by the CJEU will be followed, and the CJEU will build upon previous precedents in later judgements. As such, the interpretation of EU law found in case law serves as an important source of interpretive authority.

To interpret EU law will be necessary to consider a combination of a literal interpretation of the wording found in the source of law, a contextual

interpretation which requires placing the provision in its context and taking consideration of its relationship with other provisions, and the purpose of the legislation in question.

When an EU legal act implements international conventions, it might also be useful to consult those conventions to gain an understanding of the meaning within the context of EU law.

Opinions found in the juridical literature can also provide a better understanding of law if built around convincing arguments on how a provision ought to be interpreted.

### **1.3 Delimitations**

The analysis of this thesis has been delimited to assess if data in a big data context can be protected as works under the Infosoc Directive or protected by copyright or the *sui generis* database right in the Database Directive as well as a trade secret under the Trade Secret Directive.

This means that other EU sources of law that might influence data but not offer protection for the actors utilizing data in the big data context have been excluded from the analysis. These other include *inter alia* competition law, EU general principles, the charter of fundamental rights, and sector specific legal acts. Another important delimitation lies in that the general data protection regulation (GDPR) has not been considered. The main reason for these delimitations is that these sources of law does not really serves as a ground of protecting data for the actor's analysis within a big data context, or in the case of competition law, mainly deal with the possibility of requiring sharing of data.

## 1.4 Outline

The thesis will start of with a bit of background information providing some context and inspirations for the research questions of this thesis.

After presenting the background, a description of big data will be provided, followed by an explanation of the big data process and on various data types. These should serve to provide a basic level of technical understanding and provide a necessary technical understanding for analysing the research questions.

Then an overview of the relevant provisions and how they should be interpreted of the Infosoc Directive, followed by an analysis. The same structure will then be applied to the Database Directive and the Trade Secrets Directive respectively.

A short reflection on contractual control of data, followed by some concluding remarks ends off the thesis.

# 2 Analysis

## 2.1 Background

Currently we are amidst what has been called the Fourth Industrial Revolution.<sup>1</sup> Just in the year of 2010 we created as much data as had been created since the dawn of civilization until 2003. Still, information keeps being created at an increasing pace.<sup>2</sup> Within the framework of Big Data and the use of AI technologies we now possess the ability to automate complex tasks through collecting, sharing, and processing the collected data.<sup>3</sup> The value of the data economy within the EU and UK exceeded 400 billion Euro in 2019 which amounted to an annual growth rate of 7,6 %. The value of the data economy is expected to keep increasing and reaching a value of between 550 – 827 billion Euro by 2025.<sup>4</sup>

The Commission, having recognized both the exceptional growth of big data technology and services as well as the enormous potential this technology can unlock in most fields, held in its communication *Towards a thriving data economy* in 2013 that Europe cannot afford miss out on such an opportunity. The Commission recognized *inter alia* the need of making sure that the regulatory framework should lead to legal certainty for businesses as well as creating consumer trust in data technologies.<sup>5</sup> In 2016 the General Data Protection Regulation (GDPR) was adopted, and was to be transposed until 25 May 2018, to address the concerns regarding consumer trust in data

---

<sup>1</sup> Yann Ménière and others, 'Patents and the Fourth Industrial Revolution: The Global Technology Trends Enabling the Data-Driven Economy' (EPO 2020) <[www.researchgate.net/publication/353637878\\_Patents\\_and\\_the\\_Fourth\\_Industrial\\_Revolution\\_The\\_global\\_technology\\_trends\\_enabling\\_the\\_data-driven\\_economy](http://www.researchgate.net/publication/353637878_Patents_and_the_Fourth_Industrial_Revolution_The_global_technology_trends_enabling_the_data-driven_economy)> accessed 25 May 2022, 14.

<sup>2</sup> Cristophe Geiger, Giancarlo Frosio and Oleksandr Bulayenko, 'Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU' [2019] SSRN Electronic Journal, 4.

<sup>3</sup> Ménière and others (n 1) 14.

<sup>4</sup> Gabriella Cattaneo and others, 'The European Data Market Monitoring Tool: Key Facts & Figures, First Policy Conclusions, Data Landscape and Quantified Stories D2.19 Final Study Report' (2020 Publications Office of the European Union) 8–9.

<sup>5</sup> Commission, 'Towards a thriving data-driven economy' COM (2014) 442 final.

technology. Furthermore, the Commission found a need to remove national restrictions to the free flow of data which was emerging from national requirements to keep data within the national territory.<sup>6</sup> In 2018 a regulation on the free flow of non-personal data was adopted to create free movement of data within the Union by prohibiting unjustified data localization requirements.<sup>7</sup>

In 2015 the Commission followed up with a communication on *A Digital Single Market Strategy for Europe*. The data economy was recognized as a ‘as a catalyst for economic growth, innovation and digitisation across all economic sectors, particularly for SMEs (and start-ups) and for society as a whole.’<sup>8</sup> It was held that to benefit fully from the data economy legislative barriers such as *inter alia* fragmented copy right rules and a lack of clarity on rules governing the use of data needed to be removed.<sup>9</sup> To address these issues a directive was adopted in 2019, and to be transposed by 7 June 2021, laying out certain limitations to copyright and related in regards to *inter alia* data mining.<sup>10</sup>

In 2017 the Commission released as communication on *Building a European Data Economy* recognizing that data has become an ‘essential resource for economic growth, job creation and societal progress’<sup>11</sup>. While being an essential resource issues was identified relating to access, exchange and sharing of non-personal data. To resolve those issues a proposal on European data governance (Data Governance Act) has been laid out but which has yet to be adopted.<sup>12</sup>

---

<sup>6</sup> Commission (n 5) 14.

<sup>7</sup> See Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union.

<sup>8</sup> Commission, ‘A Digital Single Market Strategy for Europe’ COM (2015) 192 final 14.

<sup>9</sup> Commission (n 8) 14.

<sup>10</sup> See Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC.

<sup>11</sup> Commission, ‘Building a European Data Economy’ COM (2017) 9 final 2.

<sup>12</sup> Commission, ‘Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act)’ COM (2020) 767 final.

In summary, the focus on adaptations to the legal framework in regard to the data economy has been to improve the free movement of data between Member States, increasing consumer trust and enabling access to data. At the same time, a study has found that small to medium-sized enterprises (SME) captured almost all the revenue in the data supply market. A study has also shown that one of the main barriers for SMEs that don't participate in the data sharing market is related to the legal uncertainty surrounding 'ownership rights' over data.<sup>13</sup> Investigating the legal issues in regard to data ownership amounts to the backdrop for the purpose of this thesis.

## 2.2 What is Big Data?

While this study aims to investigate the legal framework surrounding ownership in data, it is necessary for a proper analysis to also examine the context in which data is being used as an asset. Generally, 'Big Data' is the buzzword being used to talk about the phenomenon of drawing valuable insights from the collection of data. While data previously also has been collected and used to draw conclusions the emergence of the internet and the use of digital technology seems to have necessitated the use of a new term to describe the collection and use of datasets within this context as compared to the previous traditional use.

It does not exist any generally accepted definition of Big Data but the oldest and most common way to describe Big Data is in reference to the three Vs: volume, velocity, and variety.<sup>14</sup> Volume refers to the immense scope of digital datasets. Velocity to the speed of which the data is being collected. Digital technology allows for a continuous, often real time, recording of data. Variety denotes that the big data can consist of a combination of structured, unstructured and semi-structured data. Structured data can be standardized

---

<sup>13</sup> Everis, 'Study on data sharing between companies in Europe' Final report 24-26.

<sup>14</sup> Andrea De Mauro, Marco Greco, and Michele Grimaldi 'A Formal Definition of Big Data Based on its Essential Features' (2016) Volume 65 Library Review 122.

and classified, for example numerically or alphabetically.<sup>15</sup> It can for example involve extracting some information from text, image, or video sources for the purpose of data analysis.<sup>16</sup> Unstructured data might lack the structural organization that is required for data analysis to take place and such data often consists of text, images, audio, or video.<sup>17</sup> Semi-structured data consists of unstructured data that has been provided with metadata to classify the unstructured data.<sup>18</sup> Such meta data can for example consists of a denotation of the type of data, and how it was recorded and measured.<sup>19</sup>

To the three Vs it has also been suggested to add that Big Data is exhaustive in nature and fine grained in scope. These suggested additions can be exemplified in that data can be collected and recorded in regard to any and all uses of a device; for example in mobile phones or the logging of all transactions and interactions across digital networks or homepages and measurements from sensors embedded in an object or the environment.<sup>20</sup>

Until recently it was not possible to properly analyze Big Data. The reasons for this relate *inter alia* to the abundance of data due to the volume and variety of created data, as well as due to the fact that much of the data is collected without any specific purpose or as a by-product of another activity. However, increased computational power combined with analytical techniques utilizing artificial intelligence and machine learning has enabled detecting patterns in Big Data.<sup>21</sup> As examples of how this technology is being utilized, a study found that shared data in a business-to-business setting was used in 44 % of cases to design innovative products and services, 31% to optimise the supply

---

<sup>15</sup> Annika Richterich, *The big data agenda: Data ethics and critical data studies* (University of Westminster Press 2018) 6–7.

<sup>16</sup> Divyakant Agrawal and others 'Challenges and Opportunities with Big Data A community white paper developed by leading researchers across the United States' (2012) 5.

<sup>17</sup> Amir Gandomi and Murtaza Haider, 'Beyond the Hype: Big Data Concepts, Methods, and Analytics' (2015) 35(2) *International Journal of Information Management* 137, 138; Richterich (n 15) 7.

<sup>18</sup> Richterich (n 15) 7.

<sup>19</sup> Agrawal and others (n 16) 4.

<sup>20</sup> Rob Kitchin, 'Big Data, New Epistemologies and Paradigm Shifts' (2014) 1 *Big Data & Society* 2.

<sup>21</sup> Rob Kitchin (n 20) 2.

chain, 29% for training algorithms for Artificial Intelligence, and 26% for predictive maintenance<sup>22</sup>.

Before the output or result of the data analysis can be properly relied upon is often necessary to provide an explanation on how the result was derived and upon which inputs. Furthermore, the result can need further adaptation to assist in the interpretation of the results, such as tabular presentations or visualizations of the result.<sup>23</sup>

By transforming the data into a structure so that it is ‘useful and relevant for a specific purpose’ the data can be said to be turned into information and a knowledge asset.<sup>24</sup>

## 2.3 Big Data Process and Data Types

The process described in section 2.2 can be categorized into five phases. The first three phases can be said to relate to data management and the last two with data analysis.<sup>25</sup>

First, when talking about ‘data’ it isn’t obvious exactly how to define the term, for example a discussion can be had on drawing a distinction between data and information.<sup>26</sup> Drawing such a distinction is however not fulfilling a purpose in the context of this thesis. In the proposal for the Data Act ‘data’ has been defined as ‘any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audio-visual recording’<sup>27</sup>. As such, this thesis will settle for the following and more generalized definition of data, where it simply

---

<sup>22</sup> Commission, ‘Public Consultation on Data Act and Amended Rules on the Legal Protection of Databases: Summary report on the public consultation’.

<sup>23</sup> Agrawal and others (n 16) 7.

<sup>24</sup> Mauro, Greco and Grimaldi (n 14) 125.

<sup>25</sup> Richterich (n 15) 140.

<sup>26</sup> Mauro, Greco and Grimaldi (n 24).

<sup>27</sup> Commission ‘Proposal for a regulation on harmonised rules on fair access to and use of data (Data Act)’ COM (2022) 68 final 38.



means any content obtained in the process and for the purpose of Big Data analysis.

The first phase, data acquisition and recording, relates to the fact that the data must be recorded from some data generating source, and secondly, that much of the generated data is redundant so that it needs to be filtered and compressed to only contain useful information before being stored.<sup>28</sup> Already existing data can of course be obtained, for example by downloading a file or database.<sup>29</sup> Data can for example be acquired directly from human interaction with a technology or by reading a result from a sensor, or obtained from data brokers, exchange or sale between companies and through mergers and acquisitions.<sup>30</sup>

The data generating sources can be classified into machine generated data and human generated data.<sup>31</sup>

Data can be machine generated in the sense that its creation does not require any active human intervention to be created. Such data can be created by ‘computer processes, applications or services, or by sensors processing information received from equipment, software or machinery, whether virtual or real’<sup>32</sup>. Often this type of data will be created as a by-product and stored in a database necessitated by the operation of the technology that is collecting the data.<sup>33</sup> However, data can also be collected for other purposes, such as improving the efficiency of the technology by analysing the data to, for example, cutting energy use.<sup>34</sup>

---

<sup>28</sup> Agrawal and others (n 16) 4.

<sup>29</sup> Jean-Paul Triaille, Jérôme de Meeûs d’Argenteuil, and Amélie de Francquen, ‘Study on the Legal Framework of Text and Data Mining (TDM)’ (2014) 28.

<sup>30</sup> Vassilis Hatzopoulos, *The Collaborative Economy and EU Law* (Bloomsbury Publishing 2018) 68.

<sup>31</sup> Nestor Duch-Brown, Bertin Martens and Frank Mueller-Langer, ‘The economics of ownership, access and trade in digital data; Digital Economy Working Paper 2017-01’ (2017) JRC Technical Reports 7.

<sup>32</sup> Commission (n 11) 9.

<sup>33</sup> Commission, ‘Impact assessment report’ SWD (2022) 34 final 19.

<sup>34</sup> Alexandra Giannopoulou, ‘Access and Reuse of Machine-Generated Data for Scientific Research’ (2019) 12 *Erasmus Law Review* 155.

Data can also be human generated in the sense that it requires active human intervention. This could be human interactions on online platforms, for example a like on Facebook, or a text entered into a search engine.; or more generally, a text, video or image created by a human and that then has been collected by a digital technology.<sup>35</sup>

Returning to the phases, the second phase, information extraction and cleaning, relates to when the acquired data has not been stored in a structured manner and information needs to be extracted from the collected data so as to be fit for analysis. The collected data can also contain errors which possibly can be corrected for so as to preserve the data set with only valid data.<sup>36</sup> Storing the data will generally, but not necessarily, entail the creation of a copy of the data being created. For example, data can be ‘crawled’ through where only a small part, such as a word or small sample of words, pixels, sounds, or data, are collected and stored.<sup>37</sup>

The third phase, data integration, aggregation, and representation involve storing the collected data in a database. This involves finding a suitable database design structure and semantically expressed data to be fit for automated computer analysis.<sup>38</sup> It can involve, for example, converting terms into a common denotation, converting human readable text into a machine-readable format, or converting varying file types into a common format.<sup>39</sup>

Phase four and five relates to the data analysis part of Big Data. Phase four is the query processing, data modelling and analysis. Query processing relates to the fact that the data analysis may involve mining many data sets stored in databases with varying structures and collecting it and storing all the data for analysis in a proper manner if possible or otherwise discarding it.

---

<sup>35</sup> Ekaterina Olshannikova and others, “Conceptualizing Big Social Data” (2017) 4 Journal of Big Data.

<sup>36</sup> Agrawal and others (n 16) 4-5.

<sup>37</sup> Triaille, d’Argenteuil, and Francquen (n 29) 31.

<sup>38</sup> Agrawal and others (n 16) 5.

<sup>39</sup> Triaille, d’Argenteuil, and Francquen (n 29) 28, 45-46.

The data analysis is performed by an artificial intelligence (AI). In the proposal for the Artificial Intelligence Act a definition of AI has been suggested as software that is developed with one or more of the techniques and approaches, such as machine learning, and ‘can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with’.<sup>40</sup>

Machine learning is the main technology applied in the field of AI today. Machine learning is based on feeding data as input to an algorithm and that the algorithm then provides an output in relation to the goal given to the algorithm. A central theme of machine-learning is that the algorithm must be trained and improved by processing the input data.<sup>41</sup>

The better data, in terms of quality and quantity, that the machine learning algorithm is provided, the better the performance of the output. Machine learning can be categorized into three groups depending on the nature of the training data and the learning process: supervised learning, unsupervised learning, and reinforcement learning.<sup>42</sup>

In supervised learning the algorithm is provided with training data that has been labelled and act as a guide or reference on if predictions made on unlabeled data is correct. For example, training data can be handwritten notes with the numbers 0 to 9 which has been labelled accordingly. The learning process can in a very simplified way be described as an iterative process where the algorithm adjusts its parameters in accordance with if it correctly predicted a number from the unlabeled data in reference to the labeled data that acts as a guide of if a prediction was true or false. The human influence over the output is influenced by the quality in the selection of training data,

---

<sup>40</sup> Commission, ‘Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts’ COM(2021) 206 final, 40.

<sup>41</sup> Anthony Man-Cho So, ‘Technical Elements of Machine Learning for Intellectual Property Law’ [2020] SSRN Electronic Journal, 1–2.

<sup>42</sup> So (n 41) 2.

the choice of algorithm, the initial setup of the algorithm's parameters and in the choice of prediction rule used for classification.<sup>43</sup>

Unsupervised learning is used when the data to be used is unlabeled and with the goal of uncovering hidden structures in the data. As an example of unsupervised learning, it can be used to analyze pictures taken by a camera and classify what parts of the image belongs to the background of the picture without any training data as to provide a correct answer if the prediction was true or false.<sup>44</sup>

One task of unsupervised learning is clustering with the goal of dividing the data into clusters with similar characteristics such as customer segmentation to target advertising to customers with certain characteristics.<sup>45</sup>

Another use within unsupervised learning is what is called generative modeling, in which two algorithms interact with each other in an iterative way, where one algorithm feeds a set of fake data and a set real data to the other algorithm whose task is to correctly determine which data is fake and which is real. Both algorithms update their parameters in each iteration based on the outcome until they have been fully trained at the point where it is no longer possible to discern the real from the fake data. This model can be used to generate content such as photos, artworks, or poems.<sup>46</sup>

The last phase is interpretation of the result of the data analysis. Understanding the result of the data analysis requires interpretation and verification. There are many possible sources of error in data analysis: bugs in the computer system, the computer model may rely on false assumptions, and the result may have been based on erroneous data.<sup>47</sup>

---

<sup>43</sup> So (n 41) 3-9, 16.

<sup>44</sup> So (n 41) 9-10.

<sup>45</sup> So (n 41) 10-11.

<sup>46</sup> So (n 41) 11-13.

<sup>47</sup> Agrawal and others (n 16) 6-7.

The main forms of human involvement in unsupervised learning lies in the choice of training data and the choice and configuration of the algorithms such as the criterion used to define the clusters of data.<sup>48</sup>

Reinforced learning is based on the idea, instead of clustering, uncovering hidden structures of data or generating data that seems as real as possible, of the algorithm interacting with an environment whereas the interactions either can be punished or rewarded based on the state of the environment. As an example, reinforced learning could learn to play tic-tac-toe by rewarding a won state of the game and punishing a lost state of the game and letting all other states be neutral. The algorithm the needs to seek out the most rewarding way to interact based on a balance between its previous experience of interactions that leads to the reward vs exploring new potentially more rewarding interactions. Besides to use in teaching an AI how to play games, such as the boardgame Go, or the computer game Starcraft, it is also used in autonomous driving. The main human involvement in reinforced learning lies in carefully implementing a proper algorithm and in the significant need of computational resources as well as in the choice of which states should be rewarded or penalized.<sup>49</sup>

In summary we can talk about unstructured, semi-structured, and structured data which relates to the structure of how the data have been stored. The data itself can be computer- or human-generated. The data can be stored in the state it is collected, what can be defined as ‘raw data’ or in an altered form. After the data have been accessed and analyzed it will result in an output of form sort, which often will need to be interpreted and presented in various other forms.

## **2.4 Copyright and Related Rights Under the Infosoc Directive**

---

<sup>48</sup> So (n 41) 9–13, 16.

<sup>49</sup> So (n 41) 14–16.

## 2.4.1 Introduction

As seen in section 2.3 the Big Data process can involve many different types of data. Data can be found in the process of collecting and storing the data in the form of human- or computer-generated data. After the data has been analyzed it will result in an output that can require further adaptations.

The following analysis will go through the applicability of copyright protection to the various possible forms of data involved in the Big Data process.

## 2.4.2 Influencing International Law

The Infosoc Directive is not a full harmonization directive. Thus, the Member States of the EU retain residual domestic competence to further regulate copyright in the respective member state. However, the domestic competence is restricted by the relevant case law of the Court of Justice of the European Union (CJEU) on the Infosoc Directive as well as by the Member States obligations under international conventions.<sup>50</sup>

International conventions do not only restrict the competence of the Member States but also the interpretation of the Infosoc Directive. The CJEU has held that EU law ‘must, so far as possible, be interpreted in a manner that is consistent with international law, in particular where its provisions are intended specifically to give effect to an international agreement concluded by the Community’.<sup>51</sup> The EU, as well as all the Member States, has ratified two conventions in the field of authorial works and related rights. The agreement on trade-related aspects of intellectual property rights (TRIPS Agreement), which is administered by the World Trade Organization (WTO), and the WIPO Copyright Treaty (WCT).<sup>52</sup> Both the TRIPS Agreement and the WCT provides protection to authorial works through a reference to Article

---

<sup>50</sup> Justine Pila and Paul Torremans, *European Intellectual Property Law* (Oxford University Press 2019) 222-223.

<sup>51</sup> Case C-306/05 *Sociedad General de Autores y Editores de España (SGAE) v Rafael Hoteles SA* [2006] ECLI:EU:C:2006:764, para 35.

<sup>52</sup> Pila and Torremans (n 50) 225-227.

1 through 21 of the Berne Convention<sup>53</sup> (1971) besides an exception provided for in the TRIPS Agreement which excludes an article on the protection of authors moral rights.<sup>54</sup> Accordingly, the substantial rules on copyright protection of authorial works under international law is stipulated in the Berne Convention.

### **2.4.3 Protected Subject Matter**

Starting with international foundation of copyright, Article 2(1) in the Berne Convention defines literary and artistic works as ‘every production in the literary, scientific and artistic domain, whatever may be the mode or form of its expression’. However, what constitutes a work is not further explained in the convention itself but has been explained in the guide to the Berne convention as encompassing intellectual creations which may entail that a work possesses creativity or reflecting the personality of the maker. Works also must be original in the sense of being an author’s own intellectual creation as opposed to being a copy of another work. This does not extend to a requirement of novelty, two works that capture the same expression can be original if they have been created independently of each other.<sup>55</sup>

The guide further explains that a fundamental point of copyright protection is that ideas has been excluded from this scope of protection. The exclusion of ideas from copyright is explained by the fact that patent law is designed to protect ideas. It should however be pointed out that the expression that an idea has been dressed in can fall within the scope of copyright protection, but not the idea as such.<sup>56</sup> In the TRIPS Agreement and WCT this limitation on the scope of protection has been expressed as ‘[c]opyright protection shall extend to expressions and not to ideas, procedures, methods of operation or

---

<sup>53</sup> Berne Convention for the Protection of Literary and Artistic Works.

<sup>54</sup> Agreement on Trade-Related Aspects of Intellectual Property Rights as Amended by the 2005 Protocol Amending the TRIPS Agreement, art 9(1); Pila and Torremans (n 50) 226–227.

<sup>55</sup> Claude Masouyé, *Guide to the Berne Convention for the Protection of Literary and Artistic Works (Paris Act, 1971)* (WIPO 1978) 17.

<sup>56</sup> Masouyé (n 55) 12.

mathematical concepts as such.’<sup>57</sup> Furthermore, ‘news of the day or to miscellaneous facts having the character of mere items of press information’<sup>58</sup> has been excluded from the scope of protection of copyright in the Berne Convention. This exclusion follows from the fact that facts does not qualify as intellectual creations since facts are discovered, not created. Like ideas, facts can be expressed in a way to qualify as a work due the expression thereof being an intellectual creation, but the protection won’t include the fact itself.<sup>59</sup>

Moving on to EU law, the text of the Infosoc Directive does not stipulate any requirements for protection other than it must be a work from an author.<sup>60</sup> In the caselaw of the CJEU it is explained that the Infosoc Directive is intended to implement the Berne Convention and through the obligation in Article 1(4) of WCT follows an obligation for the EU to comply with Articles 1 to 21 of the Berne Convention.<sup>61</sup> Not surprisingly then, the legal requirements developed in the caselaw of the CJEU in order to qualify as a work under EU law follows a similar language as found in the guide to the Berne Convention.

In *Cofemel* CJEU held that the concept of work requires an original subject matter, in the sense of being the author’s own intellectual creation and that the classification as a work is reserved to the elements that qualify as such expressions.<sup>62</sup> To qualify as an original work ‘it is both necessary and sufficient that the subject matter reflects the personality of its author, as an expression of his free and creative choices’.<sup>63</sup> It is enough that creative choices has been made to meet the originality criterion without having to further take into account the intellectual effort or skill needed to create the work.<sup>64</sup> Even if the elements of a work themselves does not qualify as

---

<sup>57</sup> See Article 9(2) TRIPS Agreement and Article 2 WCT.

<sup>58</sup> Article 2(8) Berne Convention.

<sup>59</sup> Masouyé (n 55) 22–23.

<sup>60</sup> See Articles 2–4 Infosoc Directive.

<sup>61</sup> Case C-310/17 *Levola Hengelo BV v Smilde Foods BV* [2018] ECLI:EU:C:2018:899, para 38.

<sup>62</sup> Case C-683/17 *Cofemel — Sociedade de Vestuário SA v G-Star Raw CV* [2019] ECLI:EU:C:2019:721, para 29.

<sup>63</sup> *Cofemel* (n 62), para 30.

<sup>64</sup> Case C-469/17 *Funke Medien NRW GmbH v Bundesrepublik Deutschland* [2019] ECLI:EU:C:2019:623, para 23; even if a technical result of a subject matter is shown to be



original, as in the case of individual words, the CJEU has held that the sequence and combination of as few as 11 consecutive words as part of a work can leave enough scope to be regarded as an intellectual creation.<sup>65</sup> Copyright protection in regard to works is granted formless as soon as an author has given expression to an intellectual creation without any need for registration. Furthermore, a subject matter must be ‘expressed in a manner which makes it identifiable with sufficient precision and objectivity’<sup>66</sup> in order to avoid subjectivity in establishing a protected subject matter as well as to enable authorities and individuals to clearly and precisely identify protected subject matter.<sup>67</sup>

If an expression has ‘been dictated by technical considerations, rules or other constraints which have left no room for creative freedom’<sup>68</sup> it cannot be considered original. However, even when an expression contains components dictated by such constraints, if the author has made free and creative choices beyond such constraints, it may qualify as a work, albeit excluding the components that has been dictated only by the constraints as meeting the originality criterion. The reason given by the CJEU, is that where such components has been ‘dictated by their technical function, the different methods of implementing an idea are so limited that the idea and the expression become indissociable’.<sup>69</sup> Originality can also be precluded in purely informative documents, where the content essentially has been determined by the information it contains, in so far as the expression has been become indissociable from the information conveyed.<sup>70</sup> This reflects the exclusion of facts and ideas from copyright protection as found in the Berne Convention, TRIPS Agreement and WCT.

---

able to be achieved in more than one way, such freedom of choice in itself does not prove that creative choices actually has been made, see para 24.

<sup>65</sup> Even if the elements themselves does not qualify as original, as in the case of individual words, the sequence and combination of consecutive words,

<sup>66</sup> *Levola Hengelo* (n 61), para 40.

<sup>67</sup> *Levola Hengelo* (n 61), para 41.

<sup>68</sup> Case C-833/18 *SI, Brompton Bicycle Ltd v Chedech/Get2Get* [2020] ECLI:EU:C:2020:461, para 24.

<sup>69</sup> *Brompton Bicycle* (n 68), para 27.

<sup>70</sup> *Funke Medien* (n 64), para 24.

## 2.4.4 Authorship

Starting again, with the international aspects of copyright law, the Berne Convention offers rights to authors in respect of their literary and artistic works without delving deeper on the question of who can be regarded as the author.<sup>71</sup> In the guide to the convention the lack of definition on the topic of authorship is explained by the existence of diverging national legislation on the topic, where some countries only recognize natural persons, while other countries also recognize legal persons as authors.<sup>72</sup> It is generally accepted that the intended author is the natural person that creates a work.<sup>73</sup> As such, it seems safe to say to the Berne Convention impose an obligation to recognize natural persons as authors, while, as a convention on minimum levels of protection, leaves the question of recognizing authorship by legal persons a question for national law.

Regarding EU law, alike the Berne Convention, the Infosoc Directive does not offer a definition on who can be considered the author. However, in the Terms Directive<sup>74</sup>, which lays down the duration of authors' rights, refers to the authors within the meaning of Article 2 of the Berne Convention.<sup>75</sup> As already mentioned, the Berne Convention only requires recognition of natural persons as authors. While EU law only requires the recognition of natural persons as authors, member states may designate legal persons as the rightholders in relation to a work, in which case the EU law regulates the duration of the rights.<sup>76</sup> It should also be noted that CJEU has held that the Infosoc Directive 'is based on principles and rules already laid down in the

---

<sup>71</sup> Article 1 Berne Convention.

<sup>72</sup> Masouyé (n 55) 11.

<sup>73</sup> Paul Goldstein and P. Berndt Hugenholtz, *International Copyright: Principles, Law, and Practice* (Oxford University Press 2019) 176.

<sup>74</sup> Directive 2006/116/EC on the term of protection of copyright and certain related rights.

<sup>75</sup> Article 2(1) Term Directive.

<sup>76</sup> Article 2(4) Term Directive.

directives in force in the area of intellectual property'<sup>77</sup> which mandates that, due to 'the requirements of unity of the European Union legal order and its coherence, the concepts used by that body of directives must have the same meaning, unless the European Union legislature has, in a specific legislative context, expressed a different intention.'<sup>78</sup> It thus follows that only natural persons are recognized as authors under the Infosoc Directive since that directive is preceded by the Term Directive.<sup>79</sup>

## 2.4.5 Rights

Before assessing the applicability of copyright under the Infosoc Directive a brief overview of the rights and limitations of copyright will be offered.

The exclusive economic rights offered to an author lies in reproduction, communication to the public and distribution.<sup>80</sup>

The reproduction right gives authors the 'right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part'<sup>81</sup>. As mentioned in Recital 14 of the directive and repeatedly in the case law of the CJEU, the reproduction right is to be given a broad definition.<sup>82</sup>

The right of communication is stipulated as providing the author the exclusive right 'to authorise or prohibit any communication to the public of their works, by wire or wireless means, including the making available to the public of

---

<sup>77</sup> Joined Cases C-403/08 and C-429/08 *Football Association Premier League Ltd, NetMed Hellas SA, Multichoice Hellas SA v QC Leisure, David Richardson, AV Station plc, Malcolm Chamberlain, Michael Madden, SR Leisure Ltd, Philip George Charles Houghton, Derek Owen* (C-403/08) and *Karen Murphy v Media Protection Services Ltd* (C-429/08) [2011] ECLI:EU:C:2011:631, para 187.

<sup>78</sup> *Football Association Premier League* (n 77), para 188.

<sup>79</sup> It should be noted however, that the current Term Directive does not precede the Infosoc Directive, but that Directive 93/98/EEC of 29 October 1993 harmonizing the term of protection of copyright and certain related rights, which the current Term Directive replaced, offers the same definition of author as well as duration of right for legal persons as in the current directive, see Articles 1(1) and 1(4).

<sup>80</sup> Articles 2–4 Infosoc Directive.

<sup>81</sup> Article 2 Infosoc Directive.

<sup>82</sup> See for example Case C-145/10 *Eva-Maria Painer v Standard VerlagsGmbH and others* [2011] ECLI:EU:C:2011:798, para 96.

their works in such a way that members of the public may access them from a place and at a time individually chosen by them.’<sup>83</sup> As with the right of reproduction, what is considered an act of communication must be construed broadly such as it is enough that a ‘public may access it, irrespective of whether they avail themselves of that opportunity’.<sup>84</sup> Public requires ‘an indeterminate number of potential recipients and implies, moreover, a fairly large number of persons’.<sup>85</sup>

The right of distribution gives the author ‘the exclusive right to authorise or prohibit any form of distribution to the public by sale or otherwise’<sup>86</sup>. The distribution right is exhausted such as that it no longer can be exercised in relation to the specific physical works that has been sold or ownership has been transferred within the EU.<sup>87</sup>

## 2.4.6 Limitations and Exceptions

The only mandatory exception laid out in the Infosoc Directive provides an exception in respect to the reproduction right.<sup>88</sup> It covers transient or incidental acts of temporary reproduction ‘which are an integral and essential part of a technological process and the sole purpose of which is to enable’ either transmissions between third parties by an intermediary or lawful use of a work or other subject matter as long as the act of reproduction does not have an independent economic significance.

---

<sup>83</sup> Article 3(1) Infosoc Directive.

<sup>84</sup> Case C-466/12 *Nils Svensson and others v Retriever Sverige AB* [2014] ECLI:EU:C:2014:76, paras 17 and 19.

<sup>85</sup> *Svensson* (n 84), para 21; in some cases it must also be considered whether this public consists of a new public in relation to previous acts of communications from the author, but a further examination of this criteria falls outside the scope of this thesis, but more on this subject can found in Eleonora Rosatis article ‘When does a communication to the public under EU copyright law need to be to a “new public”?’ in *European Law Review* 2020, 45(6), 802–823.

<sup>86</sup> Article 4(1) Infosoc Directive.

<sup>87</sup> Article 4(2) Infosoc Directive; Case C-263/18 *Nederlands Uitgeversverbond, Groep Algemene Uitgevers v Tom Kabinet Internet BV, Tom Kabinet Holding BV, Tom Kabinet Uitgeverij BV* [2019] ECLI:EU:C:2019:1111, paras 44-45.

<sup>88</sup> Article 5(1) Infosoc Directive.

An act has been held to be transient ‘only if its duration is limited to what is necessary for the proper completion of the technological process in question, it being understood that that process must be automated so that it deletes that act automatically, without human intervention, once its function of enabling the completion of such a process has come to an end.’<sup>89</sup> Such acts include *inter alia* acts necessary for browsing or caching.<sup>90</sup> Incidental acts of reproduction are those that are incidental to the technological process in so far as they also are temporary.<sup>91</sup> To be considered as an integral and essential part of a technological process an act must be ‘necessary, in that the technological process concerned could not function correctly and efficiently without that act.’ and be carried out entirely within the context of the technological process.<sup>92</sup>

For an act to have no independent economic significance ‘the economic advantage derived from their implementation must not be either distinct or separable from the economic advantage derived from the lawful use of the work concerned and it must not generate an additional economic advantage going beyond that derived from that use of the protected work’.<sup>93</sup> This is the case if the author of the temporary reproductions ‘is likely to make a profit due to the economic exploitation of the temporary reproductions themselves’ or ‘if the acts of temporary reproduction lead to a change in the subject matter reproduced, as it exists when the technological process concerned is initiated, because those acts no longer aim to facilitate its use, but the use of a different subject matter.’<sup>94</sup>

Any exception under the Infosoc Directive must also pass the so called three step test which stipulates that the exceptions and limitations ‘shall only be applied in [(1)] certain special cases [(2)] which do not conflict with a normal

---

<sup>89</sup> Case C-5/08 *Infopaq International A/S v Danske Dagblades Forening* [2009] ECLI:EU:C:2009:465, para 64.

<sup>90</sup> Recital 15 Infosoc Directive.

<sup>91</sup> Triaille, d’Argenteuil, and Francquen (n 37) 43.

<sup>92</sup> Case C-302/10 *Infopaq International A/S v Danske Dagblades Forening* [2012] ECLI:EU:C:2012:16, para 30.

<sup>93</sup> *Infopaq II* (n 92), para 50.

<sup>94</sup> *Infopaq II* (n 92), paras 52–53.

exploitation of the work or other subject-matter and [(3)] do not unreasonably prejudice the legitimate interests of the rightholder'.<sup>95</sup>

Additionally, two exceptions to the reproduction right have been laid out in the DSM Directive<sup>96</sup>. One of the exceptions is applicable only in regard to research organisations and cultural heritage institutions while the other is applicable for all other actors.<sup>97</sup> Both exceptions aim to enhance legal certainty in the scope of 'text and data mining' which is defined 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'.<sup>98</sup> The definition of text and data mining thus seems to cover the entire Big Data process as described in section 2.3.

As a first requirement to rely on either exception is lawful access to the reproduced or extracted content which basically means that the reproduced content should be freely available, accessible through an open access policy or otherwise through a contractual arrangement.<sup>99</sup>

Starting with a brief overview of the specifics regarding the research exception. A research institution is defined as any entity whose primary goal is to conduct scientific research or to carry out educational activities also involving the conduct of scientific research if it is done on a non-profit basis or if all the profits are reinvested in its scientific research or pursuant to a public-interest mission recognized by a Member State and provided that an undertaking that exercises influential a decisive influence upon the research organisation is not given a preferential treatment to enjoy the research

---

<sup>95</sup> Article 5(5) Infosoc Directive.

<sup>96</sup> Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (DSM Directive).

<sup>97</sup> Articles 3-4 DSM Directive.

<sup>98</sup> Article 2(2) DSM Directive; Recital 8 and 12 DSM Directive.

<sup>99</sup> Article 3-4 DSM Directive; Recital 14 DSM Directive.

results.<sup>100</sup> A cultural heritage institution means a ‘publicly accessible library or museum, an archive or a film or audio heritage institution’.<sup>101</sup>

To rely on the exception, the reproduction or extraction by a research organisation or cultural heritage institution must be conducted for the purposes of text and data mining as well as for the purpose of scientific research, which includes both the natural and human sciences.<sup>102</sup>

An important aspect of the research exception is that any contractual restrictions to this exception is unenforceable.<sup>103</sup>

Any copies made in relying on the research exception must be stored with an appropriate level of security and may be retained for the purpose of scientific research, which includes keeping the copies for verification of research results.<sup>104</sup> However, rightholders of the copied works or other subject matter shall be allowed to undertake necessary measures that do not undermine the effective application of the exception in order to ensure the security and integrity of the networks or databases which hosts the copied content.<sup>105</sup>

Next, moving on to an explanation of the specifics regarding the exception of general applicability. Anyone can rely on this exception as long as the reproduction or extraction is done for the purpose of text and data mining.<sup>106</sup> However, to rely on the exception it is also necessary that the rightholders have not expressly reserved the use of the works or other subject matter in an appropriate manner, such as in a machine-readable manner for publicly

---

<sup>100</sup> Article 2(1) DSM Directive; Recital 12 of the DSM Directive exemplifies that public-interest missions could ‘be reflected through public funding or through provisions in national laws or public contracts’.

<sup>101</sup> Article 2(3) DSM Directive.

<sup>102</sup> Article 3(1) DSM Directive; Recital 12 DSM Directive.

<sup>103</sup> Article 7(1) DSM Directive.

<sup>104</sup> Article 3(2) DSM Directive.

<sup>105</sup> Article 3(3) and Recital 16 DSM Directive.

<sup>106</sup> Article 4(1) DSM Directive.

available content.<sup>107</sup> It also follows that contractual obligations may limit the reliance on the exception.

Copies made in reliance on the exception may be kept for as long as necessary for the purposes of text and data mining.<sup>108</sup> This would likely entail an obligation to delete the data obtained through the exception after the data has been analysed and a resulted in an output of some sort of information since the definition of text and data mining refers the obtaining information from digital content with any automated analytical technique.

Both exceptions in the DSM Directive to reproductions and extractions must also pass the three-step test as laid in the Infosoc Directive.<sup>109</sup>

## **2.4.7 A Second Layer of Protection: Technological Protection Measures**

Besides the exclusive rights of reproduction, communication to the public and distributions the Infosoc Directive also provides for protection against circumvention of technological protection measures (TPM) that protect a work, related right or the *sui generis* database right.<sup>110</sup> This procedure has often been regarded as granting the rightholders a second layer of protection.<sup>111</sup>

TPM is defined as ‘any technology, device or component that, in the normal course of its operation, is designed to prevent or restrict acts, in respect of works or other subject-matter’<sup>112</sup> unless authorised by the rightsholder. The way TPMs can protect a work can include *inter alia* access-controls,

---

<sup>107</sup> Article 4(3) DSM Directive.

<sup>108</sup> Article 4(2) DSM Directive.

<sup>109</sup> Article 7(2) DSM Directive; regarding the three-step test see Article 5(5) Infosoc Directive.

<sup>110</sup> Article 6(1) and 6(3) Infosoc Directive.

<sup>111</sup> Mirelle van Eechoud and others ‘Harmonizing European Copyright Law: The Challenges of Better Law Making’ (2012) Amsterdam Law School Research Paper No. 2012-07, Institute for Information Law Research Paper No. 2012-07, 156 <<http://ssrn.com/abstract=2049935>>.

<sup>112</sup> Article 6(3) Infosoc Directive.



prevention of certain uses of the work, or by preventing the integrity of the work by preventing modifications.<sup>113</sup>

The conditions for TPMs to fall within the scope of protection against circumventions is that they are effective and that the person that performs the circumvention is doing so knowingly or with reasonable grounds to know so.<sup>114</sup> Effective TPM is defined as ‘where the use of a protected work or other subject-matter is controlled by the rightholders through application of an access control or protection process, such as encryption, scrambling or other transformation of the work or other subject-matter or a copy control mechanism, which achieves the protection objective’.<sup>115</sup> If the conditions are met the Member States are to provide adequate legal protection against circumvention of the TPM.

In relation to some of the non-mandatory limitations and exceptions provided for under the Infosoc Directive if a Member State have implemented such limitations or exceptions, they must also provide for the beneficiaries of such limitations or exceptions to be able to rely on them to the necessary extent. However, this right does only apply in the absence of voluntary measures taken by the rightsholder and under the condition that the beneficiary also has legal access to the protected work.<sup>116</sup>

Furthermore, the right to rely on a limitation or exception is further conditioned upon that the protected work or other subject matter has not been ‘made available to the public on agreed contractual terms in such a way that members of the public may access them from a place and at a time individually chosen by them’<sup>117</sup> On this note, it should be considered that the right to rely upon limitations and exceptions in regard to TPMs also is provided for under the DSM Directive, in relation to the exception to the

---

<sup>113</sup> Eechoud and others (n 111) 131–132.

<sup>114</sup> Article 6(1) Infosoc Directive.

<sup>115</sup> Article 6(3) Infosoc Directive.

<sup>116</sup> Article 6(4) Infosoc Directive.

<sup>117</sup> Article 6(4) Infosoc Directive.

reproduction right for the purpose of text and data mining. However, under the DSM Directive this right is not contingent upon that the work has not been made available to the public.<sup>118</sup> If the rightsholder provides such a measure, either voluntarily or through measures taken by the Member State, the technological measure applied in doing so is also protected as a TPM.<sup>119</sup>

## **2.4.8 Assessment**

### **2.4.8.1 Human-Generated Data**

First an assessment of the applicability of copyright protection under the Infosoc Directive of human-generated data will be conducted, followed by an assessment of computer-generated data.

Human-generated data can consist of for example written texts, such as search terms on a search engine, tweets, updates on social media, linguistic content on homepages or digitalization of all books such as in the case of Google Books library project. On this note it should be held in mind that it cannot even be ruled out that as few as 11 consecutive words holds the possibility of meeting the originality criterion if free and creative choices have been made in the choice, sequence, and combination of those words.<sup>120</sup> Human-generated data could also refer to, for example photos, images, audio, or videos.

In relation to such human-generated subject-matter it is very likely that it will be protected by copyright given that the originality criterion sets a very low threshold in that the subject matter must allow for the author to make free and creative choices and that free and creative choices have come to expression. Given that part of a subject matter, such as only 11 words from a news article, may potentially meet the originality criterion, human-generated data of texts, images, photos, audio or video would likely for the most part be protected under the Infosoc Directive.

---

<sup>118</sup> Article 7(2) DSM Directive.

<sup>119</sup> Article 6(4) Infosoc Directive.

<sup>120</sup> *Infopaq* (n 89), para 45 and 48.

Since the copyright is vested in the author of a work the initial rightsholder of a work is likely not the actor that is processing the data in a Big Data context, but a user of a digital platform. To use such protected works for big data analysis it would first be necessary to obtain a license from the original author to use the data the purpose of data analysis. It would also require allowance to reproduce the work, or otherwise, to transfer the ownership of the work to the data analyser to enable the necessary storage of the data. Ownership of the work would be necessary if the intent is to sell data, in so far as the work or parts of the work would qualify as intellectual creations. A license to use the work for data analysis would suffice if the intent of handling the data do not lie in reselling the data but rather in selling or using the result of the data analysis as long as the work or parts of the analysed works aren't replicated in the resulting output. Normally such an output would not contain any content of the analysed works.<sup>121</sup>

To process works for data analysis without the need to obtain a license or transfer of ownership it would be necessary to rely on an exception or limitation to copyright. The mandatory exception in the Infosoc Directive for temporary reproductions would not be applicable in most Big Data processes. First, since reproductions in the Big Data process generally are stored in a database for loading and processing at a later stage it would likely not be able to be regarded as temporary, and it would further entail that the data also are automatically deleted at the end of the technological process. Furthermore, since much data needs to be transformed into a computer readable format this would entail a change in the subject so as it would be regarded as carrying independent economic significance. In the event that some niche small scale data analysis would allow for relying on this exception, it would only allow to acquire data in the result of the data analysis, since the analysed data itself must be automatically deleted at the end of the technological process.

---

<sup>121</sup> Triaille, d'Argenteuil, and Francquen (n 37) 49-50.

When it comes to the exception to reproductions and extractions under the DSM Directive for text and data mining the main criteria to prevent utilization of the exception would be to prevent lawful access to a work as both exceptions are contingent upon lawful access. As long as the work is not publicly available access would need to be sought through contractual terms.

As far the general exception goes, that is applicable for all actors besides research organisations and cultural heritage institutions, even if having lawful access to the work, it would in all cases still be possible to protect the work from the scope of this exception, by expressly reserving the use of the work in an appropriate manner. Such reservations would however not affect the applicability of the research exception since contractual provisions that limit the applicability of the exception are unenforceable.

As a last line of defence, a work or other subject matter could be protected by TPM. A beneficiary of an exception or limitation that falls within the scope of the TPM provision in the Infosoc Directive, such as the text and data mining exceptions, might still be effectively prevented from enjoying this right. The right to let a beneficiary of an exception also enjoy that exception even if the work has been protected by a TPM will necessarily involve transaction costs.

Firstly, because the right is dependent on that the rightholder does not voluntarily takes such measures. This would require that the beneficiary seeks out every rightholder individually, which in the case of text and data mining, could involve a very large set of rightholders. Even if the transaction costs associated with such an endeavour could be overcome, secondly, since the national measures have not been harmonized it could involve many different and varying national procedures which would entail further costs.<sup>122</sup> In effect

---

<sup>122</sup> See Marcella Favale, 'Technological Protection Measures and Copyright Exceptions in EU27: Towards the Harmonization' <[https://law.depaul.edu/academics/centers-institutes-initiatives/center-for-intellectual-property-law-and-information-technology/programs/Documents/ipsc\\_2007/paper/Marcella\\_FavalePaper.pdf](https://law.depaul.edu/academics/centers-institutes-initiatives/center-for-intellectual-property-law-and-information-technology/programs/Documents/ipsc_2007/paper/Marcella_FavalePaper.pdf)> 12–14 for an overview of the varying national implementations.

TPMs thus seems to offer a strong layer of protection to copyright protected subject matter and which goes beyond what seems to have been intended due to the potentially insurmountable transaction costs associated with overcoming TPM.

Additionally, TPM offers protection that goes beyond protecting only works or other subject matter, due to the nature of TPM in limiting access and/or use of works, as to naturally also protect access or use of all digital content even if not protected by copyright. However, a limiting factor to TPMs is that they might not be able to be effectively applied depending on the context in which the protected content is to be used.

Human-generated data could however also be generated within the Big Data process in which case it seems most likely to be the case when it comes to interpreting and presenting the result of the data analysis. This could for example be a visual representation of the interpretation or a linguistic explanation that qualifies as an intellectual creation. However, if the result is a fact, the fact as an element of the protected subject matter would itself not be protected, just the expression it forms a part of.

In principle it also seems possible that the output of data analysis could qualify as works depending on the level of creative choices allowed for and how those creative choices bear influence on the expression in the form of the output of the data analysis. It does therefore not seem obvious to rule out *per se* that certain outputs could not qualify as intellectual creations of the natural persons that made the creative choices affecting the output in the big data process.

That technology is used to produce the expression does not rule out the possibility of an expression being an intellectual creation as the reflection of the creative choices made by the author, as in the case of photography. Regarding photography, CJEU has in the *Painer* judgement held that a ‘photographer can make free and creative choices in several ways and at

various points in its production.’<sup>123</sup> The CJEU followed up by giving concrete examples of choices whereas in ‘the preparation phase, the photographer can choose the background, the subject’s pose and the lighting. When taking a portrait photograph, he can choose the framing, the angle of view and the atmosphere created. Finally, when selecting the snapshot, the photographer may choose from a variety of developing techniques the one he wishes to adopt or, where appropriate, use computer software.’<sup>124</sup>

In analogy with the reasoning of the CJEU in the *Painer* case it does not seem unreasonable to think that it should also be possible for the person making the choices in a machine learning algorithm to make such creative choices as to qualify the expression as the authors own intellectual creation depending on the task of the machine learning. The predictions made in supervised learning as to what the input data resembles the most regarding the labelled training and the clustering of data in supervised machine learning would not provide enough scope of giving expression to creative choices as the output would just reflect an idea of which cluster or which labelled data the data matches the most. It would also be hard to see how reinforced machine learning could produce an intellectual creation in the instances of playing a game or in the autonomous driving of a car. In the case of sporting events, CJEU has held that they cannot be regarded as intellectual creations, and particularly in the case of football matches since the rules of the game leaves no room for creative choices.<sup>125</sup>

However, in the case of the unsupervised generative modelling where the output of an fully trained algorithm will resemble the training data provided to the algorithm to the extent that it is not possible for the AI to distinguish between the real training data and the fake generated data it seems like it should not be ruled out *per se* that such expressions could not qualify as

---

<sup>123</sup> *Painer* (n 82), para 90.

<sup>124</sup> *Painer* (n 82), para 91.

<sup>125</sup> Joined Cases C-403/08 and C-429/08 *Football Association Premier League Ltd and others v QC Leisure and others and Karen Murphy v Media Protection Services Ltd* [2011] ECLI:EU:C:2011:631, para 98.

intellectual creations with a natural person as the author. For example, if the training data consists of texts or images made by one natural person and which would qualify as works, it would seem possible that the theme of the works provided also would be expressed in the output of the data analysis. Even if it would be close to impossible to predict the exact expression in the output, it does not seem possible to entirely rule out the possibility of such as expressions reflecting the creative choices made by the author of the training data. In analogy with photography in the *Painer* case creative choices was held to be able to be made in the selection of the snapshot in the choice of developing techniques or use of computer software. Developing techniques or filters provided for by computer software would also entail a modification of the original which likely includes unforeseen alterations while not ruling out such expression as intellectual creations. Similarly, it would seem possible that the same could potentially hold true in some instances of generative modelling. A related issue, which I won't delve deeper into, in the case of an output qualifying as a work, would be if the computer engineer(s) that made and configured the algorithm could have made creative choices reflected in the expression as well as to be considered as co-authors.

Even if it in principle does not seem possible to rule out that the output of data analysis in the case of generative modelling could be intellectual creations, it would carry a lot of legal uncertainty as no case law exists on the matter. Furthermore, the possibility seems to only relate to certain special cases.

Human-generated data can also refer to data generated from other human interactions than those that qualify as intellectual creations, such as records of certain human interactions. These types of interactions could be exemplified by *inter alia* a 'like' on Facebook, a record of a transaction taking place on a web-shop or records of when and how a user accesses and interacts with a website or app. Data consisting of records of such interactions would be a record of an empirical fact of something taking place and thus fall outside the scope of copyright protection since the notion of work does not extend to

facts and ideas. Furthermore, data represented in these types of expressions would most likely also fall within the realm of expressions that has been dictated by technical considerations or other constraints as to exclude the necessary creative freedom needed to qualify as intellectual creations.

#### **2.4.8.2 Machine-Generated Data**

As for machine-generated data it could consist of measurements of varying sorts picked up by sensors, in for example Internet of Things (IoT) devices which are physical devices embedded with connectivity software to enable such devices to connect and exchange data.<sup>126</sup> Machine-generated data could also consist of the result of a data analysis on a given data set. A main hurdle to fall within the scope of copyright protection for both these categories of data is that such measurements or results of data analysis likely would be a record of a fact, such as distance, temperature, time, date or, in the case of data analysis, a probability or categorization. The data could also be represented in an idea, such as the output in form of a single word.

Secondly, even if the computer-generated data falls within the scope of subject matter that could be protected by copyright, such data would need to have an author. As only natural persons are recognized as potential authors under the Infosoc Directive it would seem to disqualify this entire category of data from the scope of copyright. As discussed above it seems that it cannot be ruled out entirely that outputs from data analysis could have a human author in the case of human-generated data. However, in the case of machine-generated data it seems highly unlikely that the choices made by natural persons in the choice and configuration of the algorithm alone would suffice to qualify the output as being the expression of those choices.

In conclusion it seems that machine-generated data would fall outside the scope of copyright protection.

---

<sup>126</sup> Commission, 'Commission Staff Working Document Impact Assessment Report Accompanying the document Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)' SWD(2022) 34 final, 75.



## 2.5 Protection of Data under the Database Directive

### 2.5.1 Scope of the Database Directive

The Database Directive<sup>127</sup> offers two types of protection of databases. First, copyright protection in databases that ‘by reason of the selection or arrangement of their contents, constitute the author’s own intellectual creation’<sup>128</sup> and secondly, a *sui generis* right that applies to databases which require a ‘substantial investment in either the obtaining, verification or presentation of the contents’<sup>129</sup>. These two rights are independent from each other with different objects and scopes of application.<sup>130</sup> The object for copyright protection in databases is granting authors protection of databases that are intellectual creations, while the *sui generis* right aims to protection databases whose maker has made substantial investment.

The scope of the directive covers all forms of databases, i.e., both physical and digital databases.<sup>131</sup> To be considered a database under the Database Directive a database must be constituted by ‘a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means.’<sup>132</sup>

The term database is to be given a ‘wide scope, unencumbered by considerations of a formal, technical or material nature’<sup>133</sup> and rather be defined by ‘in terms of its function’<sup>134</sup>. It applies to ‘literary, artistic, musical or other collections of works or collections of other material such as texts, sound, images, numbers, facts, and data’.<sup>135</sup>

---

<sup>127</sup> Directive 96/9/EC on the legal protection of databases (Database Directive).

<sup>128</sup> Article 3(1) Database Directive.

<sup>129</sup> Article 7(1) Database Directive.

<sup>130</sup> Case C-604/10 *Football Dataco Ltd and others v Yahoo! UK Ltd and others* [2012] ECLI:EU:C:2012:115, para 27.

<sup>131</sup> Article 1(1) Database Directive; *Freistaat Bayern* (n 133), para 14.

<sup>132</sup> Article 1(2) Database Directive.

<sup>133</sup> Case C-490/14 *Freistaat Bayern v Verlag Esterbauer GmbH* [2015] ECLI:EU:C:2015:735, para 12.

<sup>134</sup> *Freistaat Bayern* (n 133), para 16.

<sup>135</sup> *Freistaat Bayern* (n 133), para 14.

To be independent, the materials should be ‘separable from one another without their informative, literary, artistic, musical or other value being affected’.<sup>136</sup> Furthermore, ‘not only an individual piece of information, but also a combination of pieces of information can constitute ‘independent material’ as long as the value is not affected by separation of such information.’<sup>137</sup> The informative value of a material is not affected if ‘if it has autonomous informative value after being extracted from the collection’.<sup>138</sup> Even though material kept in a database tend to increase the materials value, and correspondingly, a decrease in value if it is extracted from the database, such change in value should not be taken into consideration for the purpose of classifying the material as independent, as long as it retains autonomous informative value after extraction.<sup>139</sup> The autonomous informative value of a material should not be judged by only the typical user of a collection, but from any interested third party.<sup>140</sup>

It is not required that the systematic or physical arrangement of the collection is physically apparent, but the collection should be contained in a fixed base and allow for the retrieval by any means, such as a technical process or through an index, table of contents, or a plan or method for classification, of the independent materials in the collection.<sup>141</sup>

A database can simultaneously qualify for protection under both copyright protection as well as the *sui generis* right under the Database Directive.<sup>142</sup>

It should also be kept in mind that if a database exists that falls within the scope of the definition under the Database Directive, it does not entail that such an database also would meet the criteria for copyright protection or fall

---

<sup>136</sup> Case C-444/02 *Fixtures Marketing Ltd v Organismos prognostikon agonon Podosfairou AE (OPAP)* [2004] ECLI:EU:C:2004:697, para 29.

<sup>137</sup> *Freistaat Bayern* (n 133), para 20.

<sup>138</sup> *Freistaat Bayern* (n 133), para 22.

<sup>139</sup> *Freistaat Bayern* (n 133), para 23.

<sup>140</sup> *Freistaat Bayern* (n 133), para 27.

<sup>141</sup> *Fixtures Marketing Ltd* (n 136), para 30.

<sup>142</sup> Article 7(4) Database Directive.

under the *sui generis* right. It also follows that the exceptions to copyright protection or the *sui generis* right is not applicable unless a database within the meaning of the Database Directive also meets the criteria for copyright protection or the *sui generis* right.<sup>143</sup>

## 2.5.2 Copyright Protection for Databases

Article 3 of the Database Directive stipulates that ‘databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation shall be protected as such by copyright’ and that ‘[n]o other criteria shall be applied to determine their eligibility for that protection’.

The notion of intellectual creation has the same meaning as the originality criterion established in the caselaw related to the Infosoc Directive. In the case of original databases, the criterion is met if the author expresses creative ability by making free and creative choices through the selection or arrangement of the data the database contains.<sup>144</sup> As with the originality criterion, this requires that when setting up the database, such creative choices are not prevented by technical considerations or other constraints.<sup>145</sup> For a more in-depth explanation of the originality criterion, see section 2.4.3 above.

The copyright protection of databases only protects the structure of the database, and thus, does not include protection of the the content, or elements of the content, of the database.<sup>146</sup>

The author of an original database is the natural person or group of natural persons that created the base. Legal persons may also be designated as the rightholder if permitted under the national law of a Member State.<sup>147</sup>

---

<sup>143</sup> Case C-30/14 *Ryanair Ltd v PR Aviation BV* [2015] ECLI:EU:C:2015:10.

<sup>144</sup> *Football Dataco* (n 130) paras 37-38.

<sup>145</sup> *Football Dataco* (n 130), para 39.

<sup>146</sup> *Football Dataco* (n 130), paras 30-31; Article 3(2) Database Directive.

<sup>147</sup> Article 4(1) Database Directive.

The author is granted exclusive rights in temporary or permanent reproduction by any means and in any form, in whole or in part, any type of alterations, any form of distribution to the public, and any communication, display or performance to the public.<sup>148</sup>

The Database Directive only contains one mandatory exception from the exclusive rights. The exception is applicable to lawful users of a database or a copy of a database for the purposes of accessing or for the normal use of the contents.<sup>149</sup>

### 2.5.3 The *Sui Generis* Database Right

The *sui generis* database right is applicable to databases in which the maker has made a qualitatively and/or quantitatively substantial investment in either the obtaining, verification or presentation of the contents.<sup>150</sup> The maker of the database is granted the right to ‘prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database.’<sup>151</sup>

It has been held by the CJEU that the investment in obtaining, verification or presentation of the contents ‘must be understood, generally, to refer to investment in the creation of that database as such.’<sup>152</sup>

Specifically in reference to investments in obtaining contents, this should be understood as referring ‘to the resources used to seek out existing independent materials and collect them in the database, and not to the resources used for the creation as such of independent materials.’<sup>153</sup> The reason for this distinction between obtained and created data, is that the purpose for the *sui generis* right is to ‘promote the establishment of storage and processing

---

<sup>148</sup> Article 5 Database Directive.

<sup>149</sup> Article 6(1) Database Directive.

<sup>150</sup> Article 7(1) Database Directive.

<sup>151</sup> Article 7(1) Database Directive.

<sup>152</sup> Case C-444/02 *Fixtures Marketing Ltd v Organismos prognostikon agonon Podosfairou AE (OPAP)* [2004] ECLI:EU:C:2004:697, para 39.

<sup>153</sup> *OPAP* (n 152), para 40.

systems for existing information and not the creation of materials capable of being collected subsequently in a database'.<sup>154</sup> Furthermore, it is argued that recital 19 of the Database Directive supports such an interpretation, since it states that a CD that contains a compilation of musical performances does not amount to a substantial investment. Thereby CJEU drew the conclusion that it would imply that the investment into creating the musical performances is not equivalent to obtaining the contents of a database.<sup>155</sup>

This distinction between created and obtained has led to many discussions on how to interpret and understand the differences between the two. It has been suggested that it should be interpreted as not protecting so called spin-off databases, which are created as a by-product of a main activity.<sup>156</sup> Others reject such an interpretation and holds it would go against the wording of the CJEU in the *British Horseracing* judgement.<sup>157</sup> I would have to agree with the latter opinion, as the CJEU held that 'the fact that the creation of a database is linked to the exercise of a principal activity in which the person creating the database is also the creator of the materials contained in the database does not, as such, preclude that person from claiming the protection of the sui generis right, provided that he establishes that the obtaining of those materials, their verification or their presentation [...] required substantial investment in quantitative or qualitative terms, which was independent of the resources used to create those materials.'<sup>158</sup> Thus, it cannot be ruled out that a database that is created as a spin-off, still have required a substantial investment.

Another issue raised by the creation/obtaining dichotomy is if recorded data is created, collected, or obtained. Recorded data is data that occur in nature and time and that can be recorded by some instrument to make them

---

<sup>154</sup> *OPAP* (n 152), para 40.

<sup>155</sup> *OPAP* (n 152), para 42.

<sup>156</sup> Robbert Fischer and others 'Study in support of the evaluation of Directive 96/9/EC on the legal protection of databases' (2018), 20.

<sup>157</sup> Guido Noto La Diega, 'Artificial Intelligence and databases in the age of big machine data' [2019] SSRN Electronic Journal, 119.

<sup>158</sup> Case C-203/02 *The British Horseracing Board Ltd and Others v William Hill Organization Ltd* [2004] ECLI:EU:C:2004:695, para 35.

intelligible and thus exists independently of their recording. The argument is that since recorded data is pre-existing to any record of it, it cannot be created; as opposed to horse racing fixtures, that does not exist until they have been arranged.<sup>159</sup> Recalling the distinction made by the court between seeking out existing independent material as opposed to creating material, where seeking out existing material shall be regarded as obtaining the material. This reasoning seems to suggest that, although a fact might exist in nature, without a recording of the fact it would not be able carry any value since it is unknown as a fact until recorded, and thus could not be considered as independent material. However, until further clarification on the issue by the CJEU this issue is shrouded in legal uncertainty. As an example, live football data is considered as obtained in the UK while it is considered to be created in Germany.<sup>160</sup>

Investments in verification of contents should be understood as referring to the resources used to ensure the reliability of the information in the database, or to monitor the accuracy of the materials in the creation of the database or during operation of the database while investments in presentation of the contents should be understood as ‘resources used for the purpose of giving the database its function of processing information, that is to say those used for the systematic or methodical arrangement of the materials contained in that database and the organisation of their individual accessibility.’<sup>161</sup>

Investments may consist in human, technical, or financial resources. When assessing whether such investments are substantial in quantitative terms it means that the resources used should be quantifiable, such as time or money, while qualitative investment are negatively defined in that it encompasses non-quantifiable investments, such as intellectual effort or energy.<sup>162</sup> As how much of an investment is substantial, this has been left for the national

---

<sup>159</sup> Paolo Burdese, ‘AI-generated databases. Do the creation/obtaining dichotomy and the substantial investment requirement exclude the sui generis right provided for under the EU Database Directive? Reflections and proposals’ [2020] SSRN Electronic Journal, 6-7.

<sup>160</sup> Diega (n 157) 118.

<sup>161</sup> *OPAP* (n 152), para 43.

<sup>162</sup> *OPAP* (n 152), para 44.

legislation and courts of the Member States to establish so far. It has generally been applied as a low threshold to pass, with examples of a few hours of works or by compiling a list of 1650 email addresses.<sup>163</sup>

If a maker of a database meets the substantial investment criteria, the maker is granted the right to ‘prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database’<sup>164</sup> whereas extraction is defined as a ‘permanent or temporary transfer of all or a substantial part of the contents of a database to another medium by any means or in any form’<sup>165</sup> and re-utilization means ‘any form of making available to the public all or a substantial part of the contents of a database by the distribution of copies, by renting, by on-line or other forms of transmission’<sup>166</sup>. The database maker also has the right to prevent repeated and systematic extraction and/or re-utilization of insubstantial parts of the contents of the database ‘which, by their cumulative effect, would lead to the reconstitution of the database as a whole or, at least, of a substantial part of it, without the authorisation of the maker’.<sup>167</sup> In determining what constitutes a substantial part of the database, such an assessment must be made in relation to the volume of contents of the entire database.<sup>168</sup> The database rights last for 15 years from the completion of making the database, and may be renewed for another 15 years every time a alteration of the database can be considered a substantial new investment.<sup>169</sup>

It should also be kept in mind that the protection offered to prevent circumvention of TPMs under the Infosoc Directive also apply in regard to the *sui generis* right.<sup>170</sup> See section 2.4.7 for an overview of TPMs.

---

<sup>163</sup> Burdese (n 159) 8.

<sup>164</sup> Article 7(1) Database Directive.

<sup>165</sup> Article 7(2)(a) Database Directive.

<sup>166</sup> Article 7(2)(b) Database Directive; also the distribution right is exhausted with the first sale of a copy of the database within the EU so as it is no longer possible to control further sales of that copy.

<sup>167</sup> Case C-304/07 *Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg* [2008] ECLI:EU:C:2008:552, para 27; Article 7(5) Database Directive.

<sup>168</sup> *British Horseracing* (n 158), para 70.

<sup>169</sup> Article 10 Database Directive.

<sup>170</sup> Article 6(3) Infosoc Directive.

The exceptions for text and data mining under the DSM Directive also applies to extraction and reproductions of lawfully available works that otherwise would be protected by the *sui generis* right.<sup>171</sup> For more on the text and data mining exceptions see section 2.4.6.

The maker of the database is explained in recital 41 of the Database Directive to be ‘the person who takes the initiative and the risk of investing; whereas this excludes subcontractors in particular from the definition of maker’. The maker may transfer, assign, or license the *sui generis* rights.<sup>172</sup>

If a maker of a database has made the database available to the public, lawful users of that database have the right of ‘extracting and/or re-utilizing insubstantial parts of its contents, evaluated qualitatively and/or quantitatively, for any purposes whatsoever.’ unless the user is only authorized to extract and/or reutilize parts of the database, in which case the right only extends to that part.<sup>173</sup> Any contractual provisions contrary to this right is null and void.<sup>174</sup> A lawful user may however ‘not perform acts which conflict with normal exploitation of the database or unreasonably prejudice the legitimate interests of the maker of the database’.<sup>175</sup>

## 2.5.4 Assessment

Starting with the question if digital databases for storage of data in general would meet the definition of database under the Database Directive. First, while there exists many different database architectures that might offer pros and cons depending on the needs, type of use and the type of *data*, in common between all types of databases is that they can store data, and later, upon a query, retrieve the stored data individually.<sup>176</sup>

---

<sup>171</sup> Article 3(1) and 4(1) DSM Directive.

<sup>172</sup> Article 7(3) Database Directive.

<sup>173</sup> Article 8(1) Database Directive.

<sup>174</sup> Article 15 Database Directive.

<sup>175</sup> Article 8(2) Database Directive.

<sup>176</sup> Sourav Mukherjee ‘The battle between NoSQL Databases and RDBMS’ [2019] SSRN Electronic Journal.



Recalling that a database is defined as a collection of independent materials that has been arranged in a systematic or methodical way and is individually accessible. Since a material is independent if it carries autonomous value judged from the perspective of any third party, data stored for the purpose of data analysis should meet this criterion because even if the value of a single material is low, and the true value only can be revealed through analysis such data at an aggregated level, every data record still carries autonomous informative value. That a database must consist of a collection of independent materials imply that the database must consist of more than one material. Since the point of a database is to store data for enabling retrieval at a later point, all digital databases should meet the criterion of the material being individually accessible.

As such it seems that almost any collection of data that has been digitally stored in a database would fall within the scope of being classified as a database under the Database Directive, unless the data itself carries no value whatsoever, or unless the database only has stored on a single material.

Next, I will assess data protection in a Big Data context from the perspective of copyright protection of original databases, followed by protection under the *sui generis* right.

As the copyright protection only has the scope of the arrangement/selection of the contents of the database direct protection of data is excluded. However, as the rights are construed, by having the right to prevent reproductions of such an arrangement/selection of contents, the contents could potentially be protected indirectly from reproduction as well in instances where a reproduction of the contents also reproduces at least a part of the arrangement/selection that qualifies as an intellectual creation.

Keeping in mind that 11 words can potentially be considered an intellectual creation, even quite small parts of the structure/selection of contents from a

database could possibly been seen as an intellectual creation. However, unlike a physical collection, the structure of a digital database has a hidden structure that is not reflected by reproducing the individual contents in the database. As such, it seems highly unlikely that the structure of a database would be protected from reproduction of the contents of the database.

In regard to the selection of the contents, at the stage when the data is first generated and stored, the selection would either be done through sensor readings and records thereof in case of machine-generated data, or data collected due to human interactions outside the control of the author. The influence for the author of a database at this stage thus seems to be limited to the choice of what type of content should be stored in the database. For example, a sensor in a device capable of reading the temperature and air humidity might store those reading in separate databases. To me it seems unlikely that only that the authors choice in type of content, but without influence over the actual content, would be enough as to qualify as an intellectual creation. However, at the data analysis stage in the case of supervised learning where training data is needed, or in the case of generative modelling where real data is needed, a database containing such a selection could probably be considered an original database where the selection of those contents are protected from, for example, reproductions.

Overall, the scope for copyright protection in databases for the purposes of protecting Big Data seems to be minimal.

Moving on to assessing protection of data under the *sui generis* right.

A main issue for machine-generated data would be in relation to fulfilling the criteria of substantial investment in obtaining the contents of the database. In the case of sensor data, the sensors themselves may require substantial investments, but that would most likely be regarded as an investment in the creation of data, rather than in obtaining the data at the time the data is first stored into a database. That would likely hold true even for human-generated

data, since the investments in the digital platform that records and stores the data, would most likely have been investments in the operation and functionality of the platform, rather than in obtaining the data. Likewise, when it comes to the data that is generated from the data analysis, this would also be creation rather than obtaining material since the analysis servers to find new patterns from the analysed dataset. However, even if data at those points are stored in databases that would not fall under the *sui generis* right it could still be that in a later phase of the Big Data process, that for example unstructured data needs to be further processed by an AI for it to be transformed into machine-readable structured data and then stored in a new database. It could also be that the collected data are stored in separate databases, but that at a later stage the contents of those databases are merged into a larger aggregated database. Such activities might require substantial investments in obtaining the data. Another way that has been suggested in acquiring substantial investment in obtaining created data could be to transfer the ownership of the data to a sister company in which case the fee could be a substantial investment in obtaining the data.<sup>177</sup>

Another way of acquiring data is through text and data mining, which could entail purchasing licenses, in order to get lawful access to databases or other sources of data. If data is acquired for data analysis through a purchase it opens the possibility of obtaining the data, even if the data was created in the database it was extracted from during the mining process.

In regard to substantial investments in verification it seems unlikely to require substantial investment. In regard to machine generated data it might be necessary to have some sort of verification process to sort out data resulting from malfunctions in the sensor readings. However, if that is the case and the size of investment that would entail is an empirical question.

---

<sup>177</sup> Burdese (n 159) 11.

When it comes to investments in processing the data, which refers to the capability of processing information of the database it seems hard to make any general assessments. While a database might be available for free as open source, certain data analysis project might require custom databases that require financial investment in purchasing it or its creation, or technical investments for the computer hardware to operate and store the database, or otherwise in human effort to create the database.

Even though there is opportunities to have a database in a Big Data context protected under the *sui generis* right, it has been pointed out that it is not necessarily always favourable. The reason for the *sui generis* to not necessarily benefit the database maker is that that lawful users of a *sui generis* database has the right to extract or reutilize insubstantial parts of the database and that contractual restrictions of this right is null and void. Therefore it could be favourable to protect to database fully from contractual provisions that restricts such uses.<sup>178</sup> Additionally, along the same line of reasoning, a database that falls outside the scope of a *sui generis* database also prevents extractions from the database based on the exceptions for text and data mining. However, with the text and data mining exceptions, the contractual override only refers to research organisations and cultural heritage institutions, while the general exception still allows to reserve the use against text and data mining.

Furthermore, as already established under the assessment under the Infosoc Directive, applied TPMs to the database would offer a strong second layer of defence due to the associated transaction costs. On this note, it could also be challenging to identify who is the maker of a database and thus the person that has to be contacted to see if they are willing to voluntarily remove the TPM that prevents utilization of an exception.

---

<sup>178</sup> Fischer (n 156) 22.

## 2.6 Protection of Big Data under the Trade Secret Directive

The Trade Secrets Directive<sup>179</sup> protects trade secrets against unlawful acquisition, use and disclosure. The directive harmonizes minimum levels of protection of trade secrets meaning that national legislation in the Member States may offer stronger protection than required by the directive.<sup>180</sup>

For information to fall within the scope of a trade secret three cumulative conditions must be fulfilled.

First, that ‘it is secret in the sense that it is not, as a body or in the precise configuration and assembly of its components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question’.<sup>181</sup>

Secondly, the information must have commercial value because it is secret.<sup>182</sup> It is explained in recital 14 that the value can be actual or potential. Recital 14 further exemplifies instances where information has commercial value, such as ‘where its unlawful acquisition, use or disclosure is likely to harm the interests of the person lawfully controlling it, in that it undermines that person's scientific and technical potential, business or financial interests, strategic positions or ability to compete.’

Thirdly, that the information ‘has been subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret’.<sup>183</sup> The directive lacks further clarification on what is considered to be reasonable steps. Until the concept has been clarified by the CJEU it remains

---

<sup>179</sup> Directive (EU) 2016/943 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure (Trade Secrets Directive)

<sup>180</sup> Article 1(1) Trade Secrets Directive.

<sup>181</sup> Article 2(1)(a) Trade Secrets Directive.

<sup>182</sup> Article 2(1)(b) Trade Secrets Directive.

<sup>183</sup> Article 2(1)(c) Trade Secrets Directive.

open for interpretation. However, that the steps should be reasonable under the circumstances suggests that it is a relative concept that must include an assessment of the context of the situation at hand.

The scope of a trade secret should, according to recital 14, 'cover know-how, business information and technological information where there is both a legitimate interest in keeping them confidential and a legitimate expectation that such confidentiality will be preserved' and exclude 'trivial information and the experience and skills gained by employees in the normal course of their employment, and also excludes information which is generally known among, or is readily accessible to, persons within the circles that normally deal with the kind of information in question'.

Trade secrets does not create any 'exclusive right to know-how or information'<sup>184</sup> and is regarded as an alternative or complement to intellectual property rights.<sup>185</sup> Trade secrets holders are granted rights to 'prevent, or obtain redress for, the unlawful acquisition, use or disclosure of their trade secret'.<sup>186</sup> In turn, a trade secret holder means any natural or legal person lawfully controlling a trade secret.<sup>187</sup> This suggest that all natural or legal persons controlling a trade secret, such as through a license, would be regarded as trade secret holders concurrently.<sup>188</sup>

A trade secret may be lawfully acquired if it is discovered or created independently which reflects the fact that trade secrets does not grant exclusive right to information or know-how.<sup>189</sup> Trade secret can also be lawfully acquired through reverse engineering by 'observation, study, disassembly or testing of a product or object that has been made available to

---

<sup>184</sup> Recital 16 Trade Secrets Directive.

<sup>185</sup> Recital 2 Trade Secrets Directive.

<sup>186</sup> Article 4(1) Trade Secrets Directive.

<sup>187</sup> Article 2(2) Trade Secrets Directive.

<sup>188</sup> Gert Würtenberger 'Protection of trade secrets and know-how in the European Union: the EU Trade Secrets Directive (EU) 2019/943' <<https://ip-iurisdictio.org/protection-of-trade-secrets-and-know-how-in-the-european-union-the-eu-trade-secrets-directive-eu-2019-943/>> accessed 25 May 2022.

<sup>189</sup> Article 3(1)(a) Trade Secrets Directive.

the public or that is lawfully in the possession of the acquirer of the information who is free from any legally valid duty to limit the acquisition of the trade secret'.<sup>190</sup> An aspect of legal uncertainty that has been brought up in relation to reverse engineering is that it grants lawful acquisition of a trade secret, but it is unclear whether such an lawful acquisition also would entail lawful use and disclosure of the trade secret.<sup>191</sup> Acquiring information through the exercise 'of the right of workers or workers' representatives to information and consultation in accordance with Union law and national laws and practices' is also considered lawful.<sup>192</sup> Trade secrets can furthermore be acquired lawfully through practices, that under the circumstances, conforms with honest commercial practices.<sup>193</sup>

Additionally, trade secrets may be lawfully acquired, used, or disclosed to the extent that is allow or required by Union or national law.<sup>194</sup>

On the other hand, in the case of unlawful acquisition, use or disclosure of a trade secret the trade secret holder shall have access measures, procedures and remedies under the national laws of the Member States to prevent, or obtain redress for, the unlawful acquisition, use or disclosure of their trade secret.<sup>195</sup>

Acquisition of a trade secret is unlawful when carried out through 'unauthorised access to, appropriation of, or copying of any documents, objects, materials, substances or electronic files, lawfully under the control of the trade secret holder, containing the trade secret or from which the trade secret can be deduced' or by any conduct contrary to honest commercial practices.<sup>196</sup> What constitutes honest commercial practices is not explained,

---

<sup>190</sup> Article 3(1)(b) Trade Secrets Directive.

<sup>191</sup> Tanya Aplin, 'The Limits of EU Trade Secret Protection' [2020] SSRN Electronic Journal, 12-14.

<sup>192</sup> Article 3(1)(c) Trade Secrets Directive.

<sup>193</sup> Article 3(1)(d) Trade Secrets Directive.

<sup>194</sup> Article 3(2) Trade Secrets Directive.

<sup>195</sup> Article 4(1) Trade Secrets Directive.

<sup>196</sup> Article 4(2)(a) and 4(2)(b) Trade Secrets Directive.

but in the TRIPS Agreement which the EU is bound by, and which also provides the rights implemented through the Trade Secrets Directive it is explained mean ‘at least practices such as breach of contract, breach of confidence and inducement to breach, and includes the acquisition of undisclosed information by third parties who knew, or were grossly negligent in failing to know, that such practices were involved in the acquisition’.<sup>197</sup>

The Trade Secrets Directive also provides for four exceptions to unlawful acquisitions, uses or disclosures: in exercising the charter right to freedom of expression and information as well as freedom and pluralism of the media; for revealing misconduct or illegal activity if carried out in the general public interest; disclosure by worker or worker representatives if part the legitimate exercise in accordance with Union or national law in so far as such a disclosure is necessary; and lastly, for protecting a legitimate interest recognized by Union or national law.<sup>198</sup>

### **2.6.1 Assessment**

Starting with an assessment of the requirement that a trade secret cannot be generally known or readily accessible to persons that normally deal with the type of information in question.

Some data will certainly be excluded from be regarded as trade secrets due to being generally known or readily accessible. Data for the purpose of selling, or due to the context of its use is publicly available would thus exclude protection as trade secrets. As for data consisting of empirical records of sensor measurements, while many of such measurements may be readily accessible to do at the time of their recording, if only one or a few readings was recorded, and they have not been made publicly available, it suggests that after their recording they might not be readily accessible any longer and thus open up the possibility of being considered a trade secret. Even though data

---

<sup>197</sup> Aplin (n 191) 16.

<sup>198</sup> Article 5 Trade Secrets Directive.



obtained for data analysis not necessarily might be able to be considered trade secrets due to being readily available, the new data gathered through data analysis might not necessarily be so readily accessible. It could also be the case, that the data analysed, and the applied data analysis would be readily accessible for person active on the circle of Big Data analysis depending on the complexity of the algorithm and how readily accessible the analysed data is. In the end it seems to boil down to quite complex *in casu* assessments whether data would be considered readily available.

Regarding that the information must have actual or potential commercial value, this criterion could potentially exclude protection for very small data sets since the value in analysis require a huge volume of data, which means data gains it's true value first on a aggregated level. In any case, very small data sets will likely in many cases be regarded as trivial information. Besides the issue that the data might have to reach a certain volume to carry commercial value, it is obvious that data at a volume that enable meaningful data analysis carries value through the insights that can be gained through the analysis. However, data in lesser volumes than can readily be analysed would carry, at least potentially, commercial value since a third party might want to acquire smaller volumes of data from several data holders for data analysis.

As for having to take reasonable steps, under the circumstances, to keep the information a secret, again seems to come down to an *in casu* assessment, but at a general level it seems safe to assume to non-disclosure agreements should be signed with employees accessing such information and with any third-party obtaining a license to use the information. From a technical level the data should most likely be kept stored in a secure computer behind firewalls, and the data could also be protected as a secret by TPM, such as encryption.

## **2.7 Contractual Protection**

Even if the data falls outside the scope of copyright protection, the *sui generis* database right, and as a trade secret it can still be possible to exert *de facto* control over the data. This can be due to for example being the only actor that

have recorded a certain data, or it could also be that new data have been obtained through data analysis. By not making the data publicly available means that the controller of such data has the possibility of entering into contracts and license or sell the data or the choice to keep it private. The data can be further protected by TPM. The freedom of contract would then enable to apply terms and conditions and limits the use of the data by a contracting party. EU law does not regulate contract law in general, so for the most part it would be a question of national law of the extent of the contractual freedom. However, even though it falls outside the scope of this thesis, it should be observed that EU law might interact with the contract law through for example competition law or sector-specific directives or regulations.

One advantage with data falling outside the scope of all the assessed protections is that it will be possible to control the data to a larger extent since the exceptions under the various protection schemes does not apply. On the other hand, having to tailor contracts the keeps control over the data might entail high transactions costs.

Even if the data is protected by copyright, as a trade secret or the *sui generis* right to a large extent it would still be needed to further control the data through contractual provisions, such as reserving the right to resell the data in the case of a transaction of data.

## **2.8 Concluding remarks**

As seen, data within a Big Data process struggles in most cases to fall within the scope of the assessed legal protections. Even when the possibility for legal protection exists, such protection is often surrounded by legal uncertainty which would entail complex legal assessments to evaluate whether the assessed data could fall within in the scope of some legal protection. The most realistic legal protection for possible utilization seems to be offered under the *sui generis* right. The reason being that making substantial investments in the creation of a database even if not occurring naturally such investments could be orchestrated through varying arrangements. As for protection of data as

trade secrets this would be a possibility in regard to non-public data as long as measures are undertaken to keep it secret.

However, given the complexity and legal uncertainty in the legal protection of data, the most practical way of protecting data seems to be to try to obtain de facto control over it and regulating terms and conditions for its use through contractual arrangements. This also carries the upside up not exposing the data to the exception and limitations otherwise available to protected data. The legal uncertainty does however not only relate to if some data is legally protected, but also to if an exception or limitation could be relied upon to use data. In most cases TPMs seem to offer one of the strongest protections of data, even in the case of works in relation to beneficiaries of exceptions due to the associated transaction costs of exercising that right.

Finally, to address the legal uncertainty in ownership in big data it would seem beneficial for if a legal act was enacted to set up a framework of data ownership rights to enhance the market through increased use and trade in big data.



# Bibliography

## BOOKS

Goldstein P and Hugenholtz PB, *International Copyright: Principles, Law, and Practice* (Oxford University Press 2019)

Hatzopoulos V, *The Collaborative Economy and EU Law* (Bloomsbury Publishing 2018)

Masouyé C, *Guide to the Berne Convention for the Protection of Literary and Artistic Works* (Paris Act, 1971) (WIPO 1978)

Pila J and Torremans P, *European Intellectual Property Law* (Oxford University Press 2019)

Richterich A, *The big data agenda: Data ethics and critical data studies* (University of Westminster Press 2018)

## ARTICLES

Aplin T, 'The Limits of EU Trade Secret Protection' [2020] SSRN Electronic Journal

Burdese P, 'AI-generated databases. Do the creation/obtaining dichotomy and the substantial investment requirement exclude the sui generis right provided for under the EU Database Directive? Reflections and proposals' [2020] SSRN Electronic Journal

Diega G, 'Artificial Intelligence and databases in the age of big machine data' [2019] SSRN Electronic Journal

Echoud M and others 'Harmonizing European Copyright Law: The Challenges of Better Law Making' (2012) Amsterdam Law School Research Paper No. 2012-07, Institute for Information Law Research Paper No. 2012-07, 156 <<http://ssrn.com/abstract=2049935>> accessed 25 May 2022

Gandomi A and Haider M, 'Beyond the Hype: Big Data Concepts, Methods, and Analytics' (2015) 35(2) International Journal of Information Management

Geiger C, Frosio G and Bulayenko O, 'Text and Data Mining: Articles 3 and 4 of the Directive 2019/790/EU' [2019] SSRN Electronic Journal

Giannopoulou A, 'Access and Reuse of Machine-Generated Data for Scientific Research' (2019) 12 Erasmus Law Review 155

Kitchin R, 'Big Data, New Epistemologies and Paradigm Shifts' (2014) 1 Big Data & Society

Mauro A, Greco M and Grimaldi M 'A formal Definition of Big Data based on its essential features' (2016) 65(3) Library Review

Mukherjee S, 'The battle between NoSQL Databases and RDBMS' [2019] SSRN Electronic Journal

Olshannikova E and others, "Conceptualizing Big Social Data" (2017) 4 Journal of Big Data.

So A, 'Technical Elements of Machine Learning for Intellectual Property Law' [2020] SSRN Electronic Journal

## **OTHER SOURCES**

Agrawal D and others 'Challenges and Opportunities with Big Data A community white paper developed by leading researchers across the United States' (2012)

Cattaneo G and others, 'The European Data Market Monitoring Tool: Key Facts & Figures, First Policy Conclusions, Data Landscape and Quantified Stories D2.19 Final Study Report' (2020 Publications Office of the European Union)

Commission, 'Public Consultation on Data Act and Amended Rules on the Legal Protection of Databases: Summary report on the public consultation'

Commission, 'Towards a thriving data-driven economy' COM (2014) 442 final

Commission, 'A Digital Single Market Strategy for Europe' COM (2015) 192 final 14

Commission, 'Building a European Data Economy' COM (2017) 9 final

Commission, 'Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act)' COM (2020) 767 final

Commission, 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence

(Artificial Intelligence Act) and Amending Certain Union Legislative Acts’  
COM(2021) 206 final

Commission, ‘Commission Staff Working Document Impact Assessment  
Report Accompanying the document Proposal for a Regulation of the  
European Parliament and of the Council on harmonised rules on fair access  
to and use of data (Data Act)’ SWD(2022) 34 final

Commission ‘Proposal for a regulation on harmonised rules on fair access to  
and use of data (Data Act)’ COM (2022) 68 final

Ménière Y and others, ‘Patents and the Fourth Industrial Revolution: The  
Global Technology Trends Enabling the Data-Driven Economy’ (EPO  
2020)

<[www.researchgate.net/publication/353637878\\_Patents\\_and\\_the\\_Fourth\\_Industrial\\_Revolution\\_The\\_global\\_technology\\_trends\\_enabling\\_the\\_data-driven\\_economy](http://www.researchgate.net/publication/353637878_Patents_and_the_Fourth_Industrial_Revolution_The_global_technology_trends_enabling_the_data-driven_economy)> accessed 25 May 2022

Duch-Brown N, Martens B and Mueller-Langer F, ‘The economics of  
ownership, access and trade in digital data; Digital Economy Working Paper  
2017-01’ (2017) JRC Technical Reports

Everis, ‘Study on data sharing between companies in Europe’ Final report

Favale M, ‘Technological Protection Measures and Copyright Exceptions in  
EU27: Towards the Harmonization’

<[https://law.depaul.edu/academics/centers-institutes-initiatives/center-for-intellectual-property-law-and-information-technology/programs/Documents/ipsc\\_2007/paper/Marcella\\_FavalePaper.pdf](https://law.depaul.edu/academics/centers-institutes-initiatives/center-for-intellectual-property-law-and-information-technology/programs/Documents/ipsc_2007/paper/Marcella_FavalePaper.pdf)>

Fischer R and others ‘Study in support of the evaluation of Directive  
96/9/EC on the legal protection of databases’ (2018)

Triaille JP, d’Argenteuil J, and Francquen A, ‘Study on the Legal  
Framework of Text and Data Mining (TDM)’ (2014)

Würtenberger G, ‘Protection of trade secrets and know-how in the European  
Union: the EU Trade Secrets Directive (EU) 2019/943’ <<https://ip-iridictio.org/protection-of-trade-secrets-and-know-how-in-the-european-union-the-eu-trade-secrets-directive-eu-2019-943/>> accessed 25 May 2022





# Table of Cases

‘Cofemel — Sociedade de Vestuário SA v G-Star Raw CV’ (C-683/17)

‘Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg’ (C-304/07)

‘Eva-Maria Painer v Standard VerlagsGmbH and others’ (C-145/10)

‘Fixtures Marketing Ltd v Organismos prognostikon agonon Podosfairou AE (OPAP)’ (C-444/02)

‘Football Association Premier League Ltd and others v QC Leisure and others and Karen Murphy v Media Protection Services Ltd’ (C-403/08 and C-429/08)

‘Football Dataco Ltd and others v Yahoo! UK Ltd and others’ (C-604/10)

‘Freistaat Bayern v Verlag Esterbauer GmbH’ (C-490/14)

‘Funke Medien NRW GmbH v Bundesrepublik Deutschland’ (C-469/17)

‘Infopaq International A/S v Danske Dagblades Forening’ (C-5/08)

‘Levola Hengelo BV v Smilde Foods BV’ (C-310/17)

‘Nederlands Uitgeversverbond, Groep Algemene Uitgevers v Tom Kabinet Internet BV, Tom Kabinet Holding BV, Tom Kabinet Uitgeverij BV’ (C-263/18)

‘Nils Svensson and others v Retriever Sverige AB’ (C-466/12)

‘Ryanair Ltd v PR Aviation BV’ (C-30/14)

‘SI, Brompton Bicycle Ltd v Chedech/Get2Get’ (C-833/18)

‘Sociedad General de Autores y Editores de España (SGAE) v Rafael Hoteles SA’ (C-306/05)

‘The British Horseracing Board Ltd and Others v William Hill Organization Ltd’ (C-203/02)

