Everything Counts in Large Amounts

Protection of big data under the Database Directive

Author: Martin Zeitlin
Supervisor: Professor Sanna Wolk
Summary

Everything Counts in Large Amounts refers to the important and growing role of data in the modern digital economy. This thesis assesses the protection of big data under the Database Directive. Big data databases consist of two main components; the contents, often in the form of raw data, and the big data analytics. What is characterised by big data is, apart from the distinguishing vast volume and dynamic elements of it, the way that analytics is applied to extract meaning. Notwithstanding the vast amount of investments made in big data and its outstanding potential, the protection of big data under the Database Directive has never been tried by the CJEU. In order for big data to be protected under the Directive, there are certain obstacles to overcome. Firstly, a big data database must meet the database definition, which it can with the help of big data analytics. Nevertheless, the big data analytics itself falls outside the scope of protection of the Directive. The Directive furthermore provides two layers of protection; copyright and sui generis right. Copyright is not possible for big data since it requires originality. Sui generis protection is possible provided that there has been investment in the obtaining, verification or presentation – a requirement that the CJEU has interpreted narrowly.

Under the sui generis right, there are two forms of infringements; extraction or re-utilisation of substantial or insubstantial parts of the database. The assessment of infringement in light of big data is similar to infringements in regular databases. Nevertheless, one striking difficulty is the definition of re-utilisation and the so-called Google-exception that causes insecurity regarding the scope of the concept of re-utilisation. In case big data cannot be protected, the CJEU has warranted contractual protection that can be more generous than the Directive permits and accordingly, instead of finding an infringement, it is possible to protect the databases through the contractual liability of its users. The conclusions of this thesis are essentially that big on a general level is eligible for protection under the Database Directive, provided that suitable analytics are applied and that the investments made in the database production can be distinguished from the creation of its contents. The importance of big data will increase further in the future and the role of the Database Directive might become more important than ever before.
# Table of Contents

**ABBREVIATIONS** ......................................................................................................................... 1

**1 INTRODUCTION** ....................................................................................................................... 2
  1.1 Background ................................................................................................................................. 2
  1.2 Aim and scope ............................................................................................................................. 3
  1.3 Approach .................................................................................................................................... 4
    1.3.1 Outline ................................................................................................................................. 4
    1.3.2 Method for the pursued aim ................................................................................................. 4
    1.3.3 Selection and use of material ............................................................................................... 5
    1.3.4 On the nature of directives ..................................................................................................... 7

**2 DEFINING BIG DATA** .................................................................................................................. 9
  2.1 Main characteristics ...................................................................................................................... 9
  2.2 The importance of big data analytics .......................................................................................... 11
  2.3 The use of big data ..................................................................................................................... 12

**3 THE NATURE OF THE DATABASE DIRECTIVE** ........................................................................ 16
  3.1 Legislative background ............................................................................................................... 16
  3.2 The purposes of the Directive .................................................................................................... 17

**4 THE DATABASE DEFINITION** .................................................................................................... 18
  4.1 Article 1 – defining the scope of the Database Directive ............................................................ 18
  4.2 Substantive requirements on the contents of databases .............................................................. 19
    4.2.1 Collections of independent works, data or other materials ............................................. 19
    4.2.2 Arranged in a systematic or methodical way .................................................................... 21
    4.2.3 Individually accessible ........................................................................................................ 22
  4.3 Can the contents of big data constitute a database? .................................................................... 23
    4.3.1 On a general level ................................................................................................................. 23
    4.3.2 As to the substantive requirements of the database definition ........................................ 23
  4.4 Big data analytics and the database definition .......................................................................... 24
    4.4.1 The exclusion of computer programs in the Directive .................................................... 24
    4.4.2 Is big data analytics excluded? .......................................................................................... 25

**5 COPYRIGHT PROTECTION** ........................................................................................................ 27
  5.1 The originality requirement ....................................................................................................... 27
5.2 Computer-generated works and originality ................................................. 28
5.3 Can big data be eligible for copyright protection? ...................................... 29

6 THE SUI GENERIS RIGHT ............................................................................. 31
6.1 What is this novelty? .................................................................................. 31
6.2 Qualitatively and/or quantitatively investment ........................................... 32
6.3 The investment as an object of protection .................................................. 33
   6.3.1 The meaning of substantial investment ............................................... 33
   6.3.2 The investments’ required connection to the making of databases ......... 34
   6.3.3 Investments in the obtaining ............................................................ 34
   6.3.4 Distinguishing between creation and obtaining .................................... 36
   6.3.5 Investments in the verification or presentation ...................................... 39
6.4 Can big data be protected by the sui generis right? ..................................... 41
   6.4.1 The critical issue .............................................................................. 41
   6.4.2 Investments in the obtaining ............................................................ 41
   6.4.3 Investments in the verification or presentation ...................................... 42

7 THE SCOPE OF PROTECTION ....................................................................... 44
7.1 Forms of infringements under the Directive .............................................. 44
   7.1.1 The legal framework ....................................................................... 44
   7.1.2 Defining extraction ....................................................................... 45
   7.1.3 The users’ right to consult a database .............................................. 46
   7.1.4 Defining re-utilisation ................................................................... 47
7.2 Big data and the forms of infringements .................................................... 51
7.3 Substantial parts of the contents ............................................................... 53
   7.3.1 Substantiality ............................................................................... 53
   7.3.2 Qualitative substantial parts .......................................................... 53
   7.3.3 Quantitative substantial parts ......................................................... 54
7.4 Substantiability and big data ..................................................................... 55
   7.4.1 Qualitatively or quantitatively substantial ........................................ 55
   7.4.2 The relation between the content and the analytics ......................... 56
7.5 Insustantial parts of the contents ............................................................... 56
   7.5.1 Interpreting Article 7(5) .................................................................. 56
   7.5.2 Insustantiality and big data ............................................................. 58

8 PROTECTING DATABASES WITHOUT SUI GENERIS PROTECTION ............ 59
8.1 The case of Ryanair .................................................................................. 59
8.1.1 A new intellectual property right created? .................................................. 59
8.1.2 Big data – better off alone? ........................................................................... 61

9 CONCLUDING DISCUSSION .................................................................................. 62

9.1 On the aim and method .................................................................................... 62
   9.1.1 Can big data be protected under the Directive? ............................................ 62
   9.1.2 Personal reflections on the aim and method ............................................... 63

9.2 Some prospective outlooks .............................................................................. 64
   9.2.1 Data – a priority in the European Digital Single Market ......................... 64
   9.2.2 Evaluation of the Database Directive ......................................................... 65
   9.2.3 Technology-driven change ......................................................................... 66

BIBLIOGRAPHY ......................................................................................................... 67

Articles ....................................................................................................................... 67
Internet sources ........................................................................................................ 69
Literature .................................................................................................................. 71
Official documents .................................................................................................. 73
   European Commission ......................................................................................... 73
   Other public authorities ....................................................................................... 74

TABLE OF CASES ..................................................................................................... 75

Court of Justice of the European Union ................................................................. 75
United Kingdom ..................................................................................................... 76
United States of America ......................................................................................... 76
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AG</td>
<td>Advocate General.</td>
</tr>
<tr>
<td>Berne Convention</td>
<td>The Berne Convention for the Protection of Literary and Artistic Works, 9 September 1886</td>
</tr>
<tr>
<td>CJEU</td>
<td>Court of Justice of the European Union.</td>
</tr>
<tr>
<td>Commission</td>
<td>The European Commission.</td>
</tr>
<tr>
<td>Ed./eds.</td>
<td>Editor/editors.</td>
</tr>
<tr>
<td>EIPR</td>
<td>European Intellectual Property Review.</td>
</tr>
<tr>
<td>EU</td>
<td>European Union.</td>
</tr>
<tr>
<td>IPQ</td>
<td>Intellectual Property Quarterly.</td>
</tr>
<tr>
<td>Para/paras</td>
<td>Paragraph/paragraphs.</td>
</tr>
<tr>
<td>S./Ss.</td>
<td>Section/Sections.</td>
</tr>
<tr>
<td>WCT</td>
<td>WIPO Copyright Treaty, 23 December 1996.</td>
</tr>
</tbody>
</table>
1 Introduction

The grabbing hands grab all they can.
All for themselves after all
it’s a competitive world.
Everything counts in large amounts.
– Depeche Mode 1983

1.1 Background

Depeche Mode could hardly grasp the accuracy of their lyrics when they in 1983 concluded that everything counts in large amounts. Even though they most likely did not attempt to make a prophecy regarding the oncoming development of the data-driven economy, their lyrics essentially tells story of it all in today’s digital economy. Today, we are experiencing an ever-growing data-driven economy creating completely new business models where large accumulations of data have become valuable commodities. The term ‘big data’ is virtually the whole concept of data economy incarnated since it can be compared to an actual vault where the vast amount of the commodity is stored. The European Commission has considered big data, cloud services and the Internet of Things to be central for Europe’s competitiveness in its 2015 strategy for the European Digital Single Market.1 Nevertheless, the main legal framework for protection of databases within the European Union has not changed since enacted in 1996.

The leader’s section of The Economist argued in 2017 that data in this digital era has become the equivalent of what oil once was. In support of this, it was argued that the world’s five most valuable listed companies, namely Alphabet (owner of Google), Amazon, Apple, Facebook and Microsoft, all use data as a vital part of their business model. The corporations are indeed lucrative and collectively saw a net profit totalling 25 billion USD in Q1 2017.2 In the same issue, The Economist further wrote:

Flows of data have created new infrastructure, new businesses, new monopolies, new politics and – crucially – new economics. Digital information is unlike any previous resource; it is extracted, refined, valued,

2 The Economist – The world’s most valuable resources is no longer oil, but data (06/05/2017).
bought and sold in different ways. It changes the rules for markets and it demands new approaches from regulators. Many a battle will be fought over who should own, and benefit from, data.³

It is evident that there is immense value in data and that its nature is completely different from other commodities. Using the words of Depeche Mode – in a competitive world where the grabbing hands grab all they can – there surely must be protection available to the owners of data. If there is, like the Economist wrote, an oncoming battle over who should own and benefit from data, new light will be shed on the available legal framework. These issues are far from legal certainties and even more so when putting the buzzword big data into the equation. These legal uncertainties constitute important issues given the vast amounts of capital being spent on buzzwords like big data, which to a significant extent can be seen as a driving force behind the modern digital economy.

The European Commission stated in 2017, on the topic of building a European data economy, that raw machine-generated machine generated data may be protected under the sui generis right of the Database Directive although it generally does not meet the required conditions. The Commission also concluded that there is a lack of policy framework for machine-generated content that does not qualify as general data and that the topic is largely left to contractual solutions.⁴ Not all big data is raw machine-generated, but a significant part is which will be further elaborated in this thesis. The statements of the Commission still leave a lot of uncertainty regarding how, and to what extent, big data can be protected under the Database Directive.

1.2 Aim and scope
Nobody has probably escaped the recent attention given to personal data and the strengthening of the rights of registered individuals. However, in the perspective of rights the owners of databases, it has been essentially quiet. The intellectual property protection of databases in Europe is regulated in the Database Directive introduced in 1996. There is a great interest in analysing the Directive based big data in order to see how it can be adopted today in the conditions of the digital economy.

This thesis aims to analyse how, and to what extent, big data can be protected within the EU under the Database Directive. Hence, the territorial scope is

³ The Economist – Data is giving rise to a new economy (06/05/2017).
limited to the European Union only. Accordingly, no emphasis will be given to the existing protection in other jurisdictions than the European nor any issues arising in light of international private law.

As to the substantive scope of this thesis, it is written completely from the perspective of the database makers and the protection of database content from infringement. No emphasis will therefore be given to the issue of personal data or ownership of data. Moreover, this thesis will only analyse the issue of eligibility for database protection and forms of infringements in light of big data databases. Issues concerning the determination of damages, licensing, ownership of data etc. will thus not be covered. Since the thesis only will analyse the Database Directive, no issues relating to other types of intellectual property rights, trade secrets, data protection, competition law or similar will be covered. There will finally be a concluding discussion about the future of protection of big data. The question of research can be summarised as follows:

I. Can big data be protected under the Database Directive and what are the major issues arising in terms of eligibility for protection?

II. How can the forms of infringements under the Database Directive be used to protect big data?

1.3 Approach

1.3.1 Outline

The thesis will begin with defining the concept of ‘big data’ based on technical terms in order to continuously use the definition for analysing its standing under the Database Directive. The structure of the thesis corresponds with the order of the articles in the Database Directive and each article is analysed in light of big data. Accordingly, following the definition of big data, the remaining chapters will give an introduction to the Database Directive and analyse the database definition, the copyright protection under the Directive, the sui generis right, the scope of the Directive’s protection and the protection of databases outside the scope of the Directive. Lastly, some concluding remarks will be given regarding the research question and future development.

1.3.2 Method for the pursued aim

The issues analysed in this thesis are to a large extent legally uncertain. Naturally, this poses a great challenge, in particular due to the ever-present risk of the thesis becoming
too speculative. As big data itself is not defined legally, a vast amount of emphasis will be vested in analysing the Database Directive itself de lege lata. The conclusions will on a continuous basis be used for guidance on how to assess big data in relation to the Directive. The method used is largely based on what in Sweden is commonly referred to as the legal dogmatic theory. The terminology is far from uniform and it is many legal traditions known as legal doctrine. Practitioners within the legal dogmatic theory aspire to reconstruct legal systems. In the words of Nils Jareborg; Lawmakers make the law, the judges judge, and legal dogmatic theory constructs a normative system that gives a meaning to it all. Compared to the field of law practice, the difference is, amongst other things, that the legal dogmatic theory uses many examples that are both actual and hypothetical, it seeks out new problems and may boldly propose new solutions to legal problems.

The application of the method to this thesis will appear in the continuous analysis where all foundation for analysis will be mostly explanatory as to what the law is, i.e. de lege lata. When the law itself lacks guidance on the issues, the analysis will be more of a normative character in the form if providing a recommendation to what the law should be, i.e. de lege ferenda. This will also be theme of the concluding remarks.

1.3.3 Selection and use of material
Given that big data is the basis for the whole thesis, it naturally needs to be defined in order to conduct any legal analysis. The notion of ‘big data’ is completely based on technical conditions and has not been conclusively decided by a court. Common sources used in informatics will therefore be used to provide a brief intelligible description of the main characteristics of big data.

The basis for all analysis will be legal sources commonly used for interpreting EU law. The prime source of interpretation in this thesis is clearly the Database Directive. Regarding directives, it can be noted that all official language versions of EU legislation are to be assessed equally. In case semantic discrepancies arise, the issue shall, without giving superiority to any specific language version, be resolved by a comparison of the different language versions. If that is without result, the individual

---

8 Ibid., p. 4.
articles shall be interpreted with regards to the general systematics and the purpose of the legal act it is part of.\footnote{Bergström & Hettne – *Introduktion till EU-rätten*, 2014, p. 385.}

The judgments of the CJEU will be the prime source of law besides the actual Database Directive. The standing of CJEU’s case-law is far more than a supplement to legislation since it as a rule is more detailed than the – generally speaking – broad definitions provided in directives. Through its interpretations, the CJEU has developed several legal principles that are not explicitly written in EU legislation with economic integration as one its leading purposes.\footnote{Ibid., p. 382.} The interpretation by the CJEU can in general terms be described as rather freely based and purpose-oriented.\footnote{Ibid., p. 389.} The CJEU rarely interprets law *e contrario*, whereas the use of analogies is more common, in particular for applying the overall purpose and systematics of the legislation at hand (*systematic-contextual interpretation*).\footnote{Ibid., p. 393.} Moreover, the CJEU often uses a teleological approach to interpretation where provisions are interpreted in light of common purposes, values or other pursued aims.\footnote{Ibid., p. 394.} The reasoning can at times be conceived as unprecedented – given the extensive use of principles – and can in that regard be considered a dream come true for creative litigators.\footnote{Reichel – *EU-rättslig metod*, in: eds. Korling & Zamboni – *Juridisk Metodlära*, 2013, p. 132.} As there is no case-law from the CJEU regarding big data, the earlier case-law based on the Database Directive and the principles thereof will serve as guidelines.

Another source of law used is the recitals of the Database Directive. The recitals of directives are not legally binding *per se*, nevertheless, they bear a substantial value for interpreting the articles of the directives.\footnote{Bergström & Hettne – *Introduktion till EU-rätten*, p. 392.} As to preparatory works, they do not bear the same interpretative value in EU law compared to the status it is traditionally given in the Scandinavian legal traditions. However, the CJEU has increasingly been referring to preparatory works of certain legislations.\footnote{Ibid., p. 61.}

The opinions of the Advocate General (hereafter referred to as AG) will moreover serve as an important source of law. The AG is integrated into the CJEU since its submission to the court are at many times required.\footnote{See Article 252 TFEU.}

\footnotesize

\footnotesize\textsuperscript{9} Bergström & Hettne – *Introduktion till EU-rätten*, 2014, p. 385.
\footnotesize\textsuperscript{10} Ibid., p. 382.
\footnotesize\textsuperscript{11} Ibid., p. 389.
\footnotesize\textsuperscript{12} Ibid., p. 393.
\footnotesize\textsuperscript{13} Ibid., p. 394.
\footnotesize\textsuperscript{15} Bergström & Hettne – *Introduktion till EU-rätten*, p. 392.
\footnotesize\textsuperscript{16} Ibid., p. 61.
\footnotesize\textsuperscript{17} See Article 252 TFEU.
expert nor an academic.\textsuperscript{18} The influence of the AG lies in the making of opinions that might influence the judges assigned to each respective case, although an AG does not take part in the deliberation process.\textsuperscript{19} The impacts of the AGs’ opinions can be difficult to assess given that the judgments of the CJEU tend to be rather brief. There is additionally a practice within the CJEU to not comment on novel points solely raised by the AG and not the parties.\textsuperscript{20} Consequently, the impact of the AG’s opinion is at many times not apparent from looking at the judgments solely. It is widely discussed what role the opinion of the AG should have when it is not acknowledged by the court. One common view is that when the court does not reach a different conclusion, the prior opinion of the AG carries interpretative value for the reasoning of the judges.\textsuperscript{21} Therefore, the opinions of AGs will in this thesis only be used to an extent where the statements are not contrary to the subsequent judgments of the CJEU.

Finally, legal literature will additionally be used as a source of law. Legal literature is never cited by the CJEU. Nonetheless, it is clear from looking at the reasoning of the AG that the legal literature is, at least indirectly, considered by the courts.\textsuperscript{22} Legal literature will be extensively used due to the scarce supply of judgments on the topic of this thesis.

\subsection*{1.3.4 On the nature of directives}

A directive is a normative act which comes with the requirement to be adopted by the necessary national provision in each Member State of the EU. Hence, a directive is ‘binding as to the result to be achieved’ and leaving it to the national authorities to choose the form and method.\textsuperscript{23} The reality can nevertheless be somewhat different where directives often regulate specific areas in great detail, particularly when there is a lot of technical aspects involved, thus leaving the Member States with less discretion.\textsuperscript{24}

As to the relationship between EU law and national legal systems, Member States are obliged to guarantee the effectiveness of EU law as established in Article 4(3)

\begin{thebibliography}{99}
\bibitem{Ibid.} Ibid., pp. 602, 607.
\bibitem{Turenne} Turenne – \textit{Advocate General’s Opinions or Separate Opinions: Judicial Engagement in the CJEU}, Cambridge Yearbook of European Legal Studies 14, 2012, p. 734.
\bibitem{Turenne2} Turenne – \textit{Advocate General’s Opinions or Separate Opinions: Judicial Engagement in the CJEU}, Cambridge Yearbook of European Legal Studies 14, 2012, p. 735.
\bibitem{Bergström} Bergström & Hettné – \textit{Introduktion till EU-rätten}, 2014, p. 63.
\bibitem{See} See Article 288 TFEU.
\end{thebibliography}
TEU. Directives cannot be applied between individuals (*direct horizontal effect*). There are exceptions to this rule, but the exceptions are immaterial to the legal issues presented in this thesis. Another important principle is the *indirect effect* (also known as EU-consistent interpretation). The principle obliges all public authorities of the Member States to interpret, as far as possible, all national law in light of and in conformity with EU law. In the case of *Pfeiffer*, the CJEU summarised earlier case-law and concluded that the principle is derived from the Member States’ obligation to achieve the results envisaged by directives. This is in particular relevant for interpreting provisions specifically adopted for the purpose of implementing a certain directive. The limitations of the indirect effect interpretative approach come with methods recognised in national law, general principles of law (in particular the one of legal certainty) and too strained interpretations of national law only to achieve consistency with EU law.

From a methodical perspective in this thesis, the characteristics of EU law come with some implications in terms of interpretation. National legal systems have their own autonomous law where the Directives are implemented in the for every jurisdiction suitable fashion. The EU’s main instrument for harmonising and maintaining compliance with directives within the Member States is the CJEU. Besides the activities of the CJEU, the national legal systems interpret the directives continuously. From the perspective of EU law, the implementation of the directives in the Member States is immaterial to the interpretation of directives on an EU level. Since this thesis only analyses the Database Directive itself, the supply of sources of interpretation will consequently be scarce compared to what can be expected in terms of the implementation of the Directive in the Member States. Still, the main instrument for interpreting EU law is the CJEU and when there is lack of case-law, the suitable way of assessing the autonomous meaning of EU law is to apply the law to an issue in a manner as similar to the CJEU as reasonably possible. Also, it is worth emphasising that the same hierarchy of legal sources that might be explicitly defined in national jurisdictions are not applied in the same manner by the CJEU.

---

25 See case C-91/92 *Paolo Faccini Dori v Recreb Sri.*, paras 23-25.
27 See case C-397/01 *Pfeiffer and Others*.
28 Ibid., paras 110-115.
2 Defining big data

2.1 Main characteristics

Databases are traditionally in informatics referred to as database systems when they encompass both actual content and any software controlling various aspects of data management.\(^{30}\) Raw data can be understood as data not accompanied by any information describing its nature that has not undergone any processing, either manually or through automated computer software. It is accordingly unstructured in its form.\(^{31}\) Another important term is dynamic database that is, contrary to a static database, a database that is constantly updated.\(^{32}\) As this chapter will show, it is important when mentioning big data to understand that there are two distinguishable components, the first constituted by the actual big data content and secondly the vital functions for data management and analysis, the so called ‘big data analytics’.

The term ‘big data’ was supposedly invented in the mid-nineties and recently has become a well-known buzzword within the digital economy. Attempting to define big data can easily give rise to confusion.\(^{33}\) That is because big data essentially does not have a definite definition. In concept, it involves an aggregated amount of data that is beyond the capacity of conventional database systems and the capacity is probably the most widely accepted trait of big data. Further definitions are difficult to make, however, there are some distinguished traits that are commonly attributed to big data databases.\(^{34}\) It is nevertheless important to emphasise that these technical definitions cannot compare to the precision and consistency that normally would be attributed to legal definitions and, in this regard, it must once again be stressed that there is no definition of big data provided in any case-law from the CJEU. Accordingly, the definitions below might be contradicted by technical definitions from other sources.

---


\(^{32}\) See cases C-203/02 BHB v Hill, Opinion of AG Stix-Hackl, paras 147-148; C-444/02 OPAP, Opinion of AG Six-Hackl, paras 168-169.

\(^{33}\) Gandomi & Haider – Beyond the hype: Big data concepts, methods, and analytics, International Journal of Information Management, 2015, p. 138.

\(^{34}\) Information Commissioner’s Office. Big data, artificial intelligence, machine learning and data protection, 2017, p. 6.
The major characteristics of big data can be described by using the three Vs; volume, velocity and variety.\textsuperscript{35} These characteristics are generally agreed upon by a large amount of data protection agencies around the world.\textsuperscript{36} Volume is arguably the most valuable aspect since it characterises the vast information that can be analysed for the creation of valuable outputs. Velocity means the pace of the flow of the data, which is commonly constant and rapid. Variety is the amount of different types of data accommodated in big data databases. Since the variety normally is immense, a great technical challenge lies within the structuring of the content in order to extract meaningful outputs.\textsuperscript{37} The existence of large unstructured amounts of data is by no means an innovation. The innovative feature that is worthy of the hype is the emergence of advanced data management technologies that enable the profitable utilisation of the unstructured data in business processes. The combination of the aspects of three V’s in big data and data management technology enables a real-time compilation of colossal amounts of data into useful and valuable information.\textsuperscript{38} It is this innovation that has transformed data into the valuable commodity it is considered to be today.

When it comes to the supply chain of big data databases, the first step is always the actual collection of the data. The sources could be everything from content created by users in social media, to news articles and weather forecasts. The actual input data can consist of nearly anything and would in the majority of cases be utterly incomprehensible to the human observer. Once the data is collected, it is normally stored in a completely unstructured form. The unstructured form of big data is a major difference compared to traditional databases. It is essentially only through the analysis that the informative nature of the data is revealed.\textsuperscript{39}

Notwithstanding the unstructured form and the variety of sources, the input data can be divided into three main categories. Data generated people to machine (P2M) is the result of humans interacting with machines, such as making a purchase online or visiting a website. Secondly, data can be the result of machine to machine interaction (M2M) which for instance can be geodata collected from a smartphone or the tracking of

\textsuperscript{36} Van Deer Sloot & van Schendel – International and comparative legal study on Big Data, 2016, p. 17.
\textsuperscript{38} Gandomi & Haider – Beyond the hype: Big data concepts, methods, and analytics, International Journal of Information Management, 2015, p. 138.
sensors. Finally, there is data generated by people to people (P2P) which can be interactions between people on social media. Furthermore, the input data is not always generated first-hand, but can be collected from multiple sources. It can be bought or licensed, created by users or be publicly available. As more devices are becoming connected to the internet, in particular through the development of Internet of Things, there will correspondingly be more sources of input data thus more overall generated data in the world.

The distinction between a traditional database and a big data database can be shown by a concrete non-computer-based example. Our brain is functioning through the interaction of billions of individual neurons. Singlehandedly, each individual neuron is worthless for processing information. A neuron can neither put a human on the moon nor even figure out how to open a can of soda. However, collectively, through a system of interactions so complex humans have yet to conclusively map its functioning, neurons enable humans to do all of these advanced tasks. The human brain itself is dynamic in its processing where the interaction can shift pathways for different tasks thus being able to process various forms of complex information. This can be compared to big data databases where individual materials are worthless, but put into interaction by big data analytics, can be used to process and produce extensive informative value. Big data, in its most advanced forms, can therefore be contrasted to a traditional database which can be compared to a shelf where different objects are merely just stored in, for instance an alphabetical order. Not all big data databases would fit in this analogy to the brain, but the most advanced forms can, which illustrates one of its most prominent characteristics.

2.2 The importance of big data analytics

The shift of paradigm that gave way to big data was most likely the point where the cost of storing data became less than the cost of erasing it. With the use of big data analytics, vast amounts of data have become valuable thus creating an incentive for storing data that certainly would have been erased otherwise.

In essence, big data analytics is the engine that enables the power of big data. A key point in understanding the value of data was concisely expressed by Gary

---

41 Information Commissioner’s Office. *Big data, artificial intelligence, machine learning and data protection*, 2017, p. 9.
42 Ibid., p. 41.
King (a professor of quantitative social science at Harvard University) when he stated that the ‘revolution’ is not about the data, but about the analytics that can be used to understand what the data says.\textsuperscript{43}

Traditional data analysis is, simplified speaking, about constructing a specific query to subtract specific information. The software then identifies the relevant entries for that specific query. Big data analytics, on the other hand, does not necessarily start with the use of a pre-constructed query. It can use a \textit{discovery phase} where it runs a large number of different algorithms against the data to discover far-flung correlations. When useful correlations have been discovered, it can be re-used by the new algorithms in the \textit{application phase}. The different phases can be differentiated in terms of ‘thinking with data’ and ‘acting with data’.\textsuperscript{44} The input data, i.e. the raw data, goes through processing operations using different platforms for analysis in big data analytics. The actual analysis of the aggregated data is conducted by advanced quantitative analysis software that is using different algorithms and progressively is becoming more advanced. It is also increasingly incorporating more advanced forms of artificial intelligence for optimising the results.\textsuperscript{45}

Without going into any further technical definitions, the key-point is understanding that big data cannot be properly described without mentioning the vital role of big data analytics. The analytics and the data are interdependent, i.e. they are both worthless without the existence of the other.

\subsection*{2.3 The use of big data}

With big data and the applied analytics, it is possible to analyse all available data regardless of the quantity. There is no longer a need to determine a statistically representative sample, which revolutionises the accuracy of statistics.\textsuperscript{46} Data is also increasingly being used for purposes other than the ones it was originally collected for. Different organisations can use the data for different purposes and data mining from

\begin{footnotesize}
\begin{itemize}
    \item The Washington Post – \textit{Gary King: Big data is not actually about the data} (20/05/2016).
    \item Information Commissioner’s Office. \textit{Big data, artificial intelligence, machine learning and data protection}, 2017, p. 10.
    \item Kemp IT Law. \textit{Legal Aspects of Managing Big Data}, 2014, p. 3.
    \item Information Commissioner’s Office. \textit{Big data, artificial intelligence, machine learning and data protection}, 2017, p. 11.
\end{itemize}
\end{footnotesize}
different databases can through analytics enable the discovery of new correlations and insights. This feature gives big data a wide scope of utilisation in many different sectors.

A fascinating example exposing the power of large amounts of data is the controversial company Cambridge Analytica. The company proclaimed to be able to “use data to change audience behaviour”. Somewhat provocative is the fact that the company not only provided its services for consumer marketing, but also tailored the services for political campaigns. Rather boastfully, Cambridge Analytica stated the following on their website:

“With up to 5,000 data points on over 230 million American voters, we build your custom target audience, then use this crucial information to engage, persuade, and motivate them to act.”

Cambridge Analytica became the subject of international attention after it was revealed that their services had been used by several political campaigns, mainly the British Brexit referendum and the American presidential election of 2016. Some of the company’s extensive data had been collected from surveys provided on Facebook that more than 100,000 people had completed. The surveys had been constructed as personality tests and its results were stored in a vast statistical database (ostensibly with more than 30 million profiles). The results of the personality tests could subsequently be correlated with data from a variety of online sources, hence being able to map online behaviour based on personality traits. Cambridge Analytica additionally bought personal data from multiple American sources, e.g. land registries, retailers, various membership clubs, religious congregations, that were willing to sell data. Such data was then, for the purposes of political campaigns, correlated with electoral rolls and analysed using psychometric profiling. Fundamentally, the company had optimised a method of turning digital traces and other data into profiles of real people. During the American presidential elections of 2016, approximately 175,000 ad variations for political arguments were allegedly tested on targeted audiences on social media in order to find the most effective ones for different

---

50 The Guardian – *Cambridge Analytica used data from Facebook and Politico to help Trump* (26/10/2017).
groups of people. Through the targeting, the campaign was able to tweak the ads to target certain people in, for instance, a limited geographical region being of a certain social class.\(^ {52} \) The business model of Cambridge Analytica is the understanding of specific individual’s attitudes and opinions in order to find the best way to engage with them.\(^ {53} \) It is worth mentioning that, during the course of writing this thesis, Facebook decided to suspend Cambridge Analytica’s access to the platform due to breach of terms and conditions.\(^ {54} \) Amidst the controversy, the company subsequently decided to shut down.\(^ {55} \)

The example of Cambridge Analytica illustrates the power of large amounts of data when combined with advanced analytics and that the utilisation of large amounts of data is far from always commercial in a traditional sense. Nonetheless, big data is mostly associated with its commercial potential. For example, big data enables effective forms of \textit{behavioural tracking}. One simple and common example of this phenomenon is when people are searching for a product online without making a purchase and the product subsequently reappears in ads whilst browsing other websites.\(^ {56} \) Moreover, smartphones that have Wi-Fi enabled are constantly searching for networks to connect to. When searching, the smartphones are transmitting signals that are detectable by nearby network routers. Each smartphone has a MAC address that works like an individual ID number. This frequent search for networks enables retailers to use network routers to track each individual smartphone’s movement in their stores. The logged data can then be used for very detailed retail analytics.\(^ {57} \) The data accumulated from behavioural tracking can be compiled and analysed on a major scale with the use of big data analytics. Correlated with data from other sources and behavioural tracking, the logged data can be used for powerful consumer insights.

It is worth mentioning that big data databases necessarily do not solely contain data that is attributable to individuals. Other examples of big data usage can be mapping world climate, predicting arrival times of public transportation through geospatial data and harvesting data from sensors on shipping containers in order to improve logistics.\(^ {58} \) Another example where big data is increasingly being utilised is in

\(^{52}\) Stanford University – \textit{The Data That Turned the World Upside Down} (28/01/2017).


\(^{54}\) Facebook Newsroom – \textit{Suspending Cambridge Analytica and SCL Group from Facebook} (16/03/2018).

\(^{55}\) Politico – \textit{Cambridge Analytica shutting down} (02/05/2018).


\(^{57}\) The Washington Post – \textit{How stores use your phone’s WiFi to track your shopping habits} (19/10/2013).

\(^{58}\) Information Commissioner’s Office. \textit{Big data, artificial intelligence, machine learning and data protection}, 2017, p. 13.
life science. Population studies, hospital collection and clinical studies can be merged into extensive biological collections that are analysed using big data analytics. New correlations can be found and used as starting points for further research.\textsuperscript{59} It is additionally being increasingly used to train different artificial intelligence services, which is a growing field that likely will increase the value of data even further. It can even be seen as artificial intelligence is a key to unlocking the true value of big data.\textsuperscript{60} With the increasing amounts of data produced in the world (that will only increase further fuelled by technological inventions like the Internet of Things and cloud computing), new areas where big data is implemented will most likely emerge.

The field of law practice is not spared from the impact of big data analytics. Advanced analytics trained through tremendous amounts of legal documents can be used for document analysis, i.e. ploughing through documents in order to identify those of importance and extracting key points from contracts.\textsuperscript{61} For instance, Lex Machina (a research company that develops legal analytics data and software) is allegedly able to, drawn from more than 100,000 past cases, automatically calculate the success of US patent litigation more accurately than litigators.\textsuperscript{62}

\textsuperscript{60} Information Commissioner’s Office. Big data, artificial intelligence, machine learning and data protection, 2017, p. 8; The Economist – Data is giving rise to a new economy (06/05/2017).
\textsuperscript{62} Ibid., p. 186.
3 The nature of the Database Directive

3.1 Legislative background

The Database Directive and the protection therein is unique for Europe. It was introduced to stimulate the growth of the database industry. In 1988, the total amount of electronic databases in the whole world had an estimated collective value amounting to 5 billion USD. Of that total value, the United States could account for 80%, whereas the European Union only could account for a mere 350 million USD. In an effort to nourish the European market, the European Commission evaluated the possibility of legislating some form of protection against unauthorised copying of databases. The national laws of the Member States varied considerably at that time. After a long legislative process with amended proposals, the Database Directive – introducing a new form of intellectual property – was finally adopted in 1996 by the EU. The Database Directive was commonly perceived as an exceptionally complex piece of legislation and only Austria, Germany, Sweden and the United Kingdom succeeded in complying with the deadline for implementation. The last country to implement the directive was Ireland as late as in 2001 – years after the deadline.

The Database Directive introduced two different types of protection for databases. The first is based on copyright and was drafted in light of international copyright treaties. Under international law, compilations of data are protected under Article 10(2) TRIPS Agreement which shall be interpreted in light of Article 2(5) Berne Convention. Therefore, the Berne Convention’s requirement of intellectual creation for eligibility to copyright protection also applies to the TRIPS provision. Compilations of data or other material that constitute intellectual creations are furthermore protected under Article 5 WCT. In consequence, the international treaties can be used for interpreting the copyright protection under the Database Directive. However, the other type of protection under the Directive, namely the sui generis right based on the investments made into the making of the database, is not covered by any international treaties.

65 See Article 9 TRIPS Agreement.
3.2 The purposes of the Directive

The Database Directive was introduced based on the recognition that the making of a database comes with a certain cost. As such, every database is the result of an investment which merits protection in order to stimulate the development of the information market in the EU. As early as in 1992, the Commission wrote in the Explanatory Memorandum to the draft of the Database Directive that information “is considered more and more as a tradeable commodity which is subject to economics of scale due to the increased cost of collecting, codifying, distributing relevant data on top of a considerable initial investment”. In light of the development of the modern digital economy, the predictions laid out by the Commission were indeed accurate.

The applicability of the Database directive is dependent on whether or not the database falls within the database definition in Article 1. Nonetheless, the applicability of the Directive does not equal to protection. The Database Directive guarantees two different kinds of protection, one in Article 3 (copyright protection) and one in Article 7 (sui generis right). Either one or both of the articles must be applicable in order for a database to benefit from protection. The different rights granted by the two articles give an overlapping protection where both copyright and sui generis protection can be applicable irrespective of the other’s existence. As copyright protection already existed in international treaties, the major difference emerging through the introduction of the Directive was the sui generis right. The right was created since the creative aspects needed for copyright protection excluded protection in databases where there had been a significant investment only. It is important to bear in mind that the sui generis right accordingly is a protection for the investments made in the creation of the database, which makes it completely independent from different types of copyright protection.

69 See Article 7(4) and Recital 58 of the Database Directive.
4 The database definition

4.1 Article 1 – defining the scope of the Database Directive

In order to assess big data under the Directive, it is naturally necessary to assess if, and depending on what conditions, it can fall within the scope of the Directive. Pursuant to Article 1, the Database Directive applies to: 1) databases in any form, 2) databases that are collections of independent works, data or other materials 3) arranged in a systematic or methodical way 4) and individually accessible and 5) not to computer programs used in the making or operation of databases. Article 1 may seem extensive at first glance given its rather extensive requirements.

Regardless of the fact that the Database Directive originates from 1996, there has been no conclusive determination of the definition of a database by European courts or legislators. It can be stressed, as a preliminary point, that there is accordingly a need for more modern case-law from the CJEU regarding the scope of the Directive. Nevertheless, this following chapter will explain the database definition of Article 1 based on the available case-law and other supplementary sources in order to apply the definition to the characteristics of big data.

The Database Directive applies to any form of database according to Article 1(1), which implies that the technical form of the database is immaterial. This is also confirmed by Recital 13 that states the protection of collections compilations by “means which include electronic, electromagnetic or electro-optical processes or analogous processes”. It is settled in the case-law of the CJEU that databases shall be defined “unencumbered by formal, technical or material nature”. It is furthermore expressed in the Explanatory Memorandum that the Directive seeks to “regulate the specific problems which arise as a result of the use of electronic data processing equipment for the storage, processing and retrieval of “information”, in the widest sense of that term”. Thus, it can be concluded as a general remark that the Directive purposely has a wide scope of application, which should be kept in mind at all times whilst interpreting its scope.

---

72 It can be emphasised in this regard that a database is by no means always accessible by electronic means only. Databases are compilations regardless of form.
73 See cases C-444/02 *OPAP*, para 20; C-30/14 *Ryanair*, para 33; C-490/14 *Verlag Esterbauer*, para 12.
4.2 Substantive requirements on the contents of databases

4.2.1 Collections of independent works, data or other materials

A database must be a collection of ‘independent works, data or other materials’ as defined in Article 1(2). It is evident from textual interpretation that the mentioned categories are alternate requirements. Amongst the categories, works is the most clearly defined and is supposedly to be understood as encompassing content meeting the requirement of originality in accordance with European copyright law. It is accordingly also rather trouble-free for categories of ‘works’ to meet the requirement of independence.

The notion of ‘data or other materials’ is more of a vague concept compared to ‘works’ and in particular to the requirement of independence. AG Stix-Hackl made an ambitious attempt to define data or other materials in the cases of Oy Veikkaus and OPAP. According to the AG, the requirement of independent data or other material means that the content of the database cannot be linked or must at least be capable of being separated without losing its informative content. The AG’s line of reasoning has subsequently been accepted by the CJEU. This might appear to be a very strict requirement, nonetheless, as will be accounted for below, there is a certain pragmatism to the reasoning.

The meaning of ‘data’ can also be somewhat ambiguous. In informatics, information is the semantic content that can be extracted from data or signals. AG Stix-Hackl seems to define ‘data’, within the meaning of being the content of an applicable database, to hold certain information. This is supported by the AG also defining data as “elementary statements with potentially informative content”. When it comes to ‘or other materials’, it is interpreted by AG Stix-Hackl as materials that can be considered “recognisable entities”. Still, it is unclear what materials actually fall within the definition. It is even more unclear from comparing the words used to describe ‘other materials’ in the different language version of the Database Directive. For instance, both the French version’s “éléments indépendants” and the German version’s “unabhängigen Elementen” refer to elements instead of materials. It is uncertain if this semantic

76 See cases C-46/02 Oy Veikkaus, Opinion of AG Stix-Hackl, para 36; C-444/02 OPAP, Opinion of AG Six-Hackl, para 39.
77 See cases C-444/02 OPAP, para 29; C-490/14 Verlag Esterbauer, para 17.
79 See cases C-46/02 Oy Veikkaus, Opinion of AG Stix-Hackl, para 33; C-444/02 OPAP, Opinion of AG Six-Hackl, para 36.
80 Ibid.
discrepancy has any significant implications for the interpretation and the discussion has largely been confined to the academia.\textsuperscript{81} AG Stix-Hackl has expressed that distinguishing content as ‘data’ or ‘other materials’ is not of big concern.\textsuperscript{82} The AG’s reasoning can be support by the fact the word ‘other’ is used, thus indicating a wide scope of definition covering all thinkable entities. In line with the overall objective to provide a broad definition of databases (as stated above in subchapter 4.1), the different forms of contents within a database should accordingly be immaterial to the scope of the Directive.

Certain pragmatism is appropriate in order to ensure a wide scope of application, which can be found in the reasoning of the CJEU. As earlier stated, the collections of works, data or other materials must be ‘independent’ in way that they are capable of being separated without losing their informative content. The CJEU has in this regard adapted a pragmatic approach by concluding that, notwithstanding that a single entity might not hold any informative value, several pieces of information can by accumulation constitute independent materials within the meaning of Article 1(2).\textsuperscript{83} Information shall furthermore be interpreted as holding independent value in a sense that it could provide, not just a typical user, but any third party with relevant information. This could be anyone interested in the extraction of the material.\textsuperscript{84} The CJEU’s conclusion settles that every material can satisfy the requirement of being independent as long as it singlehandedly, or by accumulating several entities, can provide informative value to any party.

Whether or not a certain proportion of materials must be independent has neither been stated in the Directive nor been settled by the CJEU. If one material is independent, it inherently implies that it is independent from other material hence necessitating that there has to be more than one material within the database. Nonetheless, as to the proportions, it has for instance been argued that more than half of the contents should be independent in order to fall within the database definition.\textsuperscript{85} It is an interesting aspect that a gross disproportionate amount of non-informative elements compared to the informative value might not result in the whole content being defined as a database. In

\begin{thebibliography}{99}
\bibitem{Stix-Hackl2002}See cases C-46/02 \textit{Oy Veikkaus}, Opinion of AG Stix-Hackl, para 33; C-444/02 \textit{OPAP}, Opinion of AG Six-Hackl, para 36.
\bibitem{Derclaye2005}See cases C-444/02 \textit{OPAP}, paras 33-36; C-604/10 \textit{Football Dataco and Others}, para 26; C-490/14 \textit{Verlag Esterbauer}, para 20.
\bibitem{Derclaye2005}See cases C-444/02 \textit{OPAP}, paras 34-36; C-490/14 \textit{Verlag Esterbauer}, para 27.
\end{thebibliography}
this regard, it is however worth emphasising that the database definition itself does not grant any protection to databases, which can only be granted through copyright or the sui generis right. Therefore, there should not be any need to define certain proportions of independent content since there is no question of balancing competing interests and, once again, the Directive shall purposely have a wide scope of application.

4.2.2 Arranged in a systematic or methodical way

Being arranged systematically or methodically excludes random accumulations of data and only planned collections of data fall within the databases definition. Completely unstructured accumulations of data accordingly fall outside the database definition and as a rule, the data must somehow be organised according to a specific criterion. This is supported by the fact that preparatory works of the Database Directive stated that “mere stockage of quantities of works or materials in electronic form” is excluded from the database definition. This indicates that collections of raw data in its unstructured form cannot be considered databases.

There is no need for the required arrangement to be physically apparent and the use of a method to retrieve data can also be used satisfy the requirement. Hence, if there is a software present that makes the data informative, there should no longer be a case of random accumulation at hand. AG Stix-Hackl has expressed that the requirement of being arranged in a systematic or methodical is sufficiently fulfilled with the application of a search engine that sorts and possibly indexes the data. Furthermore, not only static databases, but also dynamic, are covered by the Directive. All of this indicates a very generous scope of application for electronic databases.

As none of the above can be explicitly confirmed by case-law, it can definitely be scrutinised. Still, the arguments are undoubtedly in line with the overall objective of providing a wide definition of databases. The CJEU has emphasised, based on Recitals 12 and 10 of the Directive underlining the need for database protection for the development of the information market, that databases shall be given a definition with

---

90 See cases C-46/02 Oy Veikkaus, Opinion of AG Stix-Hackl, para 37; C-444/02 OPAP, Opinion of AG Stix-Hackl, para 40.
91 It can in this regard be noted that far from all databases covered by the database definition are electronic.
an emphasis on function over form. Also, the CJEU has not contradicted the arguments laid out by AG Stix-Hackl above. In conclusion, there is significant support that the requirement of being arranged in a systematic or methodical way can be satisfied in both static and dynamic databases merely by applying a search engine or some kind of organising function.

4.2.3 Individually accessible
The independent works, data or other material must finally also be ‘individually accessible’. The CJEU has concluded that the requirement means that each individual part of the content must be retrievable by some method or system. The requirement of being individually accessible is therefore closely linked to the previously mentioned requirement of independence. The requirement precludes, for instance, sole pixels in digital pictures from being part of database since they are not accessible by themselves. One way of satisfying the accessibility is to have the content searchable.

The requirement of being individually accessible can consequently be fulfilled directly when the user can browse the database and indirectly when the user can use a computer program or index to access the independent works, data or other material. In this regard, the requirement of the content being individually accessible is also from a practical point of view identical to the above-mentioned requirement of the content being systematically or methodically arranged. Method or system are in this regard interchangeable in a way that a database might be completely unstructured as long as there is a method of retrieving the content. A structured database in turn does not need a method in order to fall within the scope of application of the Directive. Following the same line of reasoning, unstructured content that is combined with an appropriate software is considered a database, just like with systematically or methodically arranged. The requirement of being individually accessible must also be satisfied at every given time. Therefore, unstructured data cannot be protected due to the fact that the database maker intends to apply a suitable software to it later onwards.

92 See case C-444/02 OPAP, paras 27-28.
93 Ibid., paras 32-33.
96 Ibid., pp. 105-106.
In summary, the necessity of having the requirement of being individually accessible can easily be questioned since it is truly difficult to imagine a situation where something fulfills the requirement of being independent materials and arranged in a systematic or methodical way without being individually accessible.

4.3 Can the contents of big data constitute a database?

4.3.1 On a general level

Since big data has no given definite definition, the obvious answer to whether or not it falls within the scope of the Database Directive is that it depends on the database at hand. Nevertheless, some deductions can be made based on the general traits of big data and the conclusions drawn from the conclusion above regarding the scope of the Database Directive.

Given that it is outstandingly clear that all forms of databases are covered by the Directive, the mere fact that big data is different from traditional databases is immaterial. The function over form approach taken by the CJEU and the aim to give the Directive a wide scope of application evidently supports that big data can fall within the scope, at least on a general level. It is furthermore immaterial to the database definition that big data databases may be dynamic – as opposed to static like traditional databases may be.

4.3.2 As to the substantive requirements of the database definition

When it comes to the requirement that the content of the database must constitute of independent works, data or other materials, it is evident that informative value must exist. Several pieces of information can by accumulation gain informative value thus constituting independent material. As AG Stick-Hackl has pointed out, the content can be “elementary statements with potentially informative content”.97 Quite intuitively, this is of big significance for big data since its content often is completely machine-generated and might, from the human eye, be utterly incomprehensible. Nevertheless, through the power of big data analytics, information can be extracted that is understandable and sorted. Hence, big data analytics can affect the database definition since accumulations of big data content might singlehandedly hold no informative value, but collectively does through the application of advanced forms of analytics – regardless of how unstructured.

97 See case C-46/02 Oy Veikkaus, Opinion of AG Stix-Hackl, para 33.
and incomprehensible the content is to the human beholder. Furthermore, there is no support in favour of the argument that a certain proportion of a database must hold informative value in order to qualify as a database. In this regard, it is worth noting that the informative value can be with reference to its value to any third-party. If a third-party is in possession of some state-of-the-art analytics software, it could consequently be argued that there is informative value to that third party. Therefore, the third-party’s access to analytics would determine the informative value of another database that itself is not even remotely connected to the analytics at hand. This might seem peculiar, however, using third-parties as a reference can be a way of further accommodating for technical innovation without having to change any legal definitions. As mentioned before, the overall aim of the database definition is to provide a wide scope of application.

Notwithstanding that the existence of analytics amongst third parties might help satisfying the requirement of independence, the big data must meet the other requirements of Article 1(2). In order to fulfil the requirement of being arranged systematically or methodically, the content of the big data database must not merely be randomly accumulated. Due to the largely pragmatic approach by the CJEU, this requirement should probably not be emphasised too much. It is simply sufficient that the material can be searchable with a search engine. Nonetheless, it does settle that raw data without the application of analytics is not covered. A big database only consisting of raw data with the intention to subsequently implement analytics cannot furthermore be considered a database. Thus, feeds of raw data can likewise not be considered databases.\(^98\)

In conclusion, the question whether big data can be considered a database is completely dependent on the use of analytics. Other than the points raised above, it is difficult to make a further analysis \textit{in abstracto} without any further case-law.

\subsection*{4.4 Big data analytics and the database definition}
\subsubsection*{4.4.1 The exclusion of computer programs in the Directive}
Article 1(3) states that the protection of the Directive shall not apply to ‘computer programs’ “used in the making or operation of a database accessible by electronic means”. Computer programs are also excluded from protection according to Article 2(a) of the Database Directive. The article stipulates, amongst other things, that the Directive shall not prejudice the legal protection of computer programs, which is further illuminated in

\footnote{98\ It could also be argued that raw data feeds are not considered big data since they normally lack the big data analytics.}
the recitals.\textsuperscript{99} It was implemented in order to secure priority to the Computer Programs Directive\textsuperscript{100} over the Database Directive.\textsuperscript{101}

The recitals state that the Directive “may also apply to the materials necessary for the operation or consultation of certain databases”\textsuperscript{102} and that one of its purposes is to protection “modern information storage and processing systems”.\textsuperscript{103} By comparing the articles and recitals above, it can be concluded that there is room for the interpretation that tools used for managing and analysing data might be protected provided that they are not deemed computer programs according to the definition provided in the Computer Programs Directive.

4.4.2 Is big data analytics excluded?

Big data analytics may be protected if it falls outside the scope of being a computer program. As mentioned above in subchapter 2.2, big data analytics is to a large extent dependant on algorithms. Algorithms cannot independently be protected by the Computer Programs Directive.\textsuperscript{104} However, they may be protected once they are inserted in a program and instructed to execute certain tasks.\textsuperscript{105} The protection of computer programs under the Computer Programs Directive is outside the scope of this essay, nonetheless, it can be presumed that there are situations where big data analytics are not protected as computer programs. It has been argued that the protection offered in the Database Directive can be overlapping with other rights regarding different aspects of database systems, which would make it immaterial whether the analytics is considered a computer program or not. If, for instance, a system is used to retrieve independent materials in a database, then the aspect of retrieval of independent materials could accordingly fall within the database definition.\textsuperscript{106} This reasoning has not explicitly been supported by the CJEU. Nevertheless, the overall purpose of the directive can be used in favour of also including big data analytics in the aspects attributable to ”materials necessary for the operation or consultation of database”. From looking at the wording of Articles 1(3) and

\textsuperscript{99} See Recital 23 of the Database Directive.
\textsuperscript{101} Virtanen – Database rights in safe European home: the path to more rigorous protection of information, 2005, pp. 133-134.
\textsuperscript{102} See Recital 20 of the Database Directive.
\textsuperscript{103} See Recital 12 of the Database Directive.
\textsuperscript{104} See case C-406/10 SAS Institute, para 32.
\textsuperscript{105} Wolk – Datorprogramsalster i upphovsrätten, 2016, p. 51.
\textsuperscript{106} Axhamn – Databasskydd, 2017, pp. 110-111.
2(a), and in particular the expression “used in the making or operation of a database”, the reasoning does not seem to be contrary to the exclusion of computer programs. On the other hand, there must have been a clear intention to exclude computer programs from the Directive by explicitly stating it in both Articles 1(3) and 2(a). If a particular computer program or similar falling within the substantive scope of the Computer Programs Directive is not be eligible for protection therein, for instance due to lack of originality, it would be remarkable to protect it under the Database Directive instead. This would effectively give rise to a completely new kind of intellectual property right specifically for database systems. There are no indications of any warranting of the establishment of any such right in the legislative procedures prior to the implementation of the Database Directive or by the CJEU. Accordingly, there are insufficient support for including database systems or big data analytics under the database definition, regardless of the fact that the definition itself does not grant any actual rights.

In conclusion, notwithstanding the fundamental importance of big data analytics and its necessity for enabling big data contents to fall under the scope of the Directive, it seems like the analytics itself does not fall within the scope. Nevertheless, it would surely be possible for the CJEU to interpret the Database Directive in a way that it would fall within the scope.
5 Copyright protection

5.1 The originality requirement

The copyright protection of the Database Directive is determined by Article 3(1). Eligibility for protection is determined by “reason of the selection or arrangement of their content”. The recitals further clarify that the only criteria used to determine whether the selection or arrangement can be eligible for copyright is the ‘author’s own intellectual creation’, i.e. originality. No other criteria than originality can be applied, which both Article 3(1) and the recitals state.\textsuperscript{107} This has also been explicitly confirmed by the CJEU.\textsuperscript{108}

The originality requirement is fully harmonised in the EU following the landmark case of \textit{Infopaq}\textsuperscript{109} and the requirements established therein are likewise applicable to databases.\textsuperscript{110} The intellectual effort and skill used in the creation of the material that is part of content of the database is notwithstanding to the originality of the database as such. Consequently, the fact that the content of a database is constituted by works that themselves can enjoy copyright protection does not affect whether the database itself can be protected.\textsuperscript{111}

It is important to understand originality in light of copyright history where photography is an illustrative example. In the 1800s, the United States Supreme Court concluded in a widely renowned case that a photography of Oscar Wilde was deemed to be a creation by a human and could thus be protected by copyright. The ruling was immensely controversial and even the inventor of photography opposed the Supreme Court’s conclusion. On the same topic, it was not until 1965 that Germany legislated full copyright for photographic works.\textsuperscript{112} This can serve as an example of the frequently controversial role of technology in the debate regarding what can be attributed to human creativity.

The originality requirement in Europe has been developed through a clash between the \textit{droit d’auteur} (author’s right), historically prominent in continental Europe,

\textsuperscript{107} See Recitals 15-16 of the Database Directive.
\textsuperscript{108} See case C-604/10 \textit{Football Dataco and Others}, para 42.
\textsuperscript{109} Case C-5/98 \textit{Infopaq}.
\textsuperscript{110} See case C-604/10 \textit{Football Dataco and Others}, para 38.
\textsuperscript{111} Ibid., para 33.
and the common law tradition. Droit d’auteur has been significantly more prone to only grant flesh-and-blood authors copyright than what the common-law tradition has been.113

5.2 Computer-generated works and originality

Originality is a reflection of the author’s personality which necessitates that the work itself consists of an individual expression originating from the author.114 In the words of the CJEU, the work itself must bear the author’s ‘personal touch’.115 The notion of ‘author’s own intellectual creation’ furthermore implies that there has to be a human taking part in the creative process.116 The CJEU has also concluded that the artist must be able to “express his creative abilities in the production by making free and creative choices”.117

Regardless of the fact that copyright historically has evolved to protect subject matter other than it was originally intended to, originality seems to require human creation, thus excluding all forms of computer-generated works from copyright protection. This has never been tried by the CJEU, nevertheless, there is a wide consensus in the legal literature that computer-generated works are excluded from copyright, which also has been settled by the Supreme Court of the Netherlands.118 Additionally, it can be noted that the American Copyright Office (the public authority that administers copyright applications in the United States) explicitly stipulates that works have to be created by humans in order to enjoy copyright.119

When it comes to partly computer-generated works, i.e. where only a certain part of the work created by a human, the originality requirement is more difficult to apply. According to the case of Infopaq, certain parts of a work can enjoy copyright “provided that they contain elements which are the expression of the intellectual creation of the

115 See cases C-145/10 Painer, para 92; C-604/10 Football Dataco and Others, para 38.
117 See case C-145/10 Painer, para 89.
119 United States Copyright Office. *Compendium of US Copyright Office Practices 3rd ed.*, 2017, s. 306. N.B. The United States is obviously not a member of the EU, but copyright is largely based on the same international treaties. Therefore, the Copyright Office can be given some interpretative value as to what originality means since it is a requirement in international copyright law.
author of the work”.\textsuperscript{120} Likewise, the CJEU has ruled that a software can be deconstructed to components that are completely dictated by their technical function and components with creative value. In those cases, only the components with creative value can meet the requirement of originality, hence falling within the scope of copyright. The rest of software consequently falls outside the scope of protection.\textsuperscript{121} The reasoning can be interpreted as an exclusion of the possibility to merge a work dictated by technical function with something of creative value in order to enjoy copyright for the work as a whole entity.\textsuperscript{122}

\textbf{5.3 Can big data be eligible for copyright protection?}

A human can, with the massive amounts of work hours, indeed create databases with massive quantities of content. Big data does however not only concern the volume. Other traits (as mentioned in above in subchapter 2.1) include velocity and variety. Given the velocity of data and its major variety, humans cannot reasonably create big data databases without any computer-generation. It is simply beyond human capacity. Therefore, the copyright protection based on reason of the selection or arrangement of their content for databases conducted by humans cannot be applicable to big data databases. Notwithstanding that the field of intellectual property law is ever-evolving, the legal sources are unequivocally pointing towards the conclusion that big data is excluded from copyright protection since it inherently requires machine-generated creation.

The output of the data that has been processed by big data analytics can nevertheless itself be arranged by a human. Even though the content itself will not be considered copyright works, it is possible to select and compile the output in a way that amounts to originality. In that case, only the parts where the human has made free creative choices can be protected. The rest would fall outside the scope of protection. This interaction between humans and computers in a creative process does indeed show the challenges of intellectual property rights in the digital world. Furthermore, the compilation made by a human cannot be considered big data according to its technical definition, but more similar to either a conventional database or literary work based on the results of big data analytics.

\textsuperscript{120} See case C-5/98 \textit{Infopaq}, para 39.
\textsuperscript{121} See case C-393/09 \textit{Bezpečnostní softwarová asociace}, paras 48–49.
In conclusion, big data contents cannot be protected by copyright. As to the analytics, they may be created by humans and can accordingly satisfy originality requirement. Nonetheless, as earlier pointed out (see subchapter 4.4.2 above), the software used for arranging the content of a database cannot likely be protected under the Database Directive.
6 The sui generis right

6.1 What is this novelty?
The copyright protection in Article 3(1) and sui generis right in Article 7(1) are two independent rights with their own objects and conditions. It is evident from the wording of Article 7(1) that the notion of substantial investment is the central prerequisite for the sui generis protection. Contrary to copyright that only recognises the author’s own intellectual creation, the central objective of the sui generis right is to protect the investment. The sui generis right can accordingly be understood as contrasted to copyright.

Investments as a basis for protection is similar to the *sweat of the brow* doctrine formerly prevalent in the common law tradition for determining copyright protection. According to the doctrine, the law is supposed the enable the author to enjoy the fruits of his labour without any requirement of certain creativity. Thus, basically mechanical and automatic creations could be eligible for protection according to the doctrine. The sweat of the brow doctrine suffered a striking setback when the United States Supreme Court concluded in the case of *Feist* that only databases with sufficient originality shall be eligible for copyright protection. The ruling of the court resulted in heavy aggravation amongst database producer and the introduction of sui generis right in EU is by some understood as a counter-reaction to the American ruling, even though its factual influence is questionable. The sui generis right protects databases where there has been 1) a qualitatively and/or quantitatively 2) substantial investment in 3) either the obtaining, verification or presentation of the contents of that database. Naturally, the scope of protection in accordance with Article 7(1) will be narrower than the database definition in Article 1.

123 See case C-604/10 Football Dataco and Others, para 27.
128 See cases C-46/02 *Oy Veikkaus*, Opinion of AG Stix-Hackl, para 42; C-203/02 *BHR v Hill*, Opinion of AG Stix-Hackl, para 35; C-338/02 *Svenska Spel*, Opinion of AG Stix-Hackl, para 32; C-444/02 *OPAP*, Opinion of AG Stix-Hackl, para 48.
Given that it is not possible for big data databases to be protected under copyright (see chapter 5 above), the sui generis right is accordingly the only protection that big data may be eligible for under the Database Directive. As it is the investments themselves that count, the fact that the content of big data is machine-generated is completely immaterial in regard to the sui generis right. As it will be shown in this chapter, the understanding of how the investment must relate to the creation of the database is a key-point for understanding the sui generis right.

6.2 Qualitatively and/or quantitatively investment

The investment mentioned in Article 7(1) may be substantial either in quantitative or qualitative terms. The CJEU has concluded, derived from Recitals 7, 39 and 40 of the Database Directive, that investments in the creation of a database may consist of development of human, financial or technical resources. The notion of ‘quantitative’ refers to quantifiable resources and ‘qualitative’ to efforts which cannot be quantified, such as intellectual effort or energy.\(^\text{129}\) The broad spectrum of covered investments suggests that lawmakers intended to ensure a broad interpretation of possible investments.\(^\text{130}\) The term ‘quantitative’ has not been the subject of much discussion although it should simply just refer to a large quantity of the database’s elements.\(^\text{131}\) A ‘qualitative’ investment can subsist in the choice of information and the way it is presented. It can most likely be assessed similarly to the originality requirement in copyright.\(^\text{132}\) The protection is only granted to investments used to draw up and compile the contents of a database. Thus, the investment must themselves be correlated to the activities connected with the actual making of the specific database.\(^\text{133}\) It is worth noting in this regard that the directive does not distinguish between any forms of direct or indirect investments, i.e. to what degree the investments must relate to the actual making of the database at hand. However, given that that the Database Directive’s purpose is to protect investments in databases as such, and not to establish a general intellectual property, it can be argued that such definition would be superfluous since it is obvious

\(^{129}\) See cases C-46/02 Oy Veikkaus, para 38; C-338/02 Svenska Spel, para 28; C-444/02 OPAP, para 44.


\(^{131}\) Ibid., p. 22.

\(^{132}\) Ibid.

\(^{133}\) See cases C-46/02 Oy Veikkaus, Opinion of AG Stix-Hackl, paras 45-46; C-338/02 Svenska Spel, Opinion of AG Stix-Hackl, paras 35-36; C-444/02 OPAP, Opinion of AG Stix-Hackl, paras 51-52.
that the investments must relate to the making of the specific database that is the subject of protection.\textsuperscript{134}

As to the possible implications for big data, it can be noted that if ‘qualitatively’ is to be understood as closely related to requirement of originality in copyright, protection of big data based on the qualitative investments made would consequently be difficult. Nonetheless, it is not clear that qualitative should be understood completely the same way. Nevertheless, if only a significant part of what is considered to amount to ‘author’s own intellectual creation’ is to be understood as qualitative, such investment would virtually be impossible when it comes big data. It is furthermore not clear to what extent it is possible to use tools, qualitatively speaking, in the creation process or in the analytics. In relation to protecting big data, quantitative investment is accordingly more suitable since all the financial resources and time spent can be accounted for.

6.3 The investment as an object of protection
6.3.1 The meaning of substantial investment

AG Stix-Hackl has concluded, based on Recital 19 of the Database Directive stating that music playlists on a CD-ROM as a rule cannot amount to an investment that is substantial enough, that the notion of ‘substantial’ is to be understood as a certain threshold. The threshold should partly be absolute albeit set at a very low level in order to avoid undermining the objectives of the Database Directive.\textsuperscript{135} One of the most prevalent arguments for setting a high level for investment in order to be substantial is to prevent the formation of information monopolies.\textsuperscript{136} No such intention to set a high level has however been shown by the CJEU.

There is moreover a relative side of ‘substantial’ in relation to the actual costs and to the scale, nature and contents of the database and the sector that it belongs to.\textsuperscript{137} Therefore, the meaning of substantial should be interpreted as a relative requirement depending on the situation in casu, but with a general minimum threshold. That general

\textsuperscript{135} See cases C-46/02 \textit{Oy Veikkaus}, Opinion of AG Stix-Hackl, para 49; C-338/02 \textit{Svenska Spel}, Opinion of AG Stix-Hackl, para 39; C-444/02 \textit{OPAP}, Opinion of AG Stix-Hackl, para 55.
\textsuperscript{137} See cases C-46/02 \textit{Oy Veikkaus}, Opinion of AG Stix-Hackl, para 48; C-338/02 \textit{Svenska Spel}, Opinion of AG Stix-Hackl, para 38; C-444/02 \textit{OPAP}, Opinion of AG Stix-Hackl, para 54.
minimum threshold has (yet) not been decided by the CJEU and further definitions are up to Member States to determine.138

6.3.2 The investments’ required connection to the making of databases
As mentioned above, the substantial investments must be related to the actual making of the particular database. Article 7(1) explicitly mentions that there must have been investment in either the obtaining, verification or presentation of the content. This list of explicitly mentioned categories of investments is exhaustive according to the CJEU.139 The investments in the different categories do not individually have to amount to a substantial investment and it is sufficient that they collectively can amount to it.140 There was a widespread fear that the Database Directive could give rise to a monopolisation of information. A vast amount of that fear disappeared however when the CJEU interpreted the sui generis right for the first time by delivering four Grand Chamber judgments on the same day in 2004. As will be shown below, the CJEU settled a broad definition of a database, but interpreted the type of investments that could entail sui generis protection very narrowly.141

6.3.3 Investments in the obtaining
The requirement of substantial investment in the obtaining of the contents in Article 7(1) aims at protecting the resources used for the creation of the database as such. The purpose of the sui generis right is to promote the establishment of storage and processing systems for existing information.142 As explicitly put by the CJEU, the purpose is not to protect the creation of materials capable of being subsequently collected in a database. Therefore, the meaning of ‘obtaining’ is the resources used to search for and collection of ‘existing independent materials’ and not the creation of such material.143 AG Stix-Hackl has further concluded, based on a comparative analysis of the different language versions, that

139 See cases C-46/02 Oy Veikkaus, para 57; C-338/02 Svenska Spel, para 22; C-444/02 OPAP, para 30.
142 See Recital 12 of the Database Directive.
143 See cases C-46/02 Oy Veikkaus, paras 34-36; C-203/02 BHB v Hill, paras 31-33; C-338/02 Svenska Spel, paras 24-26; C-444/02 OPAP, paras 40-42.
‘obtaining’ cannot be interpreted as the mere production of data.\textsuperscript{144} It has been argued in legal doctrine that the strict interpretation of ‘obtaining’ seeks to reduce the need to introduce compulsory licensing that otherwise would be needed for the purpose of countering the formation of information monopolies.\textsuperscript{145}

Regardless of the purpose of the CJEU’s reasoning, it cannot be stressed enough that ‘obtaining’ never shall be confused with ‘creation’ and, accordingly, all investments made in the creation of data is immaterial to the protection of the database. The problematic distinction can be completely avoided by collecting data from other external sources. For instance, data that is official and publicly accessible can still be protected given that there has been a substantial investment.\textsuperscript{146}

The Directive does moreover not require that the content has been provided by the database producer itself.\textsuperscript{147} User-generated data, i.e. data that is not inserted into the database by the database maker, can accordingly be protected, provided that there has been a substantial investment in the function or system that enables the users to insert the data. In this situation, the creation is clearly made by the user and the obtaining by the database maker. Consequently, the distinction between the obtaining and creation can easily be made. In practice, the distinction can nevertheless be more difficult to apply. An upload of some kind by any user will in most cases inescapably encompass more generation of data than what is visible at first glance. For example, if someone posts a picture on social media, the upload of the picture constitutes data generated through \textit{P2M} (people to machine) interaction. Besides the picture, the upload also generates related data, e.g. to facilitate the transfer, establish log file entries etc. That data is completely generated by the user’s device and its interaction with the database, i.e. \textit{M2M} (machine to machine). This illustrates the fact that although content might be inserted by another party than the database maker, the digital technology blurs the lines between what is generated by the user and the database maker.

Something that is additionally problematic is the situations where the creation of the information coincides with the creation of the database. In other words, the database maker might possibly also be the maker of its content. In this regard, the

\begin{itemize}
\item \textsuperscript{144} See cases C-46/02 \textit{Oy Veikkaus}, Opinion of AG Stix-Hackl, para 66; C-203/02 \textit{BHB v Hill}, Opinion of AG Stix-Hackl, para 46; C-338/02 \textit{Svenska Spel}, Opinion of AG Stix-Hackl, para 56; C-444/02 \textit{OPAP}, Opinion of AG Stix-Hackl, para 72.
\item \textsuperscript{145} Davison & Hugenholtz – \textit{Football fixtures, horse races and spin-offs: the ECJ domesticates the database right}, EIPR 3, 2005, p. 5.
\item \textsuperscript{146} See cases C-46/02 \textit{Oy Veikkaus}, paras 32-38; C-547/07 \textit{Apis-Hristovich}, para 72.
\item \textsuperscript{147} Axhamn – \textit{Databasskydd}, 2017, p. 247.
\end{itemize}
CJEU has concluded that if the creation of a database is linked to a principal activity where the database maker also creates the content, it does not necessarily preclude protection. However, the investment itself must be independently made in the obtaining, verification or presentation of the content, i.e. distinguished from to the making of the content.\(^{148}\) In other words, it is immaterial if the database maker is also a collector and a producer of the content, as long as there has been a substantial investment that can be deemed independent from the creation. This can further be clarified with a profoundly British illustrative example stated in a ruling of the England and Wales Court of Appeal:

Suppose a scholar created a database consisting of all Charles Dickens' references to law and lawyers. That would involve expenditure of significant resources and the database would qualify for protection. Now suppose the scholar added a commentary to the entries. Would he lose protection for his database? Obviously not.\(^ {149}\)

### 6.3.4 Distinguishing between creation and obtaining

Distinguishing between creation and either the obtaining, verification or presentation can be a complicated task, especially since the distinction can be artificial and of metaphysical proportions. For instance, if the notion of creation is interpreted very narrowly, even the reading of a recording might constitute a kind of creation.\(^ {150}\)

In the case of *BHB v Hill*, the CJEU determined that the investments made in the course of horseracing, where lists concerning the races were drawn up and subsequently inserted in databases, could not be deemed substantial investments in the database as such. That is because the investment, i.e. the drawing up of lists, was not attributable to the making of the database.\(^ {151}\) The court further reiterated that the purpose of the sui generis right is to promote the establishment of storage and processing systems for existing information and not the creation of materials capable of subsequently being collected in a database.\(^ {152}\) In this regard, it is worth mentioning that the BHB (the claimant) was constantly updating the database, costing approximately four million GBP

---

\(^{148}\) See cases C-46/02 *Oy Veikkaus*, para 39; C-203/02 *BHB v Hill*, para 35; C-338/02 *Svenska Spel*, para 29; C-444/02 *OPAP*, para 45.

\(^{149}\) *Football Dataco Limited v Sportradar GmbH* [2013] EWCA Civ 27, para 67.


\(^{151}\) See case C-203/02 *BHB v Hill*, para 38.

\(^{152}\) Ibid., para 31.
annually and involving around 80 employees. Plus, it had more than 20 million records and about 180,000 entries were made annually. Notwithstanding, the CJEU found those investments irrelevant to the making of BHB’s database.

In the cases of Oy Veikkaus, Svenska Spel and OPAP, the CJEU similarly concluded that the fixture lists of football matches were attributed to the organisation of the matches and not the making of databases. The CJEU therein stressed that the obtaining did not itself require any particular effort that could be deemed independent to the creation of the lists. The reasoning of the CJEU is an expression of the fact that producers do not need to deploy resources in the form of investments to obtain content that they already have access to. The main reasoning behind the judgments of the CJEU, which departed from the opinions of AG Stix-Hackl, is that the obtaining of data cannot be indivisibly linked to its creation. That distinction between creation and obtaining is somewhat of a metaphysical question. For instance, the collection of a genetic sequence is from one perspective a representation of a natural phenomenon that is existent, but on the other hand, it is not the natural phenomena itself that is collected since it is only question of capturing a representation and not the phenomenon itself. Such distinctions are yet to be conclusively settled by the CJEU and given the scarce total amount of case-law, it is unlikely that the issue will be resolved in the near future. Defining the distinction further would be nothing more than deeply uncertain legally speculations. Nonetheless, it cannot be stressed enough that the distinction is absolutely crucial.

As previously mentioned in subchapter 4.2.1, several pieces of data can collectively gain informative value thus constituting independent material in accordance with the database definition of Article 1(2). If the informative value only appears through the compilation of the data in the database, there has not been any substantial investment in the obtaining of the content since the content appeared for the first time when they acquired their informative value, i.e. the same time as the database was created. In the Directive’s terminology, there was no pre-existing material prior to the moment of being

---

153 Masson – Creation of database or creation of data: crucial choices in the matter of database protection, EIPR 28(5), 2006, p. 262.
154 See cases C-46/02 Oy Veikkaus, paras 44-45; C-338/02 Svenska Spel, paras 33-34; C-444/02 OPAP, paras 47-49.
inserted into the database. Accordingly, there has not been any investment that is independent from the creation of the informative value.

The process of the making of a database is consequently of major importance for assessing whether there has been an independent substantial investment or not.\(^{158}\) The distinction between the obtaining and the creation is closely related to the critical question of defining the notion of data. The assessment of when there is an obtaining of data is therefore inevitably an assessment of when the data starts to exist. In order to determine when data exists, one must determine what actually brings the data into existence. These intricate matters are yet to be conclusively settled by the CJEU.\(^{159}\)

Information is not synonymous with data since information can be understood as data providing information. There is naturally a subjective side to it since the informative value lies in the mind of the beholder. Since it is not satisfactory to have such a relative definition, something absolute must be defined. Derclaye has argued that data supposedly is intangible materials that bear no meaning to anyone.\(^{160}\) This is contrary to AG Stix-Hackl who defined data as “elementary statements with potentially informative content”.\(^{161}\) The AG’s reasoning corresponds with the CJEU settling that there has to be informative value in order to meet the requirement of being independent. Therefore, the question of what brings the data to it existence should be determined by what exactly gives the data its informative value.

These matters would hardly have been an issue if the CJEU would have applied the same main purpose of giving a wide scope of application similar to what they did in relation to the database definition. Instead, they chose a significantly more restrictive approach. If the sui generis right would correspond with the database definition, the whole issue of informative value would practically be solved since the content of a database would at virtually all times have informative value to a third party, thus satisfying the requirement of being independent and therefore existing prior to the point of being inserted into a database.

One further illustration of the issue can be when someone buys data from another party. In that case, there is an outstandingly clear case of investment since there is a factual purchase. Inspired by this conduct, it is unclear if a database maker can

\(^{159}\) Bygrave – The data difficulty in database protection, EIPR 35(1), 2013, p. 35.
\(^{161}\) See cases C-46/02 Oy Veikkaus, Opinion of AG Stix-Hackl, para 33; C-444/02 OPAP, Opinion of AG Stix-Hackl, para 36.
circumvent the independence criterion by forming a corporate body where the data is created in one branch and the database is owned by another. If the CJEU would accept this form of set-up, the database maker and producer of the content would effectively be two different entities and, accordingly, there would be no trouble distinguishing the creation, which is done by one entity, from the investment in the database of another entity. In the data supply chain, it is accordingly clear that a buyer enjoys protection better than a presumptive seller. One could argue that also a seller of data makes an investment when creating, however, the investment can only cover the actual making of a database and there is not support for claiming that presumptive profits also can be attributable to an investment.

6.3.5 Investments in the verification or presentation

The notion of ‘verification’ refers to the resources used “with a view to ensuring the reliability of the information contained in that database, to monitor the accuracy of the materials collected when the database was created and during its operation”.\(^{162}\) Just like with obtaining, ‘verification’ likewise needs to be an independent substantial investment in relation to the creation of the material.\(^{163}\) In other words, the resources deployed whilst creating the data cannot amount to substantial investment.\(^{164}\)

When it comes to investments in the ‘presentation’ of the content of the database, the CJEU has settled it to be interpreted as the resources used for the purpose of giving the database its function of processing information. This includes all resources used for the systematic or methodical arrangement of the materials contained in that database, and the organisation of their individual accessibility.\(^{165}\) Since, according to Recital 20 of the Database Directive, materials necessary for the operation or consultation of certain databases are protected, AG Stix-Hackl has expressed that presentation is to be understood as protecting the structuring of the contents.\(^{166}\)

In the cases of *Oy Veikkaus, Svenska Spel* and *OPAP*, the CJEU settled that fixture lists of football matches were not independent investments in the presentation of the content since the presentation itself did not require any particular effort that could be

---

\(^{162}\) See case C-203/02 *BHB v Hill*, para 34.

\(^{163}\) See cases C-46/02 *Oy Veikkaus*, para 46; C-338/02 *Svenska Spel*, para 35; C-444/02 *OPAP*, para 51.

\(^{164}\) Aplin – *The ECJ elucidates the database right*, IPQ, 2005, p. 16.

\(^{165}\) See cases C-46/02 *Oy Veikkaus*, para 37; C-203/02 *BHB v Hill*, para 36; C-348/02 *Svenska Spel*, para 27; C-444/02 *OPAP*, para 43.

\(^{166}\) See cases C-46/02 *Oy Veikkaus*, Opinion of AG Stix-Hackl, para 72; C-338/02 *Svenska Spel*, Opinion of AG Stix-Hackl, para 62; C-444/02 *OPAP*, Opinion of AG Stix-Hackl, para 78.
deemed independent to the creation of the lists. In other words, the court applied the same reasoning as they did in terms of understanding ‘obtaining’. It is not clear if presentation only encompasses the resources invested in pre-existing content or also resources invested in a structure where content is subsequently inserted. This is of major importance and exposes the connection between the database definition and the sui generis right. If a database fails to meet the database definition, the content would unquestionably thus be excluded from protection. On the other hand, if the content and a system used for presentation is created separately and subsequently merged, there are no indications that this separation would affect the assessment of the investment. From a pragmatic point of view, the assessment of whether or not there has been a substantial investment in the presentation should not account for the chain of events leading up to the investment. The question of at what time the investments actually took place should be immaterial and the only fact that reasonably matters is if the investment de facto has taken place at the time of assessment. Nevertheless, it is once again blatantly unclear how the CJEU would reason regarding this issue.

As mentioned in above in subchapter 4.4.1, pursuant to Article 1(3), computer programs used in the making or operation of databases are excluded from the scope of the Database Directive. Notwithstanding, the use of systems falling under the scope of being a computer program can be accounted for whilst assessing the substantial investment in the presentation of the content. This may seem like an oddity. However, bearing in mind that the nature of the sui generis right protects the investments, it is quite natural that any form of qualitative or quantitative investment made in a system in order to produce database can be accounted for. It does not extend any protection to the computer program as such since the investment can only be accounted for in relation to the making of the database at hand. Accordingly, the investments in a computer program can never be accounted for if it is not being used for the making and running of a database. Moreover, whether a computer program successfully meets the definitions in the Computer Programs Directive is immaterial in relation to ‘presentation’ since there is no such requirement in neither the Database Directive nor in any case-law. Therefore, also the development of individual algorithms should reasonably also be able to be accounted for as an investment in the presentation of the contents.

---

167 See cases C-46/02 Oy Veikkaus para 46; C-338/03 Svenska Spel, para 35; C-444/02 OPAP, para 51.
6.4 Can big data be protected by the sui generis right?

6.4.1 The critical issue

In order to determine if big data can be protected, the key point is to determine if there has been a substantial investment in the obtaining, verification or presentation of the content. A big data database cannot meet the database definition prior to being the subject of application of the analytics. Hence, according to the database definition, no database exists at all prior to the point of application.

6.4.2 Investments in the obtaining

As mentioned, in order satisfy ‘in the obtaining’, the material inserted into the database must be pre-existent. Thus, it is obvious that raw data that is stored in big data databases cannot be protected, which is also evident from the fact that it cannot meet the database definition in Article 1. As a consequence, big data databases that are more akin to live feeds of colossal amounts of data can never be protected under ‘obtaining’.

Investments must additionally be independently made regardless of the amount of the actual financial resources spent and the obtaining must be separate from the process of creation, regardless of the resources invested overall. The highly abstract data difficulty will accordingly be crucial for determining when big data databases can be protected. Especially problematic situations may arise in relation to the generation of data within the field of Internet of Things where information sent from sensors cannot be protected since no analytics has been applied prior to the content at hand meeting the database definition.

Another problem in light of the data difficulty can be seen in relation to buying data. It is common for database makers to collect their data from other sources than themselves. The origin of the data, i.e. the point from where the data mining takes place, must itself meet the database definition. Otherwise, following the reasoning above, the content of the database would start existing at the point where it is subjected to big data analytics, which excludes it from protection. In other words, if the data is bought from a seller that does not have a database meeting the database definition, the buyer’s database, by strict interpretation, creates the data since the data concerned successfully meets the database definition at the point when the it is inserted into the buyer’s database. An interpretation that strict would be detrimental to the protection of big data databases. The CJEU does unfortunately not give any further guidance on the issue of the data...
difficulty. Nonetheless, based on the reasoning above, big data databases that collect its’
data from raw data feeds can never be protected under investments in the obtaining.

From the reasoning above, a fundamental conclusion can be made – namely
that bought data does not itself automatically entails protection – regardless of the fact
that there undeniably has been an investment. The selling source must successfully have
met the database definition. In terms of big data, most cases where the data is bought, it
likely occurs through some kind of subscription to a raw data feed. This conduct excludes
the data from protection under obtaining. From a practical point of view, it can however
be noted that the fact-finding needed to determine what specific parts of a database
originate from what specific sources can be nearly impossible. Such questions of proof
are nevertheless neither determined by the Database Directive nor the CJEU and are
therefore singlehandedly up to the Member States themselves to assess.

In conclusion, the data difficulty is the major obstacle for protecting big
data databases on the basis of investments made in the obtaining. As many big data
databases receive their inputs from raw data feeds (e.g. from sensors in Internet of Things
devices), such protection is consequently excluded. Quite intuitively, none if this is an
issue if the input data is mined from databases successfully meeting the database
definition and such databases can be protected the exact same way as traditional databases
can.

6.4.3 Investments in the verification or presentation
Verification as a basis for investment can be interesting from a big data perspective. When
the analytics is correlating data and creating informative value, one could easily argue
that this is a kind of verification. However, the verification itself needs to be independent,
just as with obtaining, from the creation of the content. Since verification aims at
protecting investments made for assessing the reliability and accuracy of the materials
and information, the development of advanced algorithms for automated decision-making
and profiling might be attributable to investments in verification. Nevertheless, such
investments in algorithms would have to be deemed independent from the principal
activity of the database maker. Questions like there ultimately decided on a case-by-case
basis and it is difficult to draw any further conclusions on a general level.

As to investments in the presentation, it is essentially a protection of
structuring of the database. For big data, this can enable the database makers to account
for the resources spent on the big data analytics. The analytics itself will not be protected,
but the resources spent on them can be accounted for in the accumulation of the substantial investment. This can seem like an inherent paradox when it comes to protection of big data. On one hand, the content cannot be protected without the analytics – since the analytics enables it to satisfy the database definition. On the other hand, the analytics cannot be protected since it is not part of the database. At the same time, the resources spent on analytics are resources spent on the process of meeting the database definition hence being resources spent on the creation of the database.

The resources spent on the creation of the database cannot be the same resources spent on the creation of the content, since the investment must be independent, just as the obtaining and verification. Notwithstanding, the data difficulty that is apparent with ‘obtaining’ does not exist in the same way for presentation. Obtaining requires that the data is pre-existent and/or originating from a database meeting the database definition. Presentation, on the other hand, only requires that the data has informative value prior to being subjected to big data analytics. Given the generous approach by the CJEU (as mentioned in subchapter 4.2.1), it is sufficient that another party can be provided with relevant information for having informative value. Therefore, provided that the presentation is deemed independent, big data can likely be protected under the basis of substantial investments in the presentation of the contents.
7 The scope of protection

7.1 Forms of infringements under the Directive

7.1.1 The legal framework

In accordance with the previous analysis made in chapter 5, big data under the Database Directive can only be eligible for sui generis protection and never copyright protection. There are two rights that the beneficiaries of sui generis protection can enjoy. Firstly, Article 7(1) prohibits the 1) extraction and/or re-utilisation of 2) the whole or of a substantial part of the contents of the database, evaluated qualitatively and/or quantitatively. Secondly, Article 7(5) prohibits the 1) repeated and systematic 2) extraction and/or re-utilisation of 3) insubstantial parts of the contents, 3) implying acts which conflict with a normal exploitation of that database or which unreasonably prejudice the legitimate interests of the maker of the database. This following chapter will follow the structure of the articles and analyse its impact in relation to big data protection.

In other words, the sui generis right entails two different layers of protection. Extraction and re-utilisation shall be interpreted in light of the purpose of the sui generis right. Consequently, the concepts shall be given a wide definition in order to not deprive the database maker of the revenue that enables the redemption of the costs incurred by the investments. In order to be considered an infringement, an act must both be contrary to the purposes mentioned in the recitals and the requirements stipulated in Article 7. This shall not be interpreted as a requirement stating that extraction or re-utilisation must have a commercial purposes. A commercial purpose can however affect the damages awarded due to an infringement.

The infringement of a database right can furthermore occur indirectly, i.e. from the use of third-party information, and not only by actually accessing the original database. It can be objected that this causes insecurity for users that are exposed to a

---

170 See cases C-46/02 Oy Veikkaus, Opinion of AG Stix-Hackl, para 74; C-338/02 Svenska Spel, Opinion of AG Stix-Hackl, para 64; C-444/02 OPAP, Opinion of AG Stix-Hackl, para 80.
171 Ibid., para 35.
172 Ibid., para 51. Also see case C-173/11 Football Dataco and Others, para 20.
174 See cases C-203/02 BHB v Hill, para 48; C-304/07 Directmedia Publishing, para 47.
175 See case C-545/07 Apis-Hristovich, para 50.
176 See case C-203/02 BHB v Hill, paras 52-53.
risk of conducting data mining from a third-party without any knowledge of any unlawfulness. Nevertheless, the good faith of any user is immaterial.\textsuperscript{177}

7.1.2 Defining extraction

Article 7(2)(a) provides a definition where extraction shall mean the 1) permanent or temporary transfer of 1) all or a substantial part of the contents of a database 3) to another medium by any form of means. Noteworthy, it is not required that the content is made public in order to constitute an infringement, contrary to re-utilisation. Based on the recitals, it is obvious that the Directive aims to protect databases from copying and rearrangement of its content without prior authorisation from the rightholder.\textsuperscript{178} This shall however not constitute an extension of protection to mere facts or data.\textsuperscript{179} The recitals may seem contradictory, but can illustrate that the legislators attempted to strike a fair balance between the interests of the rightholders and the risk of monopolisation of information when they drafted the Directive. Determining that fair balance is naturally a complex task.

Any unauthorised act of the appropriation of the whole or part of the contents of a database falls within the concept of extraction.\textsuperscript{180} Extraction shall be understood as a concept unencumbered by formal, technical or physical criteria. It is therefore immaterial whether or not a transfer of content is accomplished by a technical process in terms of “electronic, electromagnetic or electro-optical processes or any other similar processes”.\textsuperscript{181} Extraction does not solely cover processes similar to ‘copy and paste’, but also processes where the output may lead to an arrangement that is different from the original database that the content was transferred from.\textsuperscript{182} It does not matter whether the source of the content is the original database or a third party’s database.\textsuperscript{183} The fact there has not been a rearrangement of the contents can nonetheless serve as an indicator there has been an extraction. However, the assessment of facts and proof is up to national courts to assess and not the CJEU.\textsuperscript{184}

\begin{flushleft}
\textsuperscript{177} Axhamn – Databasskydd, 2017, p. 260.
\textsuperscript{178} See Recitals 7 and 38 of the Database Directive.
\textsuperscript{179} See Recital 48 of the Database Directive.
\textsuperscript{180} See cases C-304/07 Directmedia Publishing, para 34; C-545/07 Apis-Hristovich, para 40.
\textsuperscript{181} See case C-304/07 Directmedia Publishing, paras 37-38.
\textsuperscript{182} Ibid., paras 39-40. Also see case C-545/07 Apis-Hristovich, para 47.
\textsuperscript{183} See case C-304/07 Directmedia Publishing, para 41.
\textsuperscript{184} See case C-545/07 Apis-Hristovich, para 51.
\end{flushleft}
It must be stressed that there has to be a transfer from one database to another medium in order to constitute an extraction.\(^{185}\) It is accordingly not enough to merely look at the content of the database to conduct an extraction.\(^{186}\) A transfer does not mean that the content of a database has to be moved, i.e. different acts of copying are also covered. Also, both permanent and temporary transfers are covered by the notion of ‘extraction’. The distinction between the two lies within the duration of the storage on another medium and a temporary transfer can for instance the storage in the operating memory of a computer.\(^{187}\) The extraction takes place at the time when the extracted content is placed on a medium other than the original database.\(^{188}\) The CJEU’s conclusion that storage on another medium, like the operating memory of a computer, has an ambiguous meaning. It could refer to the CPU, hard drive or even peripheral sources like cloud storage and processing. The term does not emanate from previous law but was completely invented by the CJEU in the case of *Apis-Hristovich*.\(^{189}\) A reasonable explanation can be that the CJEU may simply have referred to a medium owned by another person than the rightholder of the sui generis right.

7.1.3 The users’ right to consult a database

The CJEU settled in the case of *BHB v Hill* that users have a right to ‘consult’ a database that has been made public.\(^{190}\) Accordingly, from the moment a database is made public, the sui generis right cannot be used to hinder users from consulting the database. Interestingly, the CJEU has however stated that the database maker may reserve exclusive access rights to a specific group.\(^{191}\) The allowing of a user to make a consultation does moreover not disallow the database maker from prohibiting the user from making an extraction or re-utilisation.\(^{192}\)

The meaning of ‘consultation’ is not completely clear and is not prescribed in any article of the Database Directive. The term is only used in the cases of *BHB v Hill*\(^{193}\) and *Directmedia Publishing*\(^{194}\), and in Recital 20 of the Database Directive.

---

185 See case C-545/07 *Apis-Hristovich*, para 44.
188 See case C-545/07 *Apis-Hristovich*, paras 44-45.
190 See cases C-203/02 *BHB v Hill*, para 54; C-304/07 *Directmedia Publishing*, paras 51, 54.
191 See cases C-203/02 *BHB v Hill*; C-304/07 *Directmedia Publishing*, para 55.
192 See case C-203/02 *BHB v Hill*, para 58.
193 Ibid., para 54.
194 See case C-304/07 *Directmedia Publishing*, para 51.
Furthermore, the meaning of the word deviates amongst the different language versions of the Directive. Both the English and the French versions use the word ‘consultation’. The Dutch and Spanish versions use words bearing similar meaning. The German version, on the other hand, uses the word ‘abgefragt’, which closely translates to ‘query’. Additionally, the Swedish version uses the word ‘sökning’ and the Danish version ‘søgning’ – both meaning ‘search’. Consultation itself is moreover not a term used in data science. Based on a comparison of the language versions above, it can be concluded that consultation should mean something that is similar to ‘query’, i.e. the retrieval or request of data from a database.

Consultation does not encompass any transfer of data, even though a transfer might be temporarily. If, for instance, a screen is used to display the contents in a way that necessitates transfer of all or substantial part of the content to another medium, that could be deemed an extraction that can be prohibited by the sui generis right. Consultation can in this regard be even more confusing given that in all electronic environments, a consultation would as a rule involve a temporary transfer, which falls under the notion of extraction. The CJEU seems to have reasoned that there is need for an exception to extraction applicable in limited circumstances where the database maker has implicitly consented to the consultation of the database, mainly by making it public. The reasoning does in turn minimise the risk of the sui generis protection being used as an electronic access right.

7.1.4 Defining re-utilisation
According to Article 7(2)(b), re-utilisation means 1) any form of making available to the public 2) all or a substantial part of the contents of a database. The prerequisite of making available to ‘the public’ can be an undetermined group of persons and the number of people actually accessing the re-utilised content is immaterial. Making the content available to the public can occur by the distribution of copies, by renting, and by on-line or other forms of transmission. The first sale of a copy of a database within the EU by the rightholder or with his consent consumes the right to control resale. It does however not

---

195 See case C-304/07 Directmedia Publishing, para 53.
196 Aplin – The ECJ Elucidates the Database Right, IPQ, 2005, p. 20.
197 See case C-202/12 Innoweb, para 51.
mean that the right to prohibit extraction or re-utilisation is consumed.\textsuperscript{198} Likewise, there is no consumption because there has been an online transfer.\textsuperscript{199}

The case of \textit{Innoweb}\textsuperscript{200} concerned a Dutch search engine where users could search for car advertisements and the search results included advertisements posted on third party websites. The search engine itself was a ‘dedicated meta search engine’ that used other websites’ search engines, contrary to e.g. Google that is a general search engine that is “based on an algorithm”.\textsuperscript{201} A dedicated search engine is essentially a software technology that lets users search multiple dedicated pre-selected websites. With the search results, a coherent overview is presented to the user with the ability to query information originating from the pre-selected websites.\textsuperscript{202} The CJEU concluded that there had been a re-utilisation through an online transfer to the public of a substantial part of the content of the claimant’s protected database. That is because the search engine provided the end user with essentially the same functionality as the implemented on the original database site. It translated queries from the end user into the search engine of the original database in real time in a way that all information was searched through and finally presented all the results using the format of its website, grouping the content together, but reflecting criteria comparable to those used by the search engine of the original database.\textsuperscript{203}

In essence, the distinction between a general search engine and a dedicated meta search engine became crucial, thus creating a Google-exception. The CJEU concluded that a general search engine is to be defined as based on an algorithm, which results in a rather hollow definition, in particular given that a dedicated meta search engine can likewise use an algorithm in order to compile its search results. A hollow definition is indeed problematic given that CJEU made such an important distinction between the different types of search engines in the judgment. In the national proceedings prior to the referral to the CJEU, it was settled that the difference between dedicated meta search engines and general search engines is that the former permits searching through specific themes on pre-selected websites which is contrary to general search engines. It

\textsuperscript{198} See case C-203/02 \textit{BHB v Hill}, para 52.
\textsuperscript{199} See Recital 43 of the Database Directive.
\textsuperscript{200} Case C-202/12 \textit{Innoweb}.
\textsuperscript{201} Ibid., para 24.
\textsuperscript{203} See case C-202/12 \textit{Innoweb}, para 54.
is thus unclear from where the CJEU actually derives the notion of “based on an algorithm”, that effectively creates a Google-exception, from other than “documents placed before the court”.\textsuperscript{204} It is furthermore problematic that dedicated meta search engines and general search engines is a technical distinction that may sound simple at first glance but can lead to assessments that in \textit{casu} focuses on the technical solutions rather than consequent legal analysis.\textsuperscript{205} Additionally, it is difficult to assess the extent of the Google-exception. To avoid infringement, it seems like a search engine must implement different functionalities than the original website or adopt different criteria in the display of search results. The scope of the exception is unclear, and it would indeed be contrary to the purposes of the Database Directive if the exception could be used as a \textit{carte blanche} to re-utilise any content – to any extent – as long as the search engine at hand is a ‘general search engine’ and not a ‘dedicated meta search engine’. \textit{Innoweb} does unfortunately not give any guidance on the extent of the exception since it does not further elaborate the scope of the definition of a general search engine.

Another issue arising in light of \textit{Innoweb} can be found at paragraph 54 of the judgment where the CJEU based part of its judgment on the fact that the infringing search engine searched all of the content located on the claimant’s website. However, an effective search engine does not search all the information on a database. Accordingly, part of the judgment seems to be based on a false discourse in computer science.\textsuperscript{206}

The result of the \textit{Innoweb} judgment is, regardless of the mentioned above, that operators of search engines must be able to show that they either offer different functionalities than the original website or that they adopt different criteria in the presentation of the search results.\textsuperscript{207} The findings of the CJEU can be interpreted as a strong verification of the aim to protect the database maker’s investments. In this regard, it can be noted that the use of a dedicated search engine probably resulted in larger amounts of traffic to the original database thus not necessarily harming any investment.\textsuperscript{208} On the other hand, not only traffic, but also the duration of the attention of the visitors is to many websites vital since the amount of views of each advertisement displayed on

\begin{itemize}
\item \textsuperscript{204} See case C-202/12 \textit{Innoweb}, para 54. Also see Vousden – \textit{Innoweb, Search-Engines and Engineering Legitimacy in EU Law}, IPQ 4, 2014, pp. 293-294.
\item \textsuperscript{205} Virtanen – \textit{Innoweb v Wegener: CJEU, Sui Generis database right and making available to the public – The war against machines}, European Journal of Law and Technology 5(2), 2014, pp. 8-9.
\item \textsuperscript{206} Vousden – \textit{Innoweb, Search-Engines and Engineering Legitimacy in EU Law}, IPQ 4, 2014, p. 302.
\item \textsuperscript{207} Borghi & Karapapa – \textit{Contractual restrictions on lawful use of information: sole-source databases protected by the back door?}, EIPR 37(8), 2015, p. 515.
\item \textsuperscript{208} Husovec – \textit{The End of (Meta) Search Engines in Europe?}, Max Planck Institute for Innovation and Competition Research Paper 6(14-15), 2014, pp. 17-18.
\end{itemize}
websites is paramount for the business model as a whole. At the same time, the use of external search engines is at many times better for the users since search engines can be more effective and user-friendly than many independent search utilities implemented on individual websites. The Directive only protects investments and not the users’ interest of accessing content in a user-friendly manner, which should serve as an indicator that a loss of traffic may sometimes hinder the redemption of the resources incurred by the making of the database thus constituting an infringement.

On a more general level, apart from Innoweb, the concept of re-utilisation through online transfer to the public covers the situation where data is extracted from the content of a protected database and subsequently transferred to a user.²⁰⁹ A parallel can be drawn to the autonomous concept of an ‘act of communication to the public’ and the CJEU’s case-law concerning hyperlinks in copyright law. In the case of Svensson, the CJEU assessed the definition of communication to the public and concluded that as long as there is no circumvention of restrictions imposed on the website, put in place to restrict access by the public, there is no need receive consent of the rightholder to publish copyright-protected material that already has been published on the internet.²¹⁰ However, note that the act of communication to the public in copyright is merely a subcategory of several exclusive rights of the rightholders whereas in the Database Directive, it is the overall broad concept for re-utilisation. Nevertheless, the underlying arguments for finding copyright infringement in hyperlinks and finding sui generis infringement in re-utilisation have to a certain extent been similar.²¹¹ It is worth emphasising to a ‘certain extent’ in this regard and the ‘act of communication to the public’ and the sui generis right’s ‘transfer to the public’ have not been interpreted likewise by the CJEU. Nonetheless, given that the purposes of each right correspond and that the conduct of the CJEU where it is prone to making analogies, the comparison is definitely not far-fetched. At the same time, the CJEU found that there had been an infringement in Innoweb, regardless of the fact that the original website was already available online. Thus, a major distinction can be made between the act of communication to the public in copyright and re-utilisation; the assessment in copyright law compares if the possible recipients of the copyrighted work and the alleged infringement can be considered to be part of the same

²⁰⁹ See case C-173/11 Football Dataco and Others, para 21.
²¹⁰ See case C-466/12 Svensson, para 31.
audience, whereas in the sui generis right, the assessment only aims at determining whether the infringing database itself can be accessed by the public or not.

In summary, the use of search engine can constitute infringements due to re-utilisation, although there is an important exception introduced in the case of Innoweb and the scope thereof is not clear.

7.2 Big data and the forms of infringements

It is evident that infringements can occur even though content in an infringing database is rearranged differently compared to the original database. Inherently, as in all forms of intellectual property infringements, a distinction must be made between infringement and mere inspiration. This distinction likewise applies to big data and infringement.

Big data is all about harvesting information from vast amounts of data, not merely storing it in large volumes. The same contents can provide different information depending on the analytics applied. Therefore, if someone would copy contents from a big data database, without the analytics, and subsequently apply a new set of analytics to the content, the output can in certain cases provide essentially different information. Accordingly, what is problematic about big data is that the same content can have various meanings when applied to different set of analytics. Since the actual database definition is dependent on the content bearing informative value, the change of such could easily be argued to constitute a new database. At the same time, the form of arrangements in an infringing database is immaterial. Additionally, there is no support that an infringing database must meet the database definition in order to be infringing. Nevertheless, it is difficult to argue that the investment of the database maker is affected if another party takes the same content and gives it another meaning. This shows that the issue of the data difficulty is ever-present and that extractions offer no exception. As there is virtually no guidance at all from the CJEU on the subject, there is a golden opportunity for creative litigators to create new precedent.

As to the notion of consultation, it can be summarised as something similar to running a query, which is always allowed when a database is made public. Defining a big data database as being public can give rise to complicated definition since, due to the nature of big data, it cannot be posted on a website or similar. Many big data databases are in practice similar to feeds of raw data that are directed to pre-designated systems. In that sense, they are not available to the public and should accordingly allow the database maker to prohibit users from consulting the database.
Another aspect with implications for consultation is the deployment and training of artificial intelligence systems. Artificial intelligence is advancing exponentially and can even be described as a key to unlocking the true value of big data.\(^{212}\)

In a competitive world, there is naturally a strong incentive to deploy artificial intelligence in as many fields as possible in order to train it better. It is evident that merely reading or looking at a database cannot constitute an infringement. Nevertheless, if the reading is done by an artificial intelligence system instead of a human, it is questionable if the answer is still that evident. If there is some kind of transfer of the contents that is subsequently stored in the artificial intelligence system, the conduct would fall under the scope of extraction. If the intelligent system instead only uses the content to make its own conclusions, it would be something similar to the act of reading done by humans, which is certainly an act within the concept of consultation. Consequently, there might be situations where commercial value can be harvested by mere consultation of big data databases without any extraction or re-utilisation happening. As this topic is rather futuristic, on the verge of being science-fictitious, it will indeed be interesting to see if artificial intelligence can have the impact predicted and, if so, how the CJEU decides to tackle the legal issues arising thereof.

As to the re-utilisation, the Google-exception is important for understanding infringements in light of big data. Even though big data itself is likely not accessible by web browsers, the result of big data analytics may be. For instance, a GPS-system can combine maps with vast amounts of both live and historical empirical data on traffic flows in order to calculate the best directions at a given time as an output result. These kinds of GPS-systems can be defined as big data databases. As an example, a third-party search engine operator could embed those directions directly into its own search engine. The embedded element would then only show the directions, i.e. the output result of the processing of the GPS-system. Consequently, the search engine would itself never even “touch” the contents (i.e. the maps and empirical data on traffic flows) since the actual processing would be completely conducted by the GPS-system. If it would be possible to circumvent the whole concept of re-utilisation of contents just by using the rightholders’ own processing system, the whole purpose of the Directive would fall. Instead, the situation would likely be assessed similarly to Innoweb and, by using the terminology of the judgment, be analogous to a dedicated meta search engine, thus constituting infringement.

\(^{212}\) Information Commissioner’s Office. *Big data, artificial intelligence, machine learning and data protection*, 2017, p. 8.
7.3 Substantial parts of the contents

7.3.1 Substantiality

The prohibition of extraction and re-utilisation of substantial parts of the content not only covers the acts of competitors, but any party causing “significant detriment, evaluated qualitatively or quantitatively, to the investment”.\footnote{See Recital 42 of the Database Directive.} The meaning of substantial part is not further defined in the Database Directive. Both qualitative and quantitative terms shall according to the CJEU be interpreted as the “investment in the creation of the database and the prejudice cause to that investment by the act of extracting or re-utilising that part”.\footnote{See case C-203/02 BHB v Hill, para 69.} The CJEU’s reasoning seems to establish a minimum level of investments that is reciprocal to the investment needed in order to be eligible for sui generis protection.\footnote{Axhamn – Databasskydd, 2017, p. 277.} Consequently, it is up to the courts to assess the correlation of the investment in the database with the specific extracted portions of that database. In case the courts fail to do so, the scope of protection would be wider than the sui generis right is supposed to ensure.\footnote{Davison, & Hugenholtz – Football fixtures, horse races and spin-offs: the ECJ domesticates the database right, EIPR 3, 2005, p. 8.}

7.3.2 Qualitative substantial parts

Substantial parts of the contents evaluated in qualitative terms shall be interpreted as the scale of the investment in the obtaining, verification or presentation of the contents of the database subjected to extraction or re-utilisation. That is regardless of whether the extracted or re-utilised contents represent a substantial part of the total contents. Even a ‘negligible quantity’ may be encompassed by the concept.\footnote{See cases C-203/02 BHB v Hill, para 71; C-545/07 Apis-Hristovich, para 66.} The CJEU concluded the following regarding the assessment of qualitative substantial parts in \textit{BHB v Hill}:

\begin{quote}
In order to assess whether \textit{those materials} represent a substantial part, evaluated qualitatively, of \textit{the contents} of the BHB database, it must be considered whether the human, technical and financial efforts put in by the maker of the database in obtaining, verifying and presenting \textit{those data} constitute a substantial investment.\footnote{See case C-203/02 BHB v Hill, para 76.} [Emphasis added.]
\end{quote}
What is noteworthy about the passage from *BHB v Hill* above is the different terms used and their different meanings. Foremost, the CJEU does not include the notion of ‘independent’ in relation to data or materials required under the database definition. Since ‘those materials’ is used to describe the subject matter of the infringement, it is implied that the infringer does not have to have a database meeting the database definition in order to constitute an infringement. Furthermore, ‘those materials’ might not have the same meaning as ‘data’. According to AG Stix-Hackl, data is “elementary statements with potentially informative content” and materials can be “recognisable entities”.219 ‘Contents’ itself does not distinguish between data and material and the CJEU uses both terms interchangeably in the cited passage above. If ‘materials’ can be protected, there is no requirement of any informative value, whereas ‘data’ has informative value according to the terminology provided by AG Stix-Hackl. It is unlikely that the CJEU intended to open up a possibility to circumvent infringement by merely extracting incomplete forms of information. Hence, ‘materials’ should also be able to constitute substantial parts of the content, regardless of the fact that they may lack informative value. This conclusion is profoundly abstract and theoretical. From a practical point of view, the situation where materials lack informative value – but still have human, technical and financial efforts put into its’ obtaining, verification and presentation – should reasonably be more or less non-existent.

7.3.3 Quantitative substantial parts

Substantial parts of the contents – evaluated in quantitative terms – shall be interpreted as the volume of extracted or re-utilised materials compared to the total volume of the contents in the database concerned. If a quantitatively *significant* part of the contents is extracted or re-utilised and the creation of the database concerned required the deployment of substantial resources, the investment in that extracted or re-utilised part is proportionately equally substantial.220 The actual size of any database that the contents have been extracted or re-utilised from is immaterial.221 In other words, the court has established an abstract distinction where the investment in the protected database itself shall be compared to the proportion of that database that has been extracted or re-utilised.

---

219 See cases C-46/02 *Oy Veikkaus*, Opinion of AG Stix-Hackl, para 33; C-444/02 *OPAP*, Opinion of AG Stix-Hackl, para 36.
220 See cases C-203/02 *BHB v Hill*, para 70; C-545/07 *Apis-Hristovich*, para 59.
221 See case C-545/07 *Apis-Hristovich*, para 60.
Therefore, the assessment is neither strictly absolute nor relative.\textsuperscript{222} Furthermore, there will be no infringement if a user extracts or re-utilises a substantial part that does not represent a substantial investment. That is because in those cases, the actual investments have not been harmed and no infringement has consequently occurred.\textsuperscript{223}

### 7.4 Substantiality and big data

#### 7.4.1 Qualitatively or quantitatively substantial

As to what is qualitative and quantitative substantial when it comes big data databases, there are no major implications in the application of the law compared to traditional databases. The threshold of ‘quantitative’ can however be a bit problematic given the characteristically massive volumes of big data databases. Only parts of those databases might satisfy the conditions for being protected. The court would then have to trace where the contents being extracted or re-utilised originate from in order to determine if the extracted or re-utilised contents can amount to substantial.

One practical problem that inevitably would arise for courts whilst assessing whether there has been an infringement of a big data databases is the fact that they are dynamic to such an extent that they are more or less perpetually changing. There is no easy way for courts to conduct their fact finding in those cases and in particular for finding at what point in time an assessment of an infringement shall be focused on. It has been suggested that dynamic databases should be viewed as consisting of a series of separately protected databases.\textsuperscript{224} This would truly result in an immense burden of fact-finding for the courts since they would have to assess an act of extraction or re-utilisation based on the contents and the appearance of the big data database at a certain given time. In this regard, it is noteworthy that this is a kind of fact-finding that is up to the national courts to decide and not the CJEU.\textsuperscript{225} Nonetheless, this serves as an example of how the abstract and theoretical nature of the Database Directive and its relevant case-law give rise to practical problems when applied to big data.

---

\textsuperscript{222} Axhamn – \textit{Databasskydd}, 2017, pp. 277-278.

\textsuperscript{223} Derclaye – \textit{The Court of Justice interprets the database sui generis right for the first time}, European Law Review 30(3), 2005, p. 431.

\textsuperscript{224} Davison, & Hugenholtz – \textit{Football fixtures, horse races and spin-offs: the ECJ domesticates the database right}, EIPR no. 3, 2005, p. 10.

\textsuperscript{225} See C-31/09 \textit{Bolbol}, para 40.
7.4.2 The relationship between the content and the analytics

As previously mentioned above, whether or not a big data database can be protected is mostly dependent on whether the analytics are applied. If the analytics are not applied until a certain point in time, the courts must accordingly determine if the extraction or re-utilisation took place prior to that point – or after – since that in turn would determine if the content would be protected thus also determining whether there has been an infringement. This is naturally an almost impossible task and yet another undeniably critical issue in terms of fact-finding for the national courts.

It can further be noted that it is only the content itself that is protected and not the analytics. This illustrates an inherent paradox in relation to big data and its possible protection under the Database Directive. The contents of a big data database cannot be protected without having analytics applied to it. On the other hand, the analytics themselves are not protected from infringement since they cannot be part of the database definition (as mentioned in subchapter 4.4.2). Accordingly, the extraction or-reutilisation of big data analytics cannot be deemed an infringement.

7.5 Insubstantial parts of the contents

7.5.1 Interpreting Article 7(5)

As a rule, it is permissible to extract or re-utilise parts of a database as long as they do not amount to the ‘substantial’ threshold. An exception to that rule can be found in Article 7(5) that prohibits 1) the repeated and systematic 2) extraction and/or re-utilisation of 3) insubstantial parts of the contents 4) implying acts which conflict with a normal exploitation of that database or which unreasonably prejudice the legitimate interests of the maker of the database.

It is stated in the Explanatory Memorandum that no fixed limit can be ascribed to ‘insubstantial’ and the database maker must be able to demonstrate that the allegedly infringing acts prejudiced normal exploitation.226 Rather self-explanatory, the definition of ‘insubstantial’ is furthermore defined as contrary to ‘substantial’ and every part that does not amount to substantial is accordingly insubstantial.227 The purpose of Article 7(5) is to prevent the circumvention of the prohibitions in Article 7(1). Its aim is accordingly to prevent repeated and systematic extractions or re-utilisations of insubstantial parts of the contents where cumulative effect would seriously prejudice the

---

227 See case C-203/02 BHB v Hill, para 73.
investments of the database maker.\textsuperscript{228} It is unclear if the CJEU seeks to encompass the intention of the infringer in the assessment of circumvention.\textsuperscript{229}

Repeated and systematic extractions or re-utilisations shall, notwithstanding the wording in the English version of the Database Directive, be interpreted as alternative requirements rather than cumulative.\textsuperscript{230} It can be questioned whether the act of extraction or re-utilisation must be made on numerous occasions. As mentioned above, extraction means temporary or permanent transfers of the contents whereas re-utilisation means the making of parts of the contents available to the public. In \textit{BHB v Hill}, the CJEU defined Article 7(5) as prohibiting ‘acts’ of extraction or re-utilisation, thereby implying the need for numerous actions.\textsuperscript{231}

In the recent case of \textit{Technomed v Bluecrest}\textsuperscript{232} from the England and Wales High Court of Justice, it was concluded that a .xml-file containing classifications and variables for screening heart patients through a cloud-based software was protected by the sui generis right.\textsuperscript{233} The defendants had copied a patient report on a .xml-file and used the file to compile output reports on new .xml-files for other patients using the same .xml-schema.\textsuperscript{234} The court held, amongst other things, that there had been an infringement of Article 8(5) due to the repeated compilations of .xml-reports by the defendants.\textsuperscript{235} What is noteworthy about this example from a national court is that there only had been one factual act of transfer of the content, i.e. the original .xml-file. Nevertheless, the court found that the compilation of new reports was considered numerous acts of extractions or re-utilisations.

The British judgment is a clear example of the fact that there could be a case of several acts even though a file only has only been copied on one single occasion. This example can be used as an illustration regardless of the fact that the CJEU does not account for the judgments of the Member States’ national courts. As already mentioned in other regards, it would definitely be beneficial to have more case-law from the CJEU regarding extraction or re-utilisation of insubstantial parts of the contents.

\textsuperscript{228} See case C-203/02 \textit{BHB v Hill}, para 86.
\textsuperscript{230} See cases C-203/02 \textit{BHB v Hill}, Opinion of AG Stix-Hackl, para 123; C-338/02 \textit{Svenska Spel}, Opinion of AG Stix-Hackl, para 119; C-444/02 \textit{OPAP}, Opinion of AG Stix-Hackl, para 144.
\textsuperscript{231} See case C-203/02 \textit{BHB v Hill}, paras 87, 89.
\textsuperscript{232} \textit{Technomed v Bluecrest} [2017] EWHC 2142 (Ch).
\textsuperscript{233} Ibid., paras 75-77.
\textsuperscript{234} Ibid., paras 38, 44, 58.
\textsuperscript{235} Ibid., para 81.
7.5.2 Insubstantiality and big data

The protection against extraction or re-utilisation of insubstantial parts of the contents does not have any major distinctive significance for big data as an object of protection. It can however be noted that, given the size of big data, that there might be situations where an extraction or re-utilisation does not satisfy the substantiality threshold. That insubstantial part can still, as mentioned above, result in an infringement if it causes unreasonably prejudice to the investment of the database maker. Therefore, even an output of big data analytics that itself might not result in infringement when extracted or re-utilised – due to failure to meet the substantial threshold – may nonetheless result in infringement based on how the output information is used by the alleged infringer. As a prospective remark, it can be discussed that if the training of artificial intelligence results in a situation where commercial value can be harvested from big data databases without any extraction or re-utilisation of substantial parts taking place, the flexibility of Article 7(5) can possibly be used to protect the economic interests of the rightholders.
8 Protecting databases without sui generis protection

8.1 The case of Ryanair

8.1.1 A new intellectual property right created?

According to Article 8(1), a maker of a database that is available to the public may not prevent a lawful user of that database from extracting or re-utilising insubstantial parts of its content. If a database maker attempts to restrict this by introducing contractual limitations on the lawful users, those limitations concerned shall accordingly be considered void according to Article 15 of the Database Directive.

In the case of Ryanair v PR Aviation, the database of Ryanair was used by a price comparison website that enabled users to compare prices of flight tickets and book them through that website. In accordance with the terms and conditions of Ryanair’s website, the use of the website was restricted including a prohibition on the use of automated system or software to extract data from the website. Visitors had to consent to the terms and conditions in order to access Ryanair’s website. The CJEU did not determine if PR Aviation (defendant) de facto had consented to the terms and conditions. Additionally, the court never stated that a legally binding contract had been established, but merely recognised the existence of “contractual relations”.

It was concluded by the domestic courts, and not questioned by the CJEU, that the Ryanair’s database could not be protected under the Database Directive. The CJEU concluded that the database did satisfy the database definition albeit could neither be protected by copyright nor the sui generis right. Consequently, Articles 8 and 15 could not be applied to Ryanair’s database, thus allowing the database maker to prohibit extraction or re-utilisation of insubstantial parts also by lawful users. The CJEU based its conclusion by the fact that the Directive seeks to strike a fair balance between the rights of the database makers and its lawful users. When there is no protection of a database, there is accordingly no need to strike a fair balance. Therefore, the database maker may, according to the CJEU, impose contractual limitations on the lawful users’ right to make extractions or re-utilisations of insubstantial parts of the contents. Such contractual

---

236 Case C-30/14 Ryanair v PR Aviation.
237 Ibid., para 16.
238 Vousden – Autonomy, comparison websites, and Ryanair, IPQ 4, 2015, p. 405.
239 See case C-30/14 Ryanair v PR Aviation, paras 34-35.
240 Ibid., paras 39-40.
limitations shall be assessed in accordance with the applicable laws of the Member states.\footnote{See case C-30/14 Ryanair v PR Aviation, paras 43-45.}

The reasoning stands in stark contrast to the case of Svensson\footnote{Case C-466/12 Svensson.} and the kind of restrictions that needs to be in place in order to prohibit acts of hyperlinking copyright-protected content. Following the case of Ryanair, it is sufficient to only have terms and condition in place on the website to effectively prohibit extraction or re-utilisation of the contents of the database. This can effectively be seen as the creation of a new intellectual property right for databases that satisfy the database definition but at same time are not eligible for copyright or sui generis right. On the other hand, the case of Ryanair can also be interpreted as merely stating that the Directive does not apply whatsoever to those databases thus leaving the whole issue up to the contract law of the Member States.

There are several consequences of this judgment. Above-all, the rationale to mitigate the monopolisation of information cannot be enforced on databases that are not protected. A growing field of operators base their business models on systematic extraction and re-utilisation of materials available online that may not be protected. One of the principal aims of the Database Directive was to properly develop the European information markets. The judgment has resulted in a situation where the rules and considerations of that Directive does not apply at all and leaves everything up to contractual law that is not harmonised in Europe.\footnote{Mysoor – Protecting the unprotected database, The Law Quarterly Review 131, 2015, p. 562.} Another conclusion is that a contract may warrant stronger protection than the Database Directive.\footnote{Borghi & Karapapa – Contractual restrictions on lawful use of information: sole-source databases protected by the back door?, European Intellectual Property Review 37(8), 2015, p. 524.} In other words, some databases will be completely excluded from regulation and the only possible compulsory rules that can be enforced upon them would be competition law.

The case of Ryanair is hardly surprising since it essentially just concludes that the Database Directive does not apply at all to databases that cannot be protected by copyright pr sui generis. What is surprising is that the CJEU has warranted protection by contract law to a new category of databases that are not eligible for protection under the Directive.
8.1.2 Big data – better off alone?

Given the great obstacles under the Database Directive preventing some forms of big data from protection, it is undeniably tempting to draw the conclusion that contracts might be a better solution to protect big data databases. From the public consultation done by the European Commission, it also seems to be a widespread opinion within the industry that contract law is the most appropriate protection. From looking at the case of Ryanair, it is easy to argue that only applying terms and conditions to the use of data is easier than having the content protected as intellectual property and the obstacles thereof for big data.

It is however important to note that the case of Ryanair leaves a lot for the national legal systems to decide. The way that general terms and conditions are presented for the user in Ryanair – i.e. with no need to make an active consent – is likely not acceptable for constituting a valid contract in many jurisdictions. Additionally, the sui generis right gives automatic protection to the database as intellectual property – which a contract never can. Furthermore, with the use of contract, there would naturally be an issue of conflict of laws and jurisdiction, nonetheless, such issues are easily resolved by stipulating clauses for applicable law and form for disputes.

Notwithstanding that it is outside the substantive scope of this essay, big data may, from a business perspective, also be protected as trade secrets. According to the Directive on the Protection of Trade Secrets, it is required for protection that there is information bearing commercial value due to the fact that it is secret and that reasonable steps have been taken to keep the information secret. As this is clearly outside the Database Directive, no further attention will be given to trade secrets.

9 Concluding discussion

9.1 On the aim and method

9.1.1 Can big data be protected under the Directive?

As to the elemental purpose of this thesis – i.e. to analyse how big data can be protected under the Database Directive – the short answer is that the protection is determined by the nature of the big data database at hand and that there are many obstacles to overcome for acquiring protection. The first obstacle is the database definition, which requires the implementation of big data analytics in order for big data to be considered a database which excludes raw data feeds from protection. The second obstacle is actually meeting the requirements for protection. It is impossible for big data to be protected by copyright since it cannot satisfy the requirement of originality, given that it as a rule is machine-generated. Sui generis protection is possible provided that there has been investment in the obtaining, verification or presentation – requirements that the CJEU truly has interpreted narrowly. In relation to obtaining, a great obstacle emerges due to the fact that the data must be pre-existing. Big data probably has its best chances of being eligible for protection on a general level under ‘investments in the presentation’ since the investments made in big data analytics then can be accounted for. However, big data analytics cannot be protected solely since it might be eligible for protection as a computer program, which in turn excludes the analytics from the scope of the Directive. The standing of big data analytics under the Database Directive and the Computer Programs Directive can definitely be examined further and is a splendid question for further research. Of particular interest is the way that the two directives can overlap.

As to the scope of protection and how infringements can be assessed in light of big data, it can be concluded that the assessment is similar to infringements in regular databases under the Directive. Nevertheless, one difficulty is the definition of re-utilisation and the so-called Google-exception that causes insecurity regarding the scope of the concept of re-utilisation. Infringements can also be more difficult to assess in the digital environment that big data exists in. In case big data cannot be protected, the CJEU has warranted contractual protection that can be more generous than the Directive permits and accordingly, instead of finding an infringement, it is possible to protect the databases through the contractual liability of its users.
The conclusion above all have their basis in case-law. It is worth emphasising that this only a question of basis and not in any way concluding judgments covering the issues of big data. In total, there have only been 11 judgments in the CJEU covering the Database Directive – which is not an ideal foundation for analysis. The above-mentioned conclusions are accordingly speculations based on the available case-law and applied to big data though common interpretation methods of the CJEU and by reason. It might as well be proven wrong by the stroke of a pen if it successively reaches the CJEU.

9.1.2 Personal reflections on the aim and method
This thesis has covered law. Accordingly, legal sources have been used in a methodical way with the inherent pursued aim of highest possible legal certainty. Given the lack of extensive case-law, I might have failed securing that legal certainty. The proportion of de lege lata might have become inferior compared to the proportion of de lege ferenda. I have no problem with that conclusion. However, I do wish to raise an argument. In the field of intellectual property, and in particular with regard to emerging technology, the law is generally far behind technical innovation. It can take years for the status of new technologies to be conclusively assessed by the courts. For example, it took ten years for the CD, eight years for the DVD and seven years for .mp3-files.247 If one would argue that legal science must be confined strictly to issues with high levels of legal certainty, then one would inevitably also require an issue to be handled by the courts prior to meeting the threshold for satisfactory legal certainty thus being eligible for examination in the ivory tower of legal academia. It would also be the equivalent of legal science shooting itself in the foot since the only way it could analyse technological innovation would be after the courts have paved the way.

Moreover, I am fully aware of the fact that I am not even close to being worthy of the title of a legal scholar, thus hampering the extent of discretion I enjoy whilst giving value to my own conclusions and opinions. There is inescapably an ever-present question of striking a fair balance between legal certainty and discretion. Notwithstanding, I do believe that too much emphasis on these aspects would obstruct the ability to make predictions within the field intellectual property law and technical innovations. Such view would additionally be contrary to the axioms of the legal

dogmatic theory, i.e. to both construct an explanation of the law as it is, and also the normative aspects that support it. If, regardless of my reasoning provided above, the thesis is considered being too speculative – so be it.

9.2 Some prospective outlooks

9.2.1 Data – a priority in the European Digital Single Market

The issue of data markets has been the topic of several public consultations by the Commission. Opinions by business respondents have expressed their primary concern to maintain contractual freedom due to the complexity and uniqueness of each situation in the data-driven economy. There is furthermore a fear that regulation could risk stifling innovation and impede the development of the markets. On the topic of personal data, newly introduced regulation clarifies to some extent the issues of ownership, use and access of personal data. When it comes to personal data, nobody has hardly missed the introduction of the General Data Protection Regulation (more commonly known by its acronym ‘GDPR’). According to its Article 20, registered natural persons have the right to request to have their data transferred (so-called data portability), and to have their personal data erased according to Article 17. Personal data is not covered at all by the Database Directive and has therefore not been emphasised any further in this thesis. Nevertheless, the emphasis that the European Commission has put on the development of the digital economy, for instance by the introduction of the GDPR, indicates its importance. Also, the fact that personal data forcefully can be both removed and transferred to a competitor under the GDPR shows that EU legislators are willing to decrease the rights of database makers in favour of others’ interests.

Other than the GDPR, the EU has initiated massive legislative processes under the European Digital Single Market initiative. Especially relevant for big data, it can be mentioned that there is a proposal for a directive on copyright in the Digital Single Market and Article 3 of that proposal enables text and data mining by lawful users of databases for the purpose of scientific research. According to Recital 8 of the proposed directive, “research organisations such as universities and research institutes are

---


250 See Recital 48 of the Database Directive.

confronted with legal uncertainty as to the extent that they can perform text and data mining”. If enacted, it will serve as an evident example of how the rights of database makers are balanced against competing interests and it would effectively diminish the argument that the Database Directive results in a risk that information can be monopolised. This might, from a completely speculative perspective, justify the CJEU to stretch the rights of database makers further in the future, given that it is outside the scope of research purposes. This is because the necessity of striking a fair balance would to a larger extent already have been accounted for in the legislation hence limiting the extent of the need for the CJEU to balance the competing interests of the stakeholders.

9.2.2 Evaluation of the Database Directive

The European Commission issued an evaluation of the Database Directive in 2005. The evaluation stated that the CJEU Grand Chamber’s judgments in OPAP, Oy Veikkaus, Svenska Spel and BHB v Hill of 2004 were contrary to the Commission’s original intention to “protect databases in a wide sense”. It was additionally concluded in the evaluation that the Directive had failed to accomplish one of its main purposes, namely the protection of databases lacking originality. The additional layer of intellectual property protection established by the sui generis right had additionally not appeared to stimulate innovation and growth within the EU. This is ostensibly a very severe criticism that pierces into the whole raison d’être of the Database Directive. Nonetheless, the Directive was not amended following the evaluation.

The most significant change to the Database Directive might nevertheless be introduced very soon. On 18 May 2017, the European Commission declared the intention to make another evaluation of the Database Directive. The evaluation shall assess whether the Directive fulfils its designated policy goals and if it “still is adapted in view of development of new technologies, new business models based on data exploitation, and other emerging data-related issues, policies and legal frameworks on data access and ownership”. The evaluation was supposed to be completed during Q1 2018. It is indeed unfortunate that the results of that evaluation have not been published

254 Ibid., p. 24.
at the time of writing this thesis. Given the importance of big data, it will surely be mentioned in the evaluation and it will most likely be possible to compare the upcoming evaluation’s findings with this thesis.

9.2.3 Technology-driven change

The technological trend is moving towards a further significant increase of data in the world. For instance, in accordance with the so-called Moore’s law, the evolution of computer processors has enabled a technical development where the number of transistors in a dense circuit has doubled every second year, thus doubling the processing power of computers. As technology looks today, it is not possible to uphold this rapid increase of processing power without increasing the size of the processors. One solution is to use data centres that receive data, processes it and subsequently sends the processed data back to the user. This is called cloud computing and is already utilised by, for instance, the Siri assistant in iPhones for voice recognition. In essence, processing power can now be compared to a resource available “on tap” through cloud computing. The impact is further enhanced by the development of the Internet of Things where new devices increasingly are connected to the internet. Consequently, cloud computing will likely become embedded in more devices hence leading to the generation of additional data.256

When there is more data, there is more money and, inevitably, also more competition. On top of that – as numerous futurists also predict – an oncoming revolution of artificial intelligence, it is reasonable that the production made by artificial intelligence powered devices will increase. Since satisfying the originality requirement is needed in order to be eligible for copyright and that machine-generated works are excluded thereof, it is reasonable to assume that the incentives to have works protected by the sui generis right, that does not any require originality, will surge. The protection of big data under the Database Directive might accordingly become more important than ever before.

---

Bibliography

**Articles**


Deporteer, Ben. *Technology and Uncertainty: The Shaping Effect on Copyright Law*, University of

Derclaye, Estelle  
*The Court of Justice interprets the database sui generis right for the first time*, European Law Review 30(3), 2005, pp. 420-430.  

Farkas, Thomas J.  
*Data Created by the Internet of Things: The New Gold Without Ownership?,* Revista La Propiedad Inmaterial no. 23, 2017, pp. 5-17.

Gandomi, Amir & Haider, Murtaza  
*Beyond the hype: Big data concepts, methods, and analytics*, International Journal of Information Management, 2015, pp. 137-144.

Husovec, Martin  

Jareborg, Nils  

Masson, Antoine  

Mysoor, Poorna  

Neethu, Rajam  
<table>
<thead>
<tr>
<th>Author</th>
<th>Title</th>
</tr>
</thead>
</table>

**Internet sources**

- **Cambridge Analytica** Homepage (Retrieved on 21/01/2018), available at: https://cambridgeanalytica.org/
- **Cambridge Analytica Political** Homepage (Retrieved on 21/01/2018), available at: https://ca-political.com/ca-advantage
- **Global Security Review** Cambridge Analytica: The Darker Side of Big Data, 12/01/2018 (Retrieved on 11/02/2018), available at:

The Guardian  

The Washington Post  
*Gary King: Big data is not actually about the data*, 20/05/2016 (Retrieved on 25/01/2018), available at: https://www.washingtonpost.com/blogs/post-live/wp/2016/05/05/meet-professor-gary-king/?utm_term=.6d4f0fbd8592

*How stores use your phone’s WiFi to track your shopping habits*, 19/10/2013 (Retrieved on 11/02/2018), available at: https://www.washingtonpost.com/news/the-switch/wp/2013/10/19/how-stores-use-your-phones-wifi-to-track-your-shopping-habits/?utm_term=.d908bdb4c9e0

**Literature**

Axhamn, Johan  

Bainbridge, David I.  

Bergström, Carl Fredrik & Hettne, Jörgen  

Bobek, Michak  


Database rights in safe European home: the path to more rigorous protection of information, Lappeenranta University of Technology, 2005.

Wolk, Sanna *Datorprogramsalster i upphovsrätten*, Iustus, Stockholm 2016.

**Official documents**

*European Commission*

<table>
<thead>
<tr>
<th>Document ID</th>
<th>Title of Document</th>
</tr>
</thead>
<tbody>
<tr>
<td>COM(2017) 9 final</td>
<td><em>Building A European Data Economy</em> (10/01/2017).</td>
</tr>
<tr>
<td>COM(92) 24 final</td>
<td><em>Proposal for a council directive on the legal protection of databases</em> (13/05/1992).</td>
</tr>
</tbody>
</table>


Other public authorities


Table of cases

**Court of Justice of the European Union**

<table>
<thead>
<tr>
<th>Case Number</th>
<th>Case Title</th>
<th>Date</th>
<th>ECLI</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-490/14</td>
<td>Freistaat Bayern v Verlag Esterbauer GmbH</td>
<td>29/10/2015</td>
<td>EU:C:2015:735</td>
</tr>
<tr>
<td>C-406/10</td>
<td>SAS Institute Inc. v World Programming Ltd</td>
<td>02/05/2015</td>
<td>EU:C:2012:259</td>
</tr>
<tr>
<td>C-30/14</td>
<td>Ryanair Ltd v PR Aviation BV</td>
<td>15/01/2015</td>
<td>EU:C:2015:10</td>
</tr>
<tr>
<td>C-466/12</td>
<td>Nils Svensson and others v Retriever Sverige AB</td>
<td>13/02/2014</td>
<td>EU:C:2014:76</td>
</tr>
<tr>
<td>C-202/12</td>
<td>Innoweb BV v Wegener Mediaventions BV</td>
<td>19/12/2013</td>
<td>EU:C:2013:850</td>
</tr>
<tr>
<td>C-145/10</td>
<td>Eva-Maria Painer v Standard VerlagsGmbH and others</td>
<td>07/03/2013</td>
<td>EU:C:2011:798</td>
</tr>
<tr>
<td>C-173/11</td>
<td>Football Dataco Ltd and others v Sportradar GmbH and Sportradar AG</td>
<td>18/10/2012</td>
<td>EU:C:2012:642</td>
</tr>
<tr>
<td>C-604/10</td>
<td>Football Dataco Ltd and others v Yahoo! UK Ltd and others</td>
<td>01/03/2012</td>
<td>EU:C:2012:115</td>
</tr>
<tr>
<td>C-393/09</td>
<td>Bezpečnostní softwarová asociace - Svaz softwarové ochrany (BSA) v Ministerstvo kultury</td>
<td>22/12/2010</td>
<td>EU:C:2010:816</td>
</tr>
<tr>
<td>C-31/09</td>
<td>Nawras Bolbol v Bevándorlási és Állampolgársági Hivatal</td>
<td>17/06/2010</td>
<td>EU:C:2010:351</td>
</tr>
<tr>
<td>C-545/07</td>
<td>Apis-Hristovich EOOD v Lakorda AD</td>
<td>05/03/2009</td>
<td>EU:C:2009:132</td>
</tr>
<tr>
<td>C-304/07</td>
<td>Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg</td>
<td>09/10/2008</td>
<td>EU:C:2008:552</td>
</tr>
<tr>
<td>C-46/02</td>
<td>Fixtures Marketing Ltd v Oy Veikkaus Ab</td>
<td>09/11/2004</td>
<td>EU:C:2004:332</td>
</tr>
<tr>
<td>C-203/02</td>
<td>The British Horse Racing Board Ltd (BHB) and others v William Hill Organization Ltd</td>
<td>09/11/2004</td>
<td>EU:C:2004:694</td>
</tr>
</tbody>
</table>


C-397/01  *Bernhard Pfeiffer and others v Deutsches Rotes Kreuz and others*, 05/10/2004, ECLI:EU:C:2004:584.


**United Kingdom**

*Technomed v Bluecrest* [2017] EWHC 2142 (Ch).

*Football Dataco Limited v Sportradar GmbH* [2013] EWCA Civ 27.

**United States of America**