CREATe Working Paper 2016/07 (June 2016)

From infringement to exception: why the rules on data mining in Europe need to change

Author

Christian Geib
PhD Candidate Humanities & Social Sciences, Law School, University of Strathclyde

CREATe Working Paper Series DOI: 10.5281/zenodo.56425.

This release was supported by the RCUK funded *Centre for Copyright and New Business Models in the Creative Economy (CREATe)*, AHRC Grant Number AH/K000179/1.

From infringement to exception: why the rules on data mining in Europe need to change

Christian Geib

PhD Candidate Humanities & Social Sciences, Law School, University of Strathclyde

This research was supported by the RCUK funded Centre for Copyright and New Business Models in the Creative Economy (CREATe), AHRC Grant Number AH/K000179/1.

Abstract

As societies gather increasing amounts of data, interest is growing in technology that allows people to explore and analyse it. The process used, known as data mining, offers many potential benefits. In the pharmaceutical field, for example, it can help medical researchers to discover correlations between illnesses and treatments, as well as improve diagnostics, aid understanding of multifactorial diseases, and repurpose existing drugs. Law enforcement agencies can use data mining to spot patterns in the times and dates that crimes are committed, helping them to allocate resources more effectively and prevent crimes from being committed in first place.

But questions remain about the legal framework that governs the use of data mining. In particular, do UK and EU copyright rules and the sui generis database right law impede the use of DM? These questions are important, because evidence suggests that when it comes to data mining, Europe is lagging behind the US and a number of Asian countries. One reason for this is that copyright protection in EU and European Economic Area countries is relatively strong, which has a negative net effect on the use of data mining.

This paper considers how the relationship between UK and EU copyright and database law and data mining affects three important industries: the pharmaceutical sector, law enforcement and marketing. It concludes that data mining is prima facie infringing. The legal exception and defences currently available are at best insufficient and at worst not applicable to data mining. Such legal uncertainty is likely to have a chilling effect on data mining in Europe.

The paper concludes by considering potential legislative and regulatory solutions to ensure that data miners in Europe are able to achieve the potential benefits that data mining offers.

One potential solution is for copyright owners to grant researchers licences to mine their data. In addition to the cost and scalability challenges of obtaining licences, however, there is a risk that a licensing-based solution could be used to erode copyright rules and exceptions.

Another potential solution would be to introduce a wide, US-style fair use exception to copyright law. This has the advantage of being able to deal with emerging technological developments and is thus relatively "future-proof". However, many European legal scholars and lawmakers are likely to resist a wholesale import of US-style fair use provisions. Policy makers may be more amenable to a compromise approach, similar to that proposed by the Irish Copyright Review Committee, which recommended combining UK-style fair dealing and closed exceptions with fair use.

Given that Europe is falling behind in the crucial technology of data mining, it is in Europe's competitive interest for policy makers to err on the side of openness, even if fair use provisions seems out of reach. A closed, mandatory, data mining-specific exception rather than a technology-neutral, future-proof solution seems to be the most likely outcome of the latest round of EU copyright reform. Whatever the Commission proposes, it is crucial that policy makers are pushed as far as possible in the direction of openness and legal certainty to maximise the gains for data mining in Europe.

Acknowledgements

The author owns a significant debt of gratitude to Emma Barraclough for her hard work and skill in editing this paper, Professor Lilian Edwards and Professor Martin Kretschmer for their continuous and patient advice and feedback and to CREATe RCUK Centre for Copyright and New Business Models in the Creative Economy for the funding and support of my PhD. The author welcomes comments and can be reached at christian.geib@strath.ac.uk.

1. Introduction

As societies gather increasing amounts of data, interest is growing in technology that allows people to explore and analyse it. The process by which this is done – known as data mining (DM) – is an automatic or semi-automatic way of manipulating large quantities of data to discover patterns and rules. But what social and legal questions does DM raise? In particular, do UK and EU copyright rules and the *sui generis* database right law impede the use of DM?

This working paper seeks to offer some answers to these questions. In doing so, it will consider how the relationship between UK and EU copyright and database law and DM affects three important industries. It will also examine whether the newly introduced copyright exception s. 29 A of the UK Copyright, Designs and Patents Act 1988 (UK CDPA) has improved legal certainty for data miners. Finally, it will consider proposals to reform the law in this area.

We selected three industries for the purposes of this study: pharmaceuticals, marketing and law enforcement. These were chosen based on the copyright relevance of the data mined within each industry. Fundamentally it can be separated into two basic categories: data that is copyright protectable and data that is not copyright protectable. These two different categories can be further subdivided into commercial and non-commercial uses.

Data mined in the pharmaceutical industry largely involves information stated directly in natural language such as articles in scientific journals and their abstracts[1], both of which are copyright protected.

The marketing industry is one of the biggest users of DM. Data miners in this sector seek to mine copyright protectable data as well as data that may not be copyright protectable.

Law enforcement data miners also mine copyright protectable data and data not protectable by copyright. In addition, much of the data 'mined' in the law enforcement sector involves public sector data, the use of which may raise serious civil liberty concerns.

2. Social benefits and dangers of data mining in three industries

One of the overarching aims of this paper is to consider what legal reforms are necessary to enable the better application of DM technology. It is essential, therefore, at the outset, to establish that DM's potential benefits outweigh potential dangers.

A. Potential benefits of data mining

In the pharmaceutical industry the DM of thousands of journal articles and/or their abstracts enables researchers to discover novel treatments. Given that reading a similar volume of journal articles might be impractical for human researchers, many correlations and patterns that are the key to such novel treatments might have gone unnoticed had it not been for the use of DM. DM can also be used to discover patterns in neighbouring disciplines[2]. Data mining thousands of journal articles across disciplines can also improve diagnostics, aid understanding of multifactorial diseases, repurpose existing drugs and prevent dangerous drug-drug interactions. What makes DM of journals particularly attractive from a taxpayer's perspective is that it offers new benefits from research that has already been paid for.

In the law enforcement sector, data mining large numbers of police reports to find correlations and patterns can help the police allocate resources more effectively and efficiently, with the aim of preventing crimes from being committed in first place. Crime matching, which allows the police to discover patterns within a series of crimes, can also help them identify perpetrators. Further benefits could be the review of police conduct by data mining police reports. This could reveal, for example, unusual stop and search patterns and other hidden biases.

The task of marketing is to locate and segment customers who may be interested in a certain product and deliver communication to them to trigger the desired consumer action. The use of DM in the marketing industry may help marketers achieve higher returns on their budget and ensure that customers receive more relevant advertisements. DM can provide valuable targeting information on a potential customer's location, socioeconomic background, social networks, previous buying preferences and imminent life changing events (such as the birth of a child or graduation)[3]. This will

enable marketers to predict a consumer's preferences and income more accurately and to target them more effectively.

B. Potential dangers of data mining

While the potential benefits of DM are considerable, DM is not without its potential dangers. These vary by industry.

In pharmaceutical research, especially in clinical trials, informed consent plays a key role in any research on human subjects. It is defined as a process 'by which a subject voluntarily confirms its willingness to participate in a particular trial, after having been informed of all aspects that are relevant to the subject's decision to participate'[4]. However, a potential problem with the DM of journal articles or of raw data from 'old' studies is that some of the 'new' research objectives were never envisaged at the time the research subject gave his or her consent. Any new research objectives could involve sensitive personal data, for instance by providing insight into the susceptibility of some of the research subjects to hereditary diseases. In that context data protection issues and proper anonymisation of data becomes highly relevant. Accidental or criminally induced loss of data or security breaches could have far-reaching consequences for data subjects in terms of their employability or insurance premiums. A further potential problem is that researchers could unwittingly neglect long-established scientific principles and instead settle for mere numbers and correlations.

In the law enforcement sector, over-reliance on DM could lead to profiling and discrimination of entire neighbourhoods and ethnic groups. There are a number of potential fallacies that could falsely influence law enforcement strategies, even if law enforcers claim that they are relying on seemingly 'objective' data on crime hotspots. For example, computer algorithms that form the core of DM software employed for law enforcement purposes also can reflect the biases of their developers. The common conflation of correlation and causation might also fallaciously inform law enforcement strategy and actions. Furthermore, relying on inaccurate or out-of-date data could result in highly

disruptive false-positives. In addition, the loss of data, whether through criminal security breaches or accidental data leaks, could have serious implications for the data subjects in question.

Data protection is also a crucial issue in the marketing industry. The consequences for individuals can vary from financial losses due to leaked credit card details to social ostracism after being revealed to have visited certain websites. Even if data is not leaked, there are potential discrimination issues where entire area codes or income groups are placed in a less favourable consumer category. This is especially problematic if it is based on conflating correlation and causation or on fallacious data. Furthermore, the use of DM for the marketing sector highlights the issue of neglecting the 'data fringes'[5]. This means that DM selectively caters to certain data-rich segments of society that leave substantial data traces. These people often use smart phones, own laptops, use credit cards or online payment services such as PayPal, shop online, subscribe to online music and movie services, use online geolocation-based services and frequently use online-based social networking services. Marketers may neglect data-poor'[6] sectors of society, where people pay by cash rather than credits cards, use library computers to access the internet, own prepaid mobile phones and consume media content through DVD players.

Do the benefits of DM outweigh its risks? The answer to this question may vary depending on the sector in which DM is employed. It could be argued that the health benefits resulting for society from the mining of pharmaceutical journals outweigh any data protection or consent concerns. Moreover, data protection considerations could be addressed by new legislation. Data mining in the marketing industry may be harder to justify. Commercial gains and better targeted consumer advertising might not be found to outweigh the risks of (fallacious) data-based discrimination and data leaks.

3. Why could data mining infringe copyright and the *sui generis* database right?

Having established that data mining can offers significant benefits for business and wider society, we now turn to the question of whether the law hinders DM. We first look at the relationship between copyright and data mining.

DM involves two steps relevant to copyright law. First, the scanning (whether from a paper copy, electronic files or 'scraping' content of websites) of copyrighted works produces a computer image of text and thus creates a copy. Second, the application of Optical Character Recognition converts the image into a text file. This can be regarded as a second copy.

The copy is stored in a repository or data warehouse for the central DM process, which is the application of the algorithmic analysis of the data. One of the key exploitation rights is the copyright owner's exclusive reproduction right. Copying means reproducing the work in any material form (s. 17(2) UK CDPA) and copying of designated works needs the permission of the copyright holder. If no permission has been granted, copying is *prima facie* infringing.

This raises five basic questions, which this paper will explore:

1. What kind of data is being copied?
2. Is the copied material protectable by copyright?
3. If the copied material is not protectable by copyright, could it be protectable by the *sui generis* database right?
4. If the unauthorised copying is found to be *prima facie* infringing copyright and/or the database right, are there any applicable legal exceptions?
5. If there are no applicable legal exceptions, would licensing be an option?

4. Copyright and *sui generis* database right issues in key industries

In this section we will look at how the three industries listed in the introduction are affected by copyright and the *sui generis* database right. We will demonstrate that DM in the three industries *prima facie* infringes copyright where it mines content that is copyright protectable. It also *prima facie* infringes the *sui generis* database right where it mines content that is not copyright protectable.

A. The pharmaceutical industry

DM in the pharmaceutical industry *prima facie* infringes copyright and the *sui generis* database right. Not only journal articles are protectable by copyright but also scientific databases might be protectable by copyright if they fulfil the four criteria for databases in s. 3A(1) UK CDPA. They need to be:

1. a collection of independent works, data or other materials which
2. are arranged in a systematic or methodical way
3. are individually accessible by electronic or other means
4. are original if, and only if, by reason of the selection or arrangement of the contents of the database, they constitute the author's own intellectual creation

We will consider each criterion in turn.

1. To be independent, the works contained in such databases would need to be '…separable from one another without their informative, literary … or other value being affected[7]. Journal articles contained in scientific databases such as Westlaw, ScienceDirect, Scopus or JSTOR are indeed separable from one another without losing their informative value and fulfil the criterion of independence.

2. Recital (13) EU Database Directive[8] (EU Db. Dir.) requires that the collection follows a particular plan or method of classification to allow the retrieval of any independent material[9]. Scientific databases are organised in an organised and systematic way, thereby fulfilling this criterion.

3. With regards to individual accessibility by electronic or other means, *Fixtures Marketing Ltd v Opap* held that this was a decisive factor in setting apart a database from a mere collection[10]. The threshold appears to be low. The High Court in *Forensic Telecommunications Services Ltd v Chief Constable of West Yorkshire Police* held that a mere list of 33 articles 'through its headings and row labels … that enabled users to retrieve a particular start or end address' was sufficient[11]. Scientific databases tend to use similar arrangements to retrieve desired results and would thus likely fulfil this criterion.

4. The fourth criterion relates to the originality of a database by selection or arrangement. To be protectable by copyright according to s. 3A (2) UK CDPA the selection or arrangement of the database's contents needs to be the author's own intellectual creation[12]. The major scientific databases are arranged in such a way rather than an obvious alphabetical order. They would thus fulfil this criterion.

Databases that fail to qualify for copyright protection on the grounds that they lack sufficient originality could still be protected by the *sui generis* database right offered by s. 13(1) Copyright and Rights in the Databases Regulations 1997[13] (Db. Regs). However, even such 'unoriginal databases' would need to fulfil the criteria of being databases within the meaning of s. 3A(1)(a) UK CDPA. As discussed, scientific databases would fulfil such database criteria.

In order to be protectable by the *sui generis* database right, there must have been substantial investment in obtaining, verifying or presenting of the database's contents. In 2013 the scientific journal publisher and database owner Elsevier reported that more than one million research papers were submitted to it and that more than 10,000 editors managed the peer review of these papers, resulting in the publication of more than 350,000 articles in over 2,000 journals[14]. Such numbers suggest that substantial investment is necessary to create scientific databases, which would thus fulfil this criterion of substantial investment.

Case law raises the question of whether investment in scientific databases constitutes investment in obtaining and verifying independent materials or if it 'creates' data (*British*

Horseracing Board v William Hill[15]). In order to claim the database right the 'maker' of the database had to show substantial investment in the latter. The distinction between 'obtaining and verifying' and 'creating' data appears *prima facie* unproblematic in the case of scientific databases since that data is not 'created' in the manner of football fixtures but based on scientific measurements (*Sportradar*[16]). However, scientific publishers do not only collect articles and data for their databases but also edit thousands of journal articles, which could be regarded as 'creation'. While scientific journal articles contain considerable subjective elements, they constitute a comment on some pre-existing scientific principles and phenomena that are merely measured. Such scientific facts are not 'created' by scientists[17]'.

The *sui generis* database right is infringed if, without the consent of the database owner, all or a substantial part of the contents of the database are extracted or re-utilised (Reg. 16(1) Db. Regs). Extraction means the permanent or temporary transfer of database contents to another medium by any means or any form.

In *British Horseracing Board v William Hill*, the Court of Justice of the EU held that the quantitative meaning of 'substantial' (as per Art. 7(1) and 8(1) EU Db. Dir and s. 12(1) Db.Regs.) 'refers to the volume of data extracted from the database and/or re-utilised, and must be assessed in relation to the volume of the contents of the whole of that database'[18]. The Court also held that the quantitative meaning of "substantial" 'refers to the volume of data extracted from the database and/or re-utilised, and must be assessed in relation to the volume of the contents of the whole of that database'[19]. In the same ruling the Court further held that the qualitative meaning of 'substantial' refers to the scale of the investment in obtaining, verifying or presenting the contents of the subject of the act of extraction and/or re-utilisation, regardless of whether that subject represents a quantitatively substantial part[20]. The requirement of the substantial part, both in quantitative and qualitative terms, appears to be unproblematic in the context of DM. DM operations are mostly automated and can (and often do) extract entire databases.

Re-utilisation appears to be less relevant for DM in terms of infringing *sui generis* rights. Making databases available to the public (Reg. 12(1) Db. Regs.) does not appear to occur as a direct result of the ordinary application of DM, since the repositories or data warehouses (where the data is stored for DM purposes) are typically not made available to the public.

B. The marketing industry

DM in the marketing industry *prima facie* infringes copyright and the *sui generis* database right. While much of the data mined by the marketing sector is raw data that is not copyright protectable, marketers also mine copyright protectable content.

One example of the marketing industry's use of copyright protectable content is the DM of social media content. But before considering the copyright implications of mining this kind of content, we need to step back and address a more fundamental question: Who owns copyright protectable social media content? Facebook and Twitter, two of the biggest social media platforms, state in their terms of services (TOS) (s. 2 (1) 'Statement and Responsibilities' Facebook (SARF) and Sec. 5 'Terms of Service' Twitter (TOST)) that while the user retains intellectual property rights to any content shared on their platforms, the user grants Facebook or Twitter respectively a non-exclusive, sub-licensable worldwide licence. It is noteworthy that while Facebook does mine its users' data to offer targeted advertisements on behalf of advertisers, it does not sell data to third parties for DM purposes[21]. Twitter, however, does sell or license its data to third parties for data mining purposes[22].

However, the licensing of Twitter's data is also relevant in a non-commercial DM context, because Twitter also licenses its 'firehose' of tweets (unlike ordinary application programming interfaces, the 'firehose' grants access to all tweets in real time without any limitations on the number of search results) to a small number of academic institutions without charge[23].

The TOS used by Facebook and Twitter are so-called 'click-wrap' (CWA) licences. Before users can successfully install the required software they must click an "I agree" box to express

consent to the social media site's terms of service. But can such a broad intellectual property licence be lawfully obtained through such a CWA, given high levels of consumer protection in the EU? It appears that it can. None of the guidelines in Schedule 2 Unfair Contract Terms Act 1977 (UCTA) or the criteria in s. 5 Unfair Terms in Consumer Contracts Regulations 1999 (UCTR) and Schedule 2 UCTR or Art. 2 EU Directive on Unfair Terms in Consumer Contracts (CDUTCC) suggest that such a broad IP licences are unreasonable for consumers. It appears, therefore, that while the authors of copyright protectable content retain ownership of the copyright, social media platform providers can lawfully mine or sell the data obtained from the content.

DM operations for marketing could also 'extract' or 're-utilise' data from databases that do not enjoy copyright protection. This may happen, for example, by scraping a competing price comparison site. For reasons outlined in the pharmaceutical section above, mining databases for marketing purposes could *prima facie* infringe *sui generis* database rights.

So-called 'scraping' (the automated copying of web content) is relevant here, both for copyright and *sui generis* database content. For example, advertisers can mine (copyright protectable) tweets for consumer preferences. Similarly, price comparison websites and travel search engines such as Skyscanner scrape the websites of airlines or online travel agents (OTA) such as Expedia to obtain data (not protectable by copyright). Facebook and Twitter in their TOS (s. 3(2) SARF and s. 8 'TOST') exclude 'scraping of users' content or information' on their sites without Facebook's or Twitter's consent. The licences granted in s. 2(1) SARF and s. 5 TOST appear to be sufficiently broad to entitle Facebook and Twitter to permit external parties to utilise their users' data even beyond the usual social media activities. Both clauses use very similar wording to grant the social media platforms a worldwide, non-exclusive, sub-licensable, royalty-free licence to use any content posted on or in connection with their services. The fact that such broad licences come in the form of a CWA does not, as such, preclude them being upheld in court[24], even though data suggests that only a fraction of consumers ever read these agreements[25].

When discussing the scraping and mining of data without copyright relevance it is important to consider case law on scraping and so-called browsewrap agreements (BWA). In a BWA, users do not click a button to signal their acceptance of the terms of service. Instead, the TOS are accessible through a hyperlink. A user that moves away from the homepage is deemed to have accepted the TOS. Even under EU consumer law BWAs are not *per se* illegal[26]. However, in a consumer context an acceptance under a BWA might be regarded as unfair under the Unfair Contract Terms Directive (UCTD) Annex I (i) and UTCR Schedule (i)[27]'.

Some guidance on scraping and CWAs was provided by the Court of Justice of the EU (CJEU) in *Ryanair v PR Aviation*, which was referred to the CJEU by the Dutch Supreme Court[28] for further clarification of the interpretation of arts. 6(1), 8 and 15 Db. Dir.[29]. In this case, Dutch company PR Aviation operated a price comparison website that not only allowed customers to compare budget airline prices but also, for a fee, book flights. PR Aviation obtained the data it needed by screen-scraping the websites of budget airlines such as Ryanair. Ryanair, however, had excluded any such unauthorised use of 'automated systems or software to extract data' from its website 'for commercial purposes' in its TOS. It had also prohibited the sale or booking of Ryanair flights by any third parties. Visitors to Ryanair's website needed to consent to these TOS by means of a click-wrap agreement[30]. Ryanair argued that PR Aviation's screen-scraping and sale/booking of tickets infringed its TOS. The CJEU ruled on January 15 2015, making four key legal findings:

1. Meeting the legal definition of database does not *per se* confer protection by copyright or the *sui generis* database right[31].

2. The restriction on contractual limitations to the use of databases under art. 15 Db. Dir. does not apply to databases not protected by copyright or database rights[32].

3. Authors of such unprotected databases are therefore free to place contractual limits on the use of their databases even to lawful users[33].

4. Such contractual limitations are subject to national law[34].

This meant that Ryanair could lawfully restrict the use of screen-scraping and booking functionalities by its terms of use (TOU) without being prevented to do so by art. 15 Db. Dir. The TOU would thus only be limited by national law. Remarkably, Ryanair was in a better position by having its database protected by neither copyright nor *sui generis* database rights than it would have been with those legal protections. In other words: less IP protection resulted in more (contractual) protection for Ryanair.

C. The law enforcement sector

Data mining for law enforcement purposes *prima facie* infringes copyright and the *sui generis* database right.

The law enforcement sector mines data that is copyright protectable and data that is not copyright protectable. It is useful to provide some examples of the kind of data that law enforcers are interested in mining and to consider whether it is protected by copyright or other rights.
One obvious example is film footage. Closed circuit television (CCTV) footage and police surveillance recordings are protectable by copyright under s. 5B (1) UK CDPA.

While UK law does not require film footage to be original in order to attract copyright protection[35], originality is required for photography. Case law clarifies that for photography to be deemed original the author has to demonstrate some degree of skill[36], and make free and creative choices in the production of the photograph[37]. Photography that is produced for investigative or surveillance purposes is mostly dictated by technical imperatives[38]. The vast majority of photography in a law enforcement context would consequently not qualify for copyright protection.

Would police documents qualify as 'literary works' or be sufficiently original to be protected by copyright? Some police documents could be sufficiently original to be protectable by (Crown) copyright. For other documents, such as minutes of police interrogations, the level of originality depends on how much individual judgment went into such recordings[39]. In any case, however, it is

likely that Crown Copyright (s. 163(1) UK CDPA) would apply to most of this data. Where the Crown itself owns (or may lawfully copy content communicated to them in the course of business as per s. 48 UK CDPA) the copyright of content, then copyright does not form an impediment to DM.

With regards to data mined for law enforcement purposes from governmental data repositories, the data constitute facts and principles that are not protectable by copyright[40]. However, law enforcement databases could themselves be protectable by copyright. The deliberations above on the pharmaceutical industry regarding selection and arrangement can be applied here as well.

What about copyright protectable data that is not held by law enforcement agencies or does not otherwise fall under Crown copyright? Law enforcers mine social media data such as Twitter and Facebook. Even short social media posts can be protected by copyright if they are deemed sufficiently original. As explained above, the use of social media data is governed by basic copyright rules and the terms of service used by social media platforms.

Do law enforcement agencies consider copyright issues at all when mining data from social media platforms such as Twitter for law enforcement purposes?[41] It is difficult to provide a definitive answer because there is no case law on this issue. Nor the does academic literature offer any answers. The Regulation of Investigatory Powers Act 2000 (RIPA) mentions neither intellectual property in general nor copyright or *sui generis* database rights in particular.

However, developments in the early 1990s appear to suggest that the law enforcement industry does in fact place some importance on TOS. Some websites containing legally questionable content such as virus code posted messages at the point of access to the website stating 'law enforcement officials are not permitted to enter the system'[42]. This appears to have been effective in deterring law enforcement officials, given that in 1994 the Computer Misuse Act 1990 (CMA) was amended with s. 10(b): '…nothing designed to indicate a withholding of consent to access to any program or data from persons as enforcement officers shall have effect to make access unauthorised for the purposes of the said section 1'. However, s.10 (b) CMA is limited to mere access and would therefore not legitimise the use of 'hacking' by law enforcement agencies to obtain data that is not publicly available through

a website[43]. The only legal provision in the UK CDPA that refers to copyright in a law enforcement context is s. 296ZB (3) UK CDPA that legalises the use of devices circumventing technological protection measures (TPM) by law enforcement or intelligence agencies for investigation, detection and prevention of crimes. Neither the Facebook 'Information for Law Enforcement Authorities' (FILEA) or the 'Twitter Guidelines for Law Enforcement' (TGLE) clarify or mention the relationship between law enforcement and intellectual property.

A possible fall-back provision could be the public interest clause of s. 171(3) UK CDPA. It states that a legal rule may be introduced in order to limit copyright if it serves the public interest. While the jurisdictional interpretation of s. 171(3) UK CDPA has been restrictive,[44] in *Ashdown v Telegraph Group Ltd* the Court of Appeal of England and Wales left the public interest notion in principle open, although only within a very limited context[45]. However, even without such public interest rules, in practice law enforcers seem to ignore provisions such as the CMA when it comes to investigating grave cases such as child abuse[46].

In our discussions about the copyright exceptions that may be available to data miners, it is important to note that technological protection measures have already rendered some EU member states' copyright exceptions ineffective. For example, Art. 5(2)(b) Information Society Directive (InfoSoc) grants member states the right to introduce exceptions for 'reproductions on any medium made by a natural person for private use'. Accordingly, the Netherlands introduced exception Art. 16(b)(1) in its national copyright law ('Auteurswet'), which deals with private use. However, if a content owner decides to use TPM to prevent such copying, users cannot legally circumvent it, even though such circumvention would enable access that would be lawful under the exception. Although s. 29A (5) UK CDPA prevents parties from using contracts to undermine the new exception, the law cannot prevent rightsholders from using TPM to restrict DM of their content. This means that any new open or closed exceptions aimed at enabling data mining would need to address the issue of TPM in InfoSoc. While s. 292 ZE UK CDPA 1988 allows users to complain to the Secretary of States if TPM precludes a permitted act (for example, under an exception), this does not equal a secure right to

disable such a TPM. Furthermore, this does not appear to be feasible or scalable for researchers wanting to undertake large-scale DM operations.

5. The sui generis database right

Can law enforcers mine data that is not protectable by copyright? As discussed, the law enforcement sector mines facts such as historical crime data or official crime statistics. Such facts and principles would not be protectable by copyright. They could however, be protected by the *sui generis* database right. According to reg. 14(3) Db. Regs, where a database is made by Her Majesty or by an officer or servant of the Crown in the course of his or her duties, Her Majesty shall be regarded as the maker of the database[47].

The legal considerations on the *sui generis* database right implications of DM in the pharmaceutical sector also apply to law enforcement, including the considerations on substantial extraction/re-utilisation. Unauthorised extraction of a database in a law enforcement context could therefore be *prima facie* infringing. However, in practice infringing extraction of contents appears to be of little relevance, since law enforcement authorities appear to be mining mostly their own databases or obtain licences or warrants to commercial databases or social media sites. Furthermore, the Data Protection Act (DPA) and Freedom of Information Act (FOIA) requests provide lawful access options for external data miners. As in the case of DM in the pharmaceutical context the aspect of re-utilisation appears to be of limited relevance in the context of DM in law enforcement. Making the content of databases available to the public within the meaning of Reg. 16(1) Db. Regs. does not appear to occur as a direct result of data mining for law enforcement purposes. Researchers or journalists wishing to make law enforcement data available to the public could simply make requests under the FOIA or DPA.

6. Potential legal exceptions to unauthorised copying

After establishing in the previous section that DM in the three key industries is *prima facie* capable of infringing copyright and the *sui generis* database right, we turn to the next question: can data miners rely on any legal exceptions in defence of their activities? It appears that they cannot. This section will explain why none of the available legal exceptions suffice and demonstrate that even the newly introduced s. 29A UK CDPA (copies for text and data analysis for non-commercial research) is too narrow to give data miners legal certainty. The following section will first consider potential copyright exceptions before moving on to exceptions for *sui generis* database rights.

A. Copyright exception: temporary copying exception

One potentially applicable exception to protect data miners from a claim of copyright infringement could be s. 28A UK CDPA, which offers an exception for the making of temporary copies. However, the exception is insufficient to protect data mining, or at least the majority of DM research activities. To take advantage of this exception, the copies made in the DM process would need to fulfil the five cumulative conditions set out in the law: being 'temporary'; 'transient or incidental'; 'an integral and essential part of the technological process'; its 'sole purpose of the process being to enable a transmission network between third parties by an intermediary or the lawful use of the work or protected subject matter'; and 'having no independent economic significance'. Recital (33) InfoSoc provides some context for these conditions and refers to 'browsing' and 'caching'. It is useful here to consider how the courts have approached some of the conditions. The CJEU dealt with the first condition of being 'temporary' in *Meltwater*. The Court explained that this referred to cached on-screen copies that were deleted when an internet user moved away from a previously visited website[48].

When considering whether copies made in the data mining process are 'temporary' or not for the purposes of satisfying the exception, it is important to know if the copies made are automatically deleted and replaced by other copies, as in the case of caching. This depends on the technical process

chosen by the data miner. In the US *Ticketmaster* case, a so-called 'web crawler' software application was described that only temporarily loaded this electronic information into the random access memory (RAM) of the scraping service's computers for a period of between 10 to 15 seconds. It then extracted factual information such as events, dates and times and discarded the rest[49]. Such a *modus operandi* for DM could qualify for the temporary copy exception under UK and EU law. However, this does not appear to be the practice in most DM operations. Instead, in most DM operations the copies are placed in permanent repositories and then replaced by new copies. In general, therefore, DM does not appear to be sufficiently 'temporary' within the meaning of s. 28A UK CDPA and would not qualify for the exception.

The second condition that must be met before a data miner can take advantage of the temporary copy exception is that the copy is 'transient or incidental'. In *Meltwater*, the CJEU held that a copy is 'incidental' if it neither exists independently of, nor has a purpose independent of, the technological process of which it forms part[50]. Depending on the precise DM process used and how long the copies are stored, it is debatable whether the copies meet this test. Copies of entire journal articles stored in a repository can in fact be used independent of the DM process. Thus, the majority of the copying performed in the pharmaceutical DM context does not appear to be 'incidental'. The same would be true for marketing and law enforcement DM. In *Meltwater*, the CJEU held that to be 'transient' a copy needed to have its duration limited to what is necessary for that process to work properly and be automated in as much as it deletes automatically, without human intervention[51]. Due to the lack of automatic deletion, few DM operations across the three industries would qualify as incidental.

The third condition that must be met before a data miner can take advantage of the temporary copy exception is that the copies are an integral and essential part of a technological process. Following the reasoning of the CJEU in *Meltwater*, the copies would need to be made entirely in the context of the process of DM[52]. Most importantly, the process of DM would not function properly and efficiently without those acts. It could therefore be argued that the copying performed for the purposes of DM appears to be an integral and essential part of that technological process.

The fourth condition is that the 'sole purpose of the process being to enable a transmission network between third parties by an intermediary or the lawful use of the work or protected subject matter'. Copies made in the process of DM are not used for the purpose of enabling a transmission in any of the three industries. Is there a lawful use of the work or protected subject matter? Following the reasoning of the CJEU in *Premier League*,[53] such use should be considered lawful where it is authorised by the rightsholder and is not restricted by the applicable legislation. Lawful use is precisely what is in doubt with DM for all three industries. Thus, data miners should not rely on this condition.

The fifth condition requires that the copies have no independent economic significance. A conceivable way for the copies produced in the process of DM to gain independent economic significance would be if the data miners sold on the copied content for commercial gain. Making copied content available to third parties would be clearly infringing (for example, s. 17(1) UK CDPA). However, such bad faith uses do not form part of ordinary DM processes. Data mining in the three industries would thus not normally involve any copies of independent economic significance.

With the cumulative criteria of s. 28A UK CDPA unlikely to be met, it appears doubtful whether s. 28A could offer a defence for DM in the three industries.

B. Copyright exception: research and private study

Section 29 of the UK CDPA offers a fair dealing exception for research and private study. However, this is insufficient for most data mining activities.

S. 29 (1) UK CDPA states that fair dealing for non-commercial research does not infringe copyright. S. 29 (1C) UK CDPA states that fair dealing for private study is non-infringing. For our purposes, it is necessary to establish what would constitute 'non-commercial research'. The term 'non-commercial' could pose a problem for DM projects. Recital (42) InfoSoc specifies that the non-commercial nature of the activity in question should be determined by the nature of the activity, and not by the relevant establishment's organisational structure means of funding. Thus, research carried

out at an independent school or a private university may be non-commercial, while research at a publicly funded university can be commercial if, for instance, it pursues some commercial ends. This leads to many ambiguities for general research practice. Should, for example, research by an industry-sponsored PhD student be treated as commercial? Should research by law enforcement agencies be regarded as 'non-commercial' if it involves cooperation with private security contractors? DM for marketing purposes clearly could not claim non-commercial character.

With regards to s. 29 (1C) UK CDPA in the context of DM, what constitutes 'private study'? S. 178 UK CDPA defines 'private study' as not including 'any study which is directly or indirectly for a commercial purpose'. It is doubtful whether the copying necessary for DM would fall under the private study exception: The case *Universities UK Ltd v Copyright Licensing Agency Ltd*[54] determined that the copying of a whole textbook would not be 'fair dealing'. That suggests that the copying of large volumes of complete journal articles for data mining in the pharmaceutical industry would not constitute fair dealing. Neither the law enforcement nor the marketing industry could rely on the 'private study' exception for their activities.

C. Copyright exception: copies for text and data analysis for non-commercial research

This exception, contained in s. 29A (1) (a) UK CDPA, applies to copying done for the purposes of carrying out 'a computational analysis of anything recorded in the work'. *Prima facie* DM would be covered by the terms of 'computational analysis' and 'text and data analysis'. According to s. 29A (1) (a) UK CDPA computational analysis can be carried out on 'anything', which would include scientific journal articles and raw data repositories.

The requirement of s. 29A (1)(b) UK CDPA that the copies are accompanied by 'sufficient acknowledgement' could pose a feasibility challenge for data miners, given that they might mine thousands of journal articles. However, s. 29A (1)(b) also specifies that if such acknowledgement would be 'impossible for reasons of practicality or otherwise' this requirement would not apply. Such 'reasons of practicality' would apply to large-scale DM operations, even though there is no case law

guidance on this topic. It is therefore uncertain when the threshold of 'practicality' would be crossed. S. 29A (2)-(4) UK CDPA would only pose a problem to the applicability of the DM exception of s. 29A UK CDPA if the researchers transferred the copied journal articles to third parties or sold or rented them out for commercial gain. S. 29A (5) UK CDPA prevents publishers of scientific journals from undermining the exception of s. 29A UK CDPA via contractual agreements. Thus, licensing provisions banning lawful DM would not pose a challenge to the applicability of s. 29A UK CDPA. However, in spite of s. 29A (5) UK CDPA, journal publishers could deny DM access by means of technical protection measures. Art. 6 InfoSoc says that these cannot legally be circumvented.

A significant challenge to the applicability of the non-commercial research exception in s.29A UK CDPA is the requirement of 'lawful access'. Needing to be a lawful subscriber to the journals selected for DM poses narrow limits to DM operations. Even well-funded universities would find it challenging to subscribe to every scientific journal article in just one discipline, let alone in all disciplines. However, if scientists mine journals to which their institutions do not subscribe then their DM operations will not be protected by the exception in s. 29A UK CDPA. Furthermore, the 'lawful' criterion could also pose a problem for researchers who want to mine journals to which their institution has actually subscribed. Some journal publishers could still demand an additional licence fee for DM activity[55]. If researchers wish to mine social media, the applicable TOS would mostly preclude 'lawful' DM access.

A further challenge to the applicability of s. 29A UK CDPA is that the research be conducted for 'a non-commercial purpose'. As we have already seen, it is very difficult to draw a clear line between commercial and non-commercial research. Thus, the scope of s. 29 A UK CDPA is severely limited by the 'lawful access' and the 'non-commercial' requirements. This limitation certainly applies to the pharmaceutical industry but also to the law enforcement sector if the private sector is involved. It is clear that data mining for the purposes of marketing would not be protected by this research exception.

D. Sui generis database exception: extraction of insubstantial parts by lawful users

The database regulations entitle a lawful user of a database that has been made available to the public in any manner to extract or re-utilise insubstantial parts of the contents of the database for any purpose[56]. Reg. 19(2) Db. Regs. states that any term or condition shall be void in so far as it prevents that person from extracting or re-utilising insubstantial parts of the contents of the database. While this clearly limits the scope of DM operations, for some selective DM exercises this might be sufficient. For the pharmaceutical industry, where researchers mainly mine large volumes of journal articles and abstracts, this exception might only be of limited use. However, in the marketing industry a highly selective scraping of insubstantial parts of a database – for example from a publicly accessible price comparison website – could provide sufficient information for data miners. However, in general, it is unlikely that extracting insubstantial parts of a database could benefit the pharmaceutical, marketing or law enforcement sectors: all require substantial amounts of data to discover new correlations and form new hypotheses.

E. Sui generis database exception: extraction by lawful user of substantial part for purpose of illustration for teaching or research

Reg. 20 Db. Regs. stipulates that the database right in a database which has been made available to the public is not infringed if a substantial part is extracted by a lawful user for the purpose of illustration for teaching or research in a non-commercial purpose. The Royal Society, a scientific academy, considered the exception in reg. 20 Db. Regs. 'vague and unhelpful', in particular with concerns to its non-commercial purpose[57]. It argues that basic scientific research is 'often carried out in collaborative arrangements which can be difficult to classify as either commercial or non-commercial'[58]. In addition, the requirement to indicate sources is likely to be problematic in the context of large-scale DM operations. As already discussed, the requirement for a 'lawful' user appears most limiting to the scale of permissible DM operations, even if they could be proven to be

for non-commercial purposes. These limitations, when taken together, make it unlikely that the fair dealing exception in reg. 20 Db. Regs. would apply to DM operations.

F. Sui generis database exception: making a copy through an establishment which is accessible to the public

Reg. 12(2) Db. Regs. states that making a copy of a database available for use, on terms that it will or may be returned, through an establishment which is accessible to the public shall not constitute extraction or re-utilisation of the contents of the database, provided that there is no direct or indirect economic or commercial advantage. It is unclear what constitutes an 'establishment' since it does not refer to Art. 5(2)(c) InfoSoc. Under that provision such 'establishments' would include 'publicly accessible libraries, educational establishments or museums, or by archives, which are not for direct or indirect economic or commercial advantage'. It is also unclear whether it refers to 'dedicated terminals on the premises of establishments', in line with Art. 5(3)(n) InfoSoc. As this provision is very vague data miners in the three industries should not rely on it.

G. The Schedule 1 Db. Regs: database exceptions for public administration

Schedule 1 of the Db. Rgs. provides 'exceptions to database right for public administration'. Would those be applicable at least to law enforcement? Section 1 of Schedule 1 does not appear to be applicable as it relates to parliamentary and judicial proceedings and does not mention executive or law enforcement activities. Section 3 of Schedule 1 on 'material open to public inspection or on official register' appears to be of limited use as only a small subset of data mined by law enforcement would involve such data. The same would apply to section 4 ('material communicated to the Crown in the course of public business') and Section 5 ('public records which are open to public inspection'). A bit more complicated is Section 6 that vaguely relates to 'acts done under statutory authority'. While this in principle could include law enforcement activities as authorised by an Act of Parliament, as discussed in the previous sections on copyright exceptions, the available law remains mostly silent on

data mining in general and the relevance of IP rights for law enforcement in particular. Therefore, the exceptions available through Schedule 1 are of limited use and would only appear to apply to a very small subset of law enforcement activities.

H. Case law exceptions: database extraction by a lawful user for consultation

Case law such as *British Horseracing* clarified that once a database or a part of it was made accessible to the public, the owner's *sui generis* right would not allow him or her to prevent third parties from consulting that database[59]. The CJEU, however, cautioned that the maker of the database may still prevent the extraction and re-utilisation of substantial parts of the database's contents[60]. Could DM operations be defended legally on the grounds of being such mere 'consulting' of a database? Guidance was provided by the CJEU in *Innoweb*. In this case a meta search-engine 'translated' user queries to the search engines of specific websites. Subsequently, the meta search engine created a website displaying the results of the queries. This was stored for a limited period on the server of the operator of the meta search engine and sent to the end-user[61]. The CJEU held that this did not constitute mere consultation but re-utilisation of a substantial part of the contents of the database. The CJEU concluded that it was only the end user entering a query who consulted the meta search-engine[62]. The CJEU therefore found the operations of the meta search-engine infringing the *sui generis* database right[63]. This strongly suggests that DM operations generally would not fall under the consultation exception.

This interpretation seems to be supported by *Directmedia*. In this case, the CJEU held that transfer of data following an on-screen consultation was capable of constituting an extraction of substantial parts of a database within the meaning of Art. 7 Db. Dir.[64]. The CJEU stated that to constitute such infringing extraction of data, it is not even necessary that a physical transfer, such as photocopying or downloading, must take place[65]. Even mere manual recopying of the contents of such a database to another medium could correspond to the concept of extraction[66]. If even manual recopying following an on-screen consultation could constitute infringing extraction, it appears likely

that the automated copying of substantial parts of the contents of a database for DM purposes would not fall under the consultation exception.

Given than none of the exceptions detailed above appear to apply to DM operations in the three industries, data miners are likely to infringe *sui generis* database rights.

7. Why it matters and what can be done

Recent evidence suggests that when it comes to data mining, Europe is lagging behind the US and a number of Asian countries[67]. There may be a number of cultural and socioeconomic reasons for this, but the evidence suggests that the legislative framework in Europe, as well as European policy, is partly responsible[68]. One study entitled 'Is Europe Falling behind in Data Mining? Copyright's Impact on Data Mining in Academic Research' by Handke et al. found a substantial correlation between the use of DM and the permissiveness of a country's copyright system[69]. In general, levels of copyright protection in EU/EEA countries are relatively strong. This was found to have a negative net effect on the use of data mining[70].

So what can be done to raise levels of data mining in Europe? As we have already seen, DM is *prima facie* infringing. The legal exceptions and defences currently available are at best insufficient and at worst not applicable to DM. Below, we consider some potential solutions to enable more of this increasingly important research tool.

A. Licensing: new licence models as a solution?

If an activity is *prima facie* infringing and if there are no available exceptions then one simple solution is to obtain a licence from the rightsholder. Licences grant permission to perform 'restricted acts'[71] that would normally infringe[72].

While an in-depth analysis of licensing, especially the intricacies of Creative Commons Licensing and Open Access publishing, would go beyond the scope of this working paper, the issue of

licensing is so crucial when discussing potential solutions that its main challenges need to be highlighted here:

The current situation of licensing is one where licensing is fragmented between 28 different member state copyright regimes and collective rights organisations. The European Commission as part of its Communication of the Europe 2020 strategy has therefore emphasised the necessity to foster 'multi-territorial licences'[73] which would be 'portable'[74] to achieve a true Digital Single Market.

However, the withdrawal of many stakeholders (such as research libraries and civil society groups) from the data mining working group of the 'Licensing for Europe Structured Stakeholder Dialogue'[75] and remarks by Commissioner Kroes[76] have hinted that there might be limits to what can be achieved by solely relying on licence and thus market-based solutions.

Particularly the costs and the scalability challenge of obtaining licences could pose problems for a solely licence-based solution, for instance when an academic researcher needs to data mine 1,000 different sites from different academic journal publishers or commercial database providers[77]. Also, problems with unequal bargaining power between licensee and licensor need to be factored in when making the case for an entirely market based solution. This concern that a solely licence driven approach would lead to outcomes that would be less than socially desirable due to the high transaction costs was echoed in the report of the EU Text and Data Mining Expert group[78].

Additionally, in the literature the concern has been voiced that (over-) reliance on licensing (even open licensing provisions such as Creative Commons) could further the tendency of licences to override copyright rules and exceptions. Even in the US context of the very broad and flexible 'fair use' copyright exception the cases of *ProCD v. Zeidenberg*[79] and *Bowers v. Baystate Techs*[80] highlighted that parties could trade initially allocated rights or even waive the right of fair use even in the standard contract[81]. This has also been referred to as the privatization of the 'rule-making process regarding the use of information'[82]. The concern here is that, as in the case of the use of TPM/DRM, that this might lead to a level of protection beyond what is doctrinally intended for copyright[83]. The

recent *Ryanair* decision by the CJEU has raised concerns with regards to the degree to which content owners can use licences even if the content is not protectable by copyright or the *sui generis* database right[84].

Even if on balance licences are found to be the solution, it still merits discussion in how such licences need to be designed in order to address the issue of scalability of their use. Neither the 'Licensing for Europe Structured Stakeholder Dialogue', nor the Commission Communication 'Towards a modern, more European copyright framework' nor the (delayed) draft of the White Paper 'A Copyright Policy for Creativity and Innovation in the European Union' have, to date, yielded any concrete suggestions for guidelines on standard terms. The debate has not even gotten to a stage where the possibility of a compulsory licensing scheme appears to have been explored.

To sum up, while licensing could address some of the legal challenges currently faced by data miners, a solely licensing- (and thus market-) based solution might still be problematic from a data mining perspective, given the issues of costs, scalability, imbalance of bargaining power and 'privatisation' of rule-making.

B. Introducing broad, open exceptions: the benefits of US-style fair use

Given the lengthy legislative cycles at EU level (about 20 years) and the problems with solely licence-based approaches, one way to 'future-proof' copyright legislation for unforeseen technological developments could be to adopt a very wide exception in line with fair use in the United States. This wide exception with its four-factor test (1. purpose and character of the use, 2. nature of the copyrighted work, 3. amount and substantiality of the portion taken, 4. effect of the use upon the potential market) has repeatedly shown itself to be highly adaptive in the face of new copy-based technologies such as the VCR[85], search engines[86] and mass digitisation/digital library endeavours such as *Google Books*[87] or *HathiTrust*[88] that were unforeseen at the drafting of the US Copyright Act of 1976.

The worry that such a wide open exception would come at the expense of legal certainty has been countered by recent studies by Barton Beebe[89] and Paula Samuelson[90]. Samuelson concluded that, if grouped into certain clusters, fair use judgements became highly predictable and that this did create a high degree of legal certainty[91]. In comparison, as Senftleben has discussed, even the EU copyright model of highly specialised exceptions did not necessarily guarantee a higher degree of legal certainty. He argued that EU copyright legislation combined the worst of both worlds: the inflexibility of closed exceptions without the legal certainty that such closed exceptions are supposed to provide and introducing an open exception in Art. 5(5) that did not provide the flexibility of an open exception such as fair use[92].

Given that the degree of legal certainty of fair use has been higher than claimed by its sceptics and considering how well fair use has been able to accommodate unforeseen technologies, it appears well worth European legislators considering fair use to 'future-proof' EU copyright legislation. The knee-jerk reaction of European scholars of warning against 'systemic rejection'[93] of fair use as a legal import, should not *per se* form an argument to reject this highly adaptable exception.

This innovation-friendly flexibility with its focus on 'transformativeness' of the *prima facie* infringing use became especially apparent in the *Google Books* case where the court found fair use in spite of Google being a commercial entity and in spite of Google having digitised millions of books without the authorisation of the copyright owners. The Court held that the use was transformative in the sense that 'it has transformed book text into data for purposes of substantive research, including DM and text mining in new areas, thereby opening up new fields of research. It therefore added value to the original'.

This innovative strength of the fair use exception was reiterated by Matthew Sag in the CREATe working paper on 'Archives and Copyright: Developing an agenda for reform' in 2014:

> 'In the US it's much easier; we just say, it is fair use. It is technically copying, it is technically a reproduction, but it's not infringing.[94]'

Following this rationale of not simply focussing on copying as such but placing it the broader context of what is achieved by it and the questions raised by the four factor test makes fair use a more desirable option for EU lawmakers than introducing yet another narrow exception. Most importantly, this would be a way to future-proof legislation proactively rather than playing a futile reactive game of legal catch-up amidst ever faster technological developments.

However, given the aforementioned resistance of European legal scholars and lawmakers to such a 'wholesale import' of US-style fair use, one needs to acknowledge the *Realpolitik* of European intellectual property policy. Suggesting a legal instrument that, in spite of its tremendous benefits, is almost certain to be rejected (especially given that it does not appear to be considered in the leaked draft of the Copyright Whitepaper or the Commission Communication 'Towards a modern, more European copyright framework' [95]) would yield little benefit.

What, then, are the alternatives to 'wholesale import' of fair use?

C. Semi-open exceptions: the approach of the Wittem project

The Wittem Project was set up in 2002 as a collaboration between leading European copyright scholars. The Wittem Group had its origins in the Netherlands at the Raboud University of Nijmegen, the University of Amsterdam and the University of Leiden. It involved prominent copyright scholars such as Bently and Strowel. The main aim of the project was to improve consistency and transparency of copyright law in Europe. The Wittem Group drafted the 'European Copyright Code' as a reference document for the further harmonisation of copyright law on the European level.

Art. 5.2 of the European Copyright Code offers a semi-open/semi-closed provision. It is semi-open in the sense that unlike s. 29A UK CDPA 1988 it is not aimed at a particular technology but rather at a certain type of use. Art. 5.2 broadly applies to 'use for the purpose of freedom of expression and information' without requiring authorisation. It is semi-closed in the sense that it still demands 'only against payment of remuneration and to the extent justified by the 'purposes of scientific research' (5.2(b)). The demand for remuneration, however, raises the aforementioned

concerns about scalability and limited (financial) resources of researchers. The same limitations apply to the semi-open education exception in Art. 5.3.

Art. 5.8 addresses the problem of TPM undermining copyright exceptions, by demanding that rightsholders shall be obliged to make means available for those users benefitting from the exceptions of the 'European Copyright Code' proposed by the Wittem Project. However, Art. 5.8 demands rather narrow, cumulative conditions, such as the user having lawful access to the work, the use of the work not otherwise being possible in order to benefit from the exception(s) concerned and the rightsholders not being prevented from ensuring 'adequate measures' can be used in order to control the numbers of copies made.

While such a semi-open exception would offer greater flexibility than a highly specialised exception such as s. 29A UK CDPA, it would still raise similar uncertainties regarding the meaning of terms such as 'lawful access', 'the purposes of scientific' research or the 'remuneration' required. However, this vagueness might be due to the fact that the 'European Copyright Code' is a mere reference document and subsequent legislation might provide more details.

D. The Irish CRC: combining UK-style fair dealing and closed exceptions with fair use

An interesting new approach for exceptions was suggested by 'Modernising Copyright' report by the Irish Copyright Review Committee (CRC). In essence the report suggested the introduction of a new fair use exception in the Irish Copyright and Related Rights Act (CRRA) while at the same time it did not recommend the introduction of 'the US-style "fair use" doctrine'[96].

A number of the criteria for the finding of fair use in the report are similar to the four factor test that US courts have to apply for the finding of fair use (17 US § 107). For instance, the suggested s. 49A (3) (b) CRRA refers to the 'purpose and character of the use' while s. 49A (3)(d) CRRA refers to the amount and substantiality of the portion used. However, generally these proposed Irish CRRA provisions, especially with regards to the purpose and character of the use, tend to be considerably narrower than the US-style fair use. Also, the CRC emphasised that this fair use approach was closely

embedded into the existing exceptions that needed to be exhausted before this new fair use could be applied[97]. On top of that new fair use exception and the existing fair dealing exceptions the report also suggested a DM-specific exception (s. 50A CRRA) into the fair dealing section of the CRRA. S. 50A and S. 329 CRRA also demanded 'lawful access' (s. 50A (3)(a) and s. 329 (4)(a)). With regards to licensing, s. 329 (8) (a) stated that a licensing scheme (under s. 173 CRRA) could displace any exemption granted under s. 329 (2) (b). S. 50A (5) and s. 329 (6)(b) CRRA demanded that once the process of content-mining is completed, that the copies created in the process had to be deleted. Crucially, the Irish approach addressed the issue of TPM by ensuring that the use of TPM would not undermine DM (exception s. 50A (6) and s. 329 (7) CRRA)[98].

While the Irish CRC shows a way to introduce fair use elements into the European context without a 'wholesale import' of US-style fair use, it also shows that this would come at the price of making such a fair use exception much narrower and reducing some of its flexibility. Still, the Irish proposal could improve legal certainty for DM operations. A main concern with the Irish proposal would be the ability of licences to override its fair dealings exemptions.

E. From the original draft of the Reda Report to the Communication of the Commission

A crucial development in the debate on reducing copyright impediments to DM and introducing greater flexibility in the European context was the original draft of the so-called Reda Report[99] of the European Parliament (EP). While the original draft can be considered as very progressive and flexible, it saw its flexibility considerably reduced by numerous amendments before the final Reda Report was passed by the EP. However, even the final Reda Report was watered down by the Communication from the European Commission (EC) in December 2015 entitled 'Towards a modern, more European copyright framework'[100].

Among the crucial recommendations of the original draft of the Reda Report were:

1. Making the existing exceptions mandatory[101]. This would be a clear departure from the present structure of the InfoSoc where Art. 5 contains one mandatory exception (Art. 5 (1)

InfoSoc) compared to 20-odd optional exceptions (Art. 5 (2)-(4) InfoSoc).

2. The adoption of a flexible and open norm[102].
3. A 'technology neutral' and 'future compatible' exception[103] that would not require the drafting of a new specific exception every time a new technology emerges.
4. The need to enable data mining for all purposes, as long as a permission to read the work had been obtained[104].
5. The safeguard that TPM could not undermine the new copyright exceptions[105].

These recommendations were considerably limited in the final draft of the Reda Report. The final Reda Report no longer demanded flexible and open norms and instead contained a vague reference to a 'flexible legal base' (s. 19) and 'flexibility in the interpretation'. Furthermore, it no longer required making all exceptions mandatory across the EU[106]. Also, the TPM condition was weakened by merely demanding that TPM was disallowed vis-à-vis a private copying exemption[107]. It is noteworthy that the Final Report contains numerous references to licensing, especially territorial and multi-territorial licensing while no such reference can be found in the Draft Report[108]. This seems to reflect a preference for licensing-based solutions by the European Commission. However, reference to technology-neutral and 'future-compatible' exceptions were retained in s. 44 and 64 of the Final Report.

On 9 December 2015 the European Commission released its Communication 'Towards a modern, more European copyright framework'[109]. Like the Final Report this Communication did not insist on making all exceptions mandatory but instead contained the vague suggestion to make 'some exceptions' mandatory for member states. Interestingly, at the same time the Communication underscored the fragmentation of the European copyright rules particularly in the area of exceptions, as most exceptions in EU copyright law are optional[110]. With reference to the *Padawan*[111] case the Communication stressed that in extreme cases this resulted in exceptions that were narrower than what was permitted in EU law[112]. Remarkably, while acknowledging a deficit in competitiveness vis-à-vis DM-based research due to the absence of a clear EU provision on DM[113], the Communication

has weakened the recommendations of both the Draft Reda Report and the Final Reda Report. Instead of the wide DM exceptions of both reports, the Communication provided a very narrow, vague exception to 'public research organisations' for content that they have 'lawful access' to[114]. It is neither defined what a 'public research organisation' is nor what would constitute 'lawful access'. While acknowledged as a crucial issue by the Reda Report, no reference is made to TPM in the Communication, which would allow any potential exception to be undermined by TPM. Similarly, no reference is made to 'technology-neutral' and 'future-compatible' in the Communication.

8. Conclusion

Given the *prima facie* infringing character of DM for both copyright and the *sui generis* database right, the inadequacies of the available exceptions and the limitations of solely relying on licence-based solutions, new exceptions will play a key role in enabling DM technology.

As we have shown, one should not create a false dichotomy between needing to choose between full-on US-style fair use openness and flexibility on the one hand and narrow, closed exceptions on the other. There is a sliding scale at the legislators' disposal even if they conclude that wholesale import of US-style fair use for perceived or real systemic compatibilities is not feasible.

As the evidence and scholarly analyses cited in this working paper have shown, Europe is falling behind in the crucial technology of DM. It is therefore in Europe's competitive interest to err on the side of openness, even if fair use provisions seem out of reach. The evolution from the original Draft Reda Report to the latest Communication of the Commission seems to suggest that a closed, mandatory DM-specific exception rather than a technology-neutral, future-proof solution seems to be the most likely outcome of this EU copyright reform. Still, even if that is the best outcome researchers can expect from reform, it is still crucial to push as far as possible in the direction of openness and legal certainty to maximise the gains for DM in research in Europe. That means either defining vague terms such as 'lawful access' as specifically as possible to avoid legal uncertainty or to try to make

this concept as broad as possible, such as pushing it into the direction of 'read to read equals the right to (data) mine'.

Furthermore, as to be expected in a review focussed on copyright, the Reda Report and the Commission Communication remain mute on the question of *sui generis* database rights. However, as already discussed in this working paper and as pointed out by the two reports[115] on DM commissioned by the European Commission in 2014, if one does intend to remove or at least remedy legal impediments to DM in Europe then the issue of the *sui generis* database right must be addressed[116]. The copyright reform related documents should have at least mentioned this.

Finally, as narrow as the final DM exception may be, it is crucial that it is not undermined by TPM and/or overridden by licensing. Thus, any DM exception should clarify that it cannot be displaced by licences or TPM.

References

[1] JISC Report, *The Value and Benefits of Text Mining* (2013) pp. 3, 17, 19, 25, 27 available at http://www.jisc.ac.uk/publications/reports/2012/value-and-benefits-of-text-mining.aspx [Accessed April 28, 2016].

[2] This is summed up very well by the title of one of the earliest DM works on that subject: Don R Swanson, "Two Medical Literatures That Are Logically but Not Bibliographically Connected" (1987) 38 *Journal Of The American Society For Information Science* 228, 221.

[3] Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data A Revolution That Will Transform How We Live, Work, and Think* (Houghton Mifflin Harcourt 2013) 58 Kindle location 870; Charles Duhigg, "How Companies Learn Your Secrets" *New York Times* (NYC, NY, USA, 16 February 2012) available at http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all&_r=0 [Accessed March 29, 2016].

[4] Tom L Beauchamp and James F Childress, *Principles of Biomedical Ethics* (5th edn, Oxford University Press 2001) p. 77.

[5] Jonas Lerman, "Big Data and Its Exclusions" (2013) 66 *Stanford Law Review Online* 59 available at http://www.stanfordlawreview.org/online/privacy-and-big-data/big-data-and-its-exclusions [Accessed April 1, 2016].

[6] ibid.

[7] *Fixtures Marketing Ltd v Organismos prognostikon agonon Podosfairou AE (OPAP)* (C-444/02) [2004] European Court of Justice at [29].

[8] Directive 96/9/EC on the legal protection of databases [1996] OJ L 77.

[9] *Fixtures Marketing v OPAP* [2004] (n 7) at [30].

[10] *Fixtures Marketing v OPAP* [2004] (n 7) at [31].

[11] *Forensic Telecommunications Services Ltd v Chief Constable of West Yorkshire Police* [2011] EWHC 2892 Ch (High Court of Justice of England and Wales, Chancery Division) 88.

[12] *Football Dataco Ltd and others v Yahoo! UK Ltd and others* (C-604/10) [2012] Court of Justice of the European Union, at [24].; *Bezpečnostní softwarová asociace – Svaz softwarové ochrany v Ministerstvo kultury* (C 393/09) [2010] European Court of Justice at [50].; *Eva-Maria Painer v Standard VerlagsGmbH* (C 145/10) [2011] Court of Justice of the European Union at [92].

[13] These Regulations implement the European Database Directive 96/9/EC. The terms of those Regulations are identical to those of the Directive, cp. *The British Horseracing Board Ltd and Others v William Hill Organization Ltd* (C-203/02) [2004] European Court of Justice at [9].

[14] Reed Elsevier Annual Report, *Reed Elsevier Annual Reports and Financial Statements 2013*' (2013) p. 16 available at http://www.relx.com/investorcentre/reports%202007/Documents/2014/relxgroup_ar_2014.pdf [Accessed April 2, 2016].

[15] *British Horseracing Board Ltd v William Hill Organisation* (C-203/2) [2004] European Court of Justice at [31].

[16] *Football Dataco Ltd and others v Sportradar GmbH And Another* [2013] EWCA Civ 27 (Court of Appeal of England and Wales) 57.

[17] *Football Dataco v Sportradar* [2013] 57.

[18] *British Horseracing Board Ltd v William Hill Organisation* (n 15) [2004] at [70].

[19] *British Horseracing Board Ltd v William Hill Organisation* (n 15) [2004] at [70].

[20] *British Horseracing Board Ltd v William Hill Organisation* (n 15) [2004] at [71].

[21] *Facebook/ WhatsApp* (Case COMP/M.7217) Commission Decision 7239 [2014] at [70] available at http://ec.europa.eu/competition/mergers/cases/decisions/m7217_20141003_20310_3962132_EN.pdf [Accessed April 30, 2016].

[22] Juliette Garside, "Twitter Puts Trillions of Tweets up for Sale to Data Miners-Company Plans to Make Content Generated by Users Available to Commerce, Academia and Even Police Involved in Crowd Control" *The Guardian* (London, England, 18 March 2015) available at http://www.theguardian.com/technology/2015/mar/18/twitter-puts-trillions-tweets-for-sale-data-miners [Accessed March 20, 2016].

[23] ibid.
[24] Philippa Davies, "Access v Contract: Competing Freedoms in the Context of Copyright Limitations and Exceptions for Libraries" (2013) 35 *European Intellectual Property Review* 8.

[25] Yannis Bakos, Florencia Marotta-Wurgler and David R Trossen, "Does Anyone Read the Fine Print? Consumer Attention to Standard Form Contracts" (2014) NYU Center for Law, Economics and Organization Law & Economics Research Paper Series, Paper 195, 22 available at http://lsr.nellco.org/cgi/viewcontent.cgi?article=1199&context=nyu_lewp [Accessed April 25, 2016] .

[26] European Commission Expert Group on Cloud Computing Contracts, *Unfair Contract Terms in Cloud Computing Service Contracts Discussion Paper* (2014) p. 3 available at http://ec.europa.eu/justice/contract/files/expert_groups/discussion_paper_unfair_contract_terms_en.pdf [accessed May 1, 2016].

[27] ibid.

[28] *Ryanair Limited v PR Aviation BV* [2014] Hoge Raad der Nederlanden 12/03167.

[29] *Ryanair Ltd v PR Aviation BV-Request for a preliminary ruling from the Hoge Raad der Nederlanden (Netherlands)*(C-30/14) [2014] Court of Justice of the European Union.

[30] *Ryanair Ltd v PR Aviation BV* (C-30/14) [2015] Court of Justice of the European Union at [16].

[31] *Ryanair* [2015] at [35].

[32] *Ryanair* [2015] at [37].

[33] *Ryanair* [2015] at [39].

[34] *Ryanair* [2015] at [44].

[35] Smita Kheria and others, "(Mis)appropriation Art? Copyright and Data Protection Implications of 'CCTV Sniffing' as Art." [2013] CREATe Working Paper No.2/March 2013, 8 available at http://www.create.ac.uk/wp-content/uploads/2013/03/CREATe-Working-Paper-No-2-v1.0.pdf [Accessed April 27, 2016].

[36] *Antiquesportfolio.com Plc v Rodney Fitch & Co Ltd* [2000] FSR 23 (High Court of Justice of England and Wales) 347.

[37] *Eva-Maria Painer v Standard VerlagsGmbH* [2011] at [92]; *Temple Island Collections Ltd v New English Teas Ltd* [2012] FSR 9 (Patents County Court) 20.

[38] *Temple Island Collections Ltd v New English Teas Ltd* [2012] 66.: '…an image which is nothing more than the result of happening to click a camera in the right place at the right time…'.

[39] '…more than negligible skill and labor…', s. *Ladbroke (Football) Ltd v William Hill (Football) Ltd* [1963] 1 WLR 273 (House of Lords) 287; '…if skill, labour and judgement was put into the reporting of…words in the…newspaper, copyright will subsist…', s. *Express Newspapers pc v News (UK) Ltd* [1990] FSR 359 (High Court of Justice of England and Wales-Chancery Division) 43; It is, however, crucial to underscore the traditional 'skill, labour and justice' doctrine has come under pressure by the case law of the Court of Justice of the European Union, especially since the *Infopaq* decision, s. *Infopaq International A/S v Danske Dagblades Forening* (C5/08) [2009] European Court of Justice (ECJ) ; cp. Estelle Derclaye, "Infopaq International A/S v Danske Dagblades Forening (C-5/08): Wonderful or Worrisome? The Impact of the ECJ Ruling in Infopaq on UK Copyright Law" (2010) 32 *European Intellectual Property Review* 248 ; Andreas Rahmatian, "Originality in UK Copyright Law: The Old 'Skill and Labour' Doctrine under Pressure" (2013) 44 *International Review of Intellectual Property and Competition Law* 5–6.

[40] See Recital (45) Db. Dir.

[41] Garside (n 22).

[42] Ian Walden, *Computer Crimes and Digital Investigations* (Oxford University Press 2007) p. 317.

[43] ibid.

[44] *Hyde Park Residence Ltd v David Yelland* [2000] RPC 604 (Court of Appeal of England and Wales).

[45] *Ashdown v Telegraph Group Ltd* [2001] EWCA Civ 1142 (Court of Appeal of England and Wales) 257.

[46] Keith McDevitt and Steven Wilson, "Cybercrime- the Challenge of Investigations in the Cyber World" (Presentation, University of Glasgow, December 4, 2013).

[47] Hugh Laddie, Peter Prescott and Mary Vitoria (eds), *The Modern Law of Copyright and Designs*, vol 1 (4th edn, Reed Elsevier UK Ltd 2011) 1665.

[48] *Public Relations Consultants Association Ltd v Newspaper Licensing Agency Ltd and Others* (C 360/13) [2014] Court of Justice of the European Union at [26].

[49] *Ticketmaster Corp v TicketsCom, Inc* [2003] 2003 WL 21406289 (United States District Court, CD California) 2.

[50] *Public Relations Consultants v Newspaper Licensing Agency* (n 48) at [43].

[51] *Public Relations Consultants v Newspaper Licensing Agency* (n 48) at [15], [46].

[52] *Public Relations Consultants v Newspaper Licensing Agency* (n 48) at [28], [29].

[53] *Football Association Premier League Ltd v QC Leisure* (C-403/08) [2011] *Murphy v Media Protection Services Ltd* (C-429/08) [2011] In Joined Cases (C-403/08, C-429/08) [2011] Court of Justice of the European Union, at [168].

[54] *Universities UK Ltd v Copyright Licensing Agency Ltd* [2002] R.P.C. 36 (Copyright Tribunal United Kingdom).

[55] It is difficult to verify how challenging it is for users to obtain such authorisations. Statements made by publishers and researchers seem to partly contradict one another, s. Jean-Paul Triaille, Jérôme de Meeûs d'Argenteuil and Amélie de Francquen, "Study on the Legal Framework of Text and Data Mining (TDM)" (2014) De Wolf & Partners, funded by the European Commission, p. 45.; cp. Independent Report UK Intellectual Property Office (2012) *Copyright Works-Streamlining Copyright Licensing for the Digital Age* p. 31.; Eefke Smit and Maurits van der Graaf, "Journal Article Mining-A Research Study into Practices, Policies, Plans…..and Promises." Commissioned by the Publishing Research Consortium,5 available at http://publishingresearchconsortium.com/index.php/128-prc-projects/research-reports/journal-article-mining-research-report/160-journal-article-mining [Accessed 15 April 2016]. (This study had been commissioned by the journal publishing industry).

[56] Reg. 19(1) Db. Regs.

[57] The Royal Society Report, *Keeping Science Open: The Effects of Intellectual Property Policy on the Conduct of Science* (2003) s 5.5 available at https://royalsociety.org/~/media/Royal_Society_Content/policy/publications/2003/9845.pdf [Accessed May 3, 2016].

[58] ibid.

[59] *British Horseracing Board Ltd v William Hill* [2004] (n 15) at [55].

[60] *British Horseracing Board Ltd v William Hill* [2004] (n 15) at [58].

[61] *Innoweb BV v Wegener ICT Media BV* (C 202/12) [2013] Court of Justice of the European Union at [10]–[11].

[62] *Innoweb* [2013] at [47].

[63] *Innoweb* [2013] at [54].

[64] *Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg* (C 304/07) [2008] European Court of Justice at [60].

[65] *Directmedia* [2008] at [37], [40].

[66] *Directmedia* [2008] at [37].

[67] Christian Handke, Lucie Guibault and Joan-Josep Vallbe, "Is Europe Falling Behind In Data Mining? Copyright's Impact On Data Mining In Academic Research" (2015) Empirical Paper, 6 available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2608513 [Accessed April 30, 2016] ; European Commission Report from the Expert Group, *Standardisation in the Area of Innovation and Technological Development, Notably in the Field of Text and Data Mining* (2014) p. 35 available at http://ec.europa.eu/research/innovation-union/pdf/TDM-report_from_the_expert_group-042014.pdf [Accessed May 1, 2016].

[68] Handke, Guibault and Vallbe (n 67) 19.; The Lisbon Council Report, *Mapping Text and Data Mining in Academic and Research Communities in Europe* (2014), pp. 4, 15 available at http://www.lisboncouncil.net/publication/publication/109-mapping-text-and-data-mining-in-academic-and-research-communities-in-europe.html [Accessed April 24, 2016].

[69] Handke, Guibault and Vallbe (n 67) 19.

[70] ibid 20–21.

[71] s. 16 UK CDPA
[72] S. 16(2) UK CDPA
[73] European Commission, "Communication From The Commission Europe-2020 A Strategy for Smart, Sustainable and Inclusive Growth" COM(2010) 2020 final 14.

[74] European Commission, "Communication From The Commission-On Content in the Digital Single Market" COM(2012) 789 final 3.; Michel Barnier, "Licences for Europe: Quality Content and New Opportunities for All Europeans in the Digital Era" (Launch of the initiative "Licences for Europe", Brussels, February 4, 2013) 3 available at http://europa.eu/rapid/press-release_SPEECH-13-97_en.htm?locale=en [Accessed April 28, 2016].

[75]Letter from European Technology SMEs, Open Access Publishers and the Research Sector WG 4 to Commissioners Barnier, Geoghegan-Quinn, Kroes and Vassiliou (22 May 2013) available at https://archive.org/stream/705015-letter-of-withdrawall4e-tdm-may-24-4/705015-letter-of-withdrawall4e-tdm-may-24-4_djvu.txt [Accessed 15 April 2016].

[76] "…But keep your minds open: maybe in some cases licensing won't be the solution. Maybe it will be provided by technology and data, like the Global Repertoire Database or some wider initiative not yet on the table..", s. .Neelie Kroes, "Digital Technology and Copyright Can Fit Together " (Stakeholders Dialogue on Copyright "Licences for Europe", Brussels, February 4, 2013) available at http://europa.eu/rapid/press-release_SPEECH-13-96_en.htm [Accessed April 27, 2016].

[77] The chemist and data miner Peter Murray Rust summarised this as follows: 'Licensing destroys Text and Data Mining … Imagine as few as 1000 researchers negotiating licences with 1000 publishers. That's 1 million licence negotiations…', s. Peter Murray-Rust, "Text and Data Mining – Fighting for Our Digital Future ('Peter Murray-Rust Is the Problem')" (*petermr's blog*, October 2,

2013) available at: http://blogs.ch.cam.ac.uk/pmr/2013/10/02/text-and-data-mining-fighting-for-our-digital-future-peter-murray-rust-is-the-problem/ [Accessed May 1, 2016].

[78] European Commission Report *Standardisation in the Field of Text and Data Mining* (2014) (n 67) 52.

[79] *ProCD Incorporated v Matthew Zeidenberg and Silken Mountain Web Services* [1996] 86 F.3d 1447 (United States Court of Appeals, Seventh Circuit).

[80] *Baystate Technologies, Inc v Bowers* [2008] 283 Fed.Appx. 808, 2008 WL 2704484 (C.A.Fed. (Mass.) (United States Court of Appeals, Federal Circuit).

[81] Guy A Rub, "Contracting Around Copyright: The Uneasy Case of Unbundling of Rights in Creative Works" (2011) 78 *University of Chicago Law Review* 258.

[82] Séverine Dusollier, "The Master's Tools v. The Master's House: Creative Commons v. Copyright" (2006) 29 *Columbia Journal of Law & the Arts* 282–83.

[83] As it was discussed with regards to *Ryanair* [2015] (n 30).

[84] *Ryanair* [2015].

[85] *Sony Corp of America v Universal City Studios, Inc* [1984] 464 US 417 (Supreme Court of the United States).

[86] *Perfect 10, Inc v Amazon.com, Inc* [2007] 508 F3d 1146 9th Cir 2007 (United States Court of Appeals for the Ninth Circuit).

[87] *The Authors Guild, Inc et al v Google Inc* [2013] 05 Civ. 8136 DC (United States District Court of Southern New York) , affirmed by *Authors Guild v. Google, Inc.* [2015] 13-4829-cv (United States Court Of Appeals For The Second Circuit).

[88] *Authors Guild, Inc v HathiTrust* [2014] 755 F3d 87 2d Cir 2014 (United States Court of Appeals For the Second Circuit).

[89] Barton Beebe, "An Empirical Study Of U.S. Copyright Fair Use Opinions, 1978–2005" (2008) 156 *University of Pennsylvania Law Review*.

[90] Pamela Samuelson, "Unbundling Fair Uses" (2009) UC Berkeley Public Law Research Paper No. 1323834 available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1323834 [Accessed May 2, 2016].

[91] ibid.

[92] Martin RF Senftleben, "Bridging the Differences between Copyright's Legal Traditions-The Emerging EC Fair Use Doctrine" (2010) 57 *Journal of the Copyright Society of the U.S.A.* 529 available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1723902 [Accessed May 2, 2016].

[93] Bernt Hugenholtz and Martin RF Senftleben, "Fair Use In Europe. In Search of Flexibilities" (2011) Amsterdam Law School Research Paper, at http://ssrn.com/abstract=1959554 [Accessed May 2, 2016].

[94] Ronan Deazley and Victoria Stobo (eds), "Archives and Copyright: Developing an Agenda for Reform" (2013) CREATe Working Paper 2014/4-CREATe Working Paper Series, 90 available at http://zenodo.org/record/8423/files/CREATe-Working-Paper-2014-04.pdf [Accessed April 30, 2016] .

[95] European Commission, "Communication From The Commission To The European Parliament, The Council, The European Economic And Social Committee And The Committee Of The Regions-Towards a Modern, More European Copyright Framework" COM(2015) 626 final.

[96] Report Irish Copyright Review Committee for the Department of Jobs, Enterprise and Innovation *Modernising Copyright-A Report* (2013) p. 176 available at https://www.djei.ie/en/Publications/Publication-files/CRC-Report.pdf [Accessed April 29, 2016].

[97] ibid 177.

[98] ibid 87.

[99] Draft Report European Parliament Rapporteur Julia Reda, *On the Implementation of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the Harmonisation of Certain Aspects of Copyright and Related Rights in the Information Society 2014/2256(INI)* (2015) available at http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+COMPARL+PE-546.580+02+DOC+PDF+V0//EN [Accessed April 29, 2016].
[100] European Commission "Communication Towards a Modern, More European Copyright Framework" COM(2015) 626 (n 95).

[101] Draft Reda Report (n 99) 11.

[102] Draft Reda Report (n 99) 13.

[103] Draft Reda Report (n 99) 14.

[104] Draft Reda Report (n 99) 18.

[105] Draft Reda Report (n 99) 23.

[106] Julia Reda, "Reda Report Adopted: A Turning Point in the Copyright Debate" (*Julia Reda*, June 6, 2015) available at https://juliareda.eu/2015/06/reda-report-adopted-a-turning-point-in-the-copyright-debate/ [Accessed April 28, 2016].

[107] Report European Parliament Rapporteur Julia Reda , *Harmonisation of Certain Aspects of Copyright and Related Rights-European Parliament Resolution of 9 July 2015 on the Implementation of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the Harmonisation of Certain Aspects of Copyright and Related Rights in the Information Society P8_TA-PROV(2015)0273* (2015) s 57 available at http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+TA+P8-TA-2015-0273+0+DOC+PDF+V0//EN [Accessed April 29, 2016] ; cp. the changes to the original Draft Report as outlined on Julia Reda's website: Julia Reda, "EU

copyright evaluation report-full current text" (*Julia Reda*, July 9, 2015) available at https://juliareda.eu/copyright-evaluation-report/full/ [Accessed April 30, 2016].

[108] Only s. 21 of the Draft Reda Report (n 99) contains reference to a 'statutory licenses'.

[109] European Commission, "Communication From The Commission-Towards a Modern, More European Copyright Framework" COM(2015) 626 final (n 95).

[110] ibid 6.

[111] *Padawan SL v Sociedad General de Autores y Editores de Espania (SGAE)* (C-46708) [2010] Court of Justice of the European Union.

[112] European Commission, "Communication From The Commission-Towards a Modern, More European Copyright Framework" COM(2015) 626 final (n 95) 6 ; cp. *Eva-Maria Painer v Standard Verlags-GmbH* [2011] (n 37).

[113] European Commission, "Communication From The Commission-Towards a Modern, More European Copyright Framework" COM(2015) 626 final (n 95) 7.

[114] ibid 8.

[115] Triaille, d'Argenteuil and de Francquen "Study on the Legal Framework of Text and Data Mining (TDM)" (2014) (n 55); European Commission Report, *Standardisation in the Area of Innovation and Technological Development, Notably in the Field of Text and Data Mining* (2014) (n 67).

[116] European Commission Report, *Standardisation in the Area of Innovation and Technological Development, Notably in the Field of Text and Data Mining* (2014) (n 67) 14, 52, 54, 70; Triaille, d'Argenteuil and de Francquen "Study on the Legal Framework of Text and Data Mining (TDM)" (2014) (n 55) 85, 106.

RCUK Centre for Copyright and
New Business Models in the
Creative Economy

College of Social Sciences / School of Law
University of Glasgow
10 The Square
Glasgow G12 8QQ
Web: www.create.ac.uk