

CRIDES Working Paper Series

31 October 2018

Limitations to Text and Data Mining and Consumer Empowerment *Making the Case for a Right to “Machine Legibility”*

by Rossana Ducato and Alain Strowel

Cite as: R. Ducato, A. Strowel, Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to “Machine Legibility”, CRIDES Working Paper Series, 31 October 2018

The **CRIDES**, *Centre de recherche interdisciplinaire Droit Entreprise et Société*, aims at investigating, on the one hand, the role of law in the enterprise and, on the other hand, the function of the enterprise within society. The centre, based at the Faculty of Law – UCLouvain, is formed by four research groups: the research group in economic law, the research group in intellectual property law, the research group in social law (Atelier SociAL), and the research group in tax law.

www.uclouvain.be/fr/instituts-recherche/juri/crides



This paper © Rossana Ducato and Alain Strowel 2018 is licensed under a
Creative Commons Attribution-ShareAlike 4.0 International License
<https://creativecommons.org/licenses/by-sa/4.0/>

LIMITATIONS TO TEXT AND DATA MINING AND CONSUMER
EMPOWERMENT:
MAKING THE CASE FOR A RIGHT TO “MACHINE LEGIBILITY”

Rossana Ducato^{*} and Alain Strowel[†]

ABSTRACT

The paper focuses on the current legal barriers to text and data mining (TDM) in the context of smart disclosure systems (SDSs) whose aim is to provide consumers with improved access to the data needed to make informed decisions. The use of intellectual property rights and contracts, combined with technological protection measures, can hinder TDM and the deployment of SDSs. Further, those legal constraints can negatively impact artificial intelligence innovation that requires improved access to data. There are thus various arguments for enhanced “machine legibility”. However, the TDM exception included in the draft Copyright in the DSM Directive and the various amendments proposed by the European Parliament or the Council do not appear to clear the way for enhanced “machine legibility”. In relation to SDSs, we also argue that the principle of transparency, embedded in consumer and data protection laws, can serve as a last line of defence against prohibition of TDM.

KEYWORDS

Text and data mining – Copyright – Sui generis right – Exceptions – Consumer protection – Data protection – Smart disclosure systems

^{*} **Rossana Ducato** (rossana.ducato@uclouvain.be) is a postdoctoral researcher in Law at UCLouvain and Université Saint-Louis – Bruxelles. <http://www.rosels.eu/member/rossana-ducato/>.

The research is supported by the Innoviris grant 2016-BB2B-9. A sincere thanks to dr. Guido Noto La Diega for our constructive discussion on an early draft of this paper.

[†] **Alain Strowel** (alain.strowel@uclouvain.be) is professor at UCLouvain, Université Saint-Louis – Bruxelles, KULeuven, Munich IP Law Centre. <http://www.rosels.eu/member/alain-strowel/>

1. Introduction

Smart disclosure refers “to the timely release of complex information and data in standardized, machine readable formats in ways that enable consumers to make informed decisions”.¹ Smart disclosure systems (SDSs) allow users to get easy and timely access to the relevant pre-contractual information or even receive personalised advices based on their preferences.² Over the last two years, several initiatives have emerged worldwide offering third party services to automatically analyse websites’ contractual documents and to check their compliance with applicable consumer and data protection laws.³ One of the main goals of these projects⁴ is to increase the awareness of users towards the rights, obligations and possible risks in their online transactions, trying to reduce or overcome the well-known signing-without-reading problem.⁵ Such tools are primarily directed to consumers, as end-users of the service; however, other possible users are consumer associations or regulatory authorities, which can use them to perform periodical assessments, start investigations or verify complaints more quickly.

The functioning of SDSs is based on text and data mining (TDM). TDM uses techniques from natural language processing, machine learning, information retrieval, and knowledge

¹ Sunstein 2012.

² On the advantages of smart disclosures and targeted information, see Ben-Shahar 2009; Helberger 2013; Porat and Strahilevitz 2013; Bar-Gill 2015; Busch 2016; Helleringer and Sibony 2017; Busch forthcoming.

³ One of the first projects in the area of automated analysis of legal documents is “Usable Privacy Policy” (www.usableprivacy.org), a consortium led by Carnegie Mellon University. Their tool aims to help users to navigate through the text of privacy policy and identify the privacy options and choices available (Sadeh et al. 2013). More recently, an international team formed by researchers from the Switzerland's Federal Institute of Technology, the University of Wisconsin and the University of Michigan, has launched two tools: Polisis (<https://pribot.org/polisis>), a tool to visualise in a very effective way the content of a privacy policy, and Pribot (<https://pribot.org/bot>), a chatbot available to answer questions about a specific privacy policy (Harkous et al. 2018). In the field of the automated analysis of T&C, we must mention CLAUDETTE, a research project carried out by an interdisciplinary team at the European University Institute (<https://claudette.eui.eu>). The tool, based on machine learning techniques, assesses the fairness of consumer standard terms (<https://claudette.eui.eu/use-our-tools/>). This functionality will be extended to the analysis of privacy policy. For more information, see Contissa et al. 2018; Lippi et al. 2018. Another interdisciplinary project, SaToS (Software Aided Analysis of ToS), is conducted by the chair of Software Engineering for Business Information Systems (Sebis) at TU Munich. The German research group is developing a solution to automatically identify Terms of Services of e-commerce websites and summarise the key points of the contract in a simplified language (Braun et al. 2018).

⁴ This is precisely one of the objects of “The Internet of Platforms: an empirical research on private ordering and consumer protection in the sharing economy”, carried out at UCLouvain. The project aims to address the issue of the lack of transparency in sharing economy transactions and improve the information users receive from and about the platform (<http://www.rosels.eu/research/research-project-iop/>). This paper presents some of the preliminary results of this project.

⁵ On the users’ tendency not to read contracts, the literature is extensive. *Ex multis*, Vila et al. 2003; Wilhelmsson 2004; Hillman 2006; Ben-Shahar 2013; Radin 2013; Ayres and Schwartz 2014; Bakos et al. 2014.

management for the automated analysis of digital content (structured and unstructured data), in order to extract information, identify patterns, discover new trends, insights or correlations.⁶

Despite the development of such promising tools for enhancing the readability and understandability of the conundrum of terms, the current EU legal framework is not particularly supportive of TDM. Many rigorous studies have already analysed the barriers to TDM, current tensions and negative externalities in the context of research, contributing to the debate on the TDM exception in the ongoing copyright reform.⁷ Several of these studies highlight that the beneficial uses of TDM are not limited to scientific research, but take place in other contexts, including consumer information and protection.

In the light of the artificial intelligence (AI) applications that are emerging, it is easy to predict that TDM will become a central technique enabling new kinds of information-based services and applications (e.g., for care and medical purposes, fact checking, disaster prevention, elaboration of sustainability policies and politics, etc.). Many studies have already highlighted the need for a broad access to datasets so as to train algorithms and improve AI applications.⁸

The article aims to contribute to the discussion on TDM by adding a further perspective. We take into consideration the application of TDM in SDSs, a sector that is growing in importance. The automated analysis of contracts and privacy policy for enhancing the awareness of consumers and, ultimately, ensuring consumer empowerment, is a perfect lab to test TDM's *pierres d'achoppement*.

The article is structured as follows: following the present Introduction, we outline in Section 2 a taxonomy of the possible obstacles for TDM and SDSs, identifying intellectual property rights (IPRs), contracts and technological protection measures (TPMs) as the principal ones.

In Section 3, we analyse the interplay between copyright, the database *sui generis* right and what is called private ordering⁹, noting that the existing copyright limitations may offer little help to counterbalance the power of contracts.

⁶ In the proposal for a Directive on copyright in the Digital Single Market (hereafter “Copyright in the DSM Directive”), TDM is defined as “any automated analytical technique aiming to analyse text and data in digital form in order to generate information such as patterns, trends and correlations” (Art. 2.2, Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market, COM/2016/0593 final - 2016/0280 (COD)). The definition is sufficiently broad to embrace the current TDM application panorama. For a technical definition of text and data mining, see Hearst 2003. Specifically on text mining, Feldman and Sanger 2007. For an extensive analysis of the definition of TDM, see Triaille et al. 2014.

⁷ Ibid.; Triaille et al. 2014; Bernhardt et al. 2015; Caspers and Guibault 2016a; Margoni and Dore 2016; Stamatoudi 2016; Hilty and Richter 2017; Geiger et al. 2018; Margoni and Kretschmer 2018; Rosati 2018.

⁸ On the crucial need to train algorithms on different datasets, see Hall and Pesenti 2017.

⁹ By “private ordering”, we refer to both contractual, technological and informal measures as tools to enforce platforms’ rights and interests towards their users. In the absence of a clear legislative framework or effective (and efficient) remedies, contracts and technology can be used to expand the prerogatives and powers of

Therefore, in Section 4 we explore whether the current proposal for a specific TDM exception in the Copyright in the DSM Directive¹⁰ and the ongoing copyright reform in the EU will fill some of these gaps. Despite some merit of the text proposed by the European Commission (and of the amendments of the European Parliament), the exception as drafted will not permit to fully embrace the potentialities of Big Data analytics and AI, nor will it dispel legal uncertainty, as contemplated by the European Commission.

In Section 5 we go further and argue that the TDM and related data access issue should be framed in a perspective that takes copyright's rationale seriously. This leads to the argument that the reproduction right should in the first place not cover TDM processes.

The limitations to TDM coming from private ordering will be examined in Section 6, where we present the results of an empirical analysis conducted on a representative set of online platforms operating in the sharing economy. This analysis shows that there is a widespread trend to preclude TDM via contracts (the online Terms & Conditions) and through the embedded code. A prohibition of TDM making it impossible to run tools for automated contractual analysis would prevent consumers to access the relevant information to make informed choices. This would not be in line with the principle of transparency enshrined in both consumer and data protection.

In Section 7 we conclude by presenting the principle of "machine legibility" which transposes the transparency principle in the technological context.

2. Obstacles to TDM in the current European framework: setting the scenario for smart disclosure systems

There are at least three legal protections and tools that could limit the practice of TDM: 1) intellectual property rights (IPRs); 2) contracts; 3) data protection rules. First, at the light of the "black letter" of copyright law and its current interpretation by the Court of Justice of the EU (ECJ), many TDM activities could be considered as copyright infringements or violations of the database *sui generis* right.¹¹ Second (and this is an equally worrisome signal in the never-ending battle for the control of information), contracts may expressly prohibit or limit TDM. In addition, TPMs can enforce (and reinforce) IPRs and contractual provisions, impeding TDM in practice. Finally, if the object of the mining consists of personal data, i.e. "any information

platforms, restricting the legitimate uses and faculties of the weaker party. As noted with reference to the intellectual property domain by Dussolier 2007, p. 1393-1394.

¹⁰ Art. 3, Proposal for Copyright in the DSM Directive and the amendments presently discussed in the European Parliament.

¹¹Triaille et al. 2014; Margoni and Kretschmer 2018; Stamatoudi 2016; Montagnani and Aime 2017; Strowel 2018.

relating to an identified or identifiable natural person”¹², the processing has to be compliant with data protection law.

These three legal instruments have different scopes and objectives but, in some cases, they produce the same consequences for TDM. Imagine a mining activity carried out by an insurance company on a national electronic health record system: TDM would be forbidden and several legal instruments could be invoked to justify it, in particular the protection of personal data. In other cases, data protection and copyright could admit the mining for research purposes (as both protections are subject to limitations for research), but TPMs, impeding the bulk download of the content, could affect the conduct of research in practice. Meanwhile, some mandatory copyright exceptions and lawful uses cannot be overridden by contractual provisions. It is therefore a complex and dynamic scenario that needs to be further explored if we want to unleash the potential of SDSs and, beyond, AI applications.

In the case of SDSs, the mining covers legal documents, such as Terms & Conditions (T&C) and privacy policy. Therefore, the issue of data protection will be left out from the analysis. Furthermore, for the purpose of this paper, the focus will be on IP and contractual obstacles.

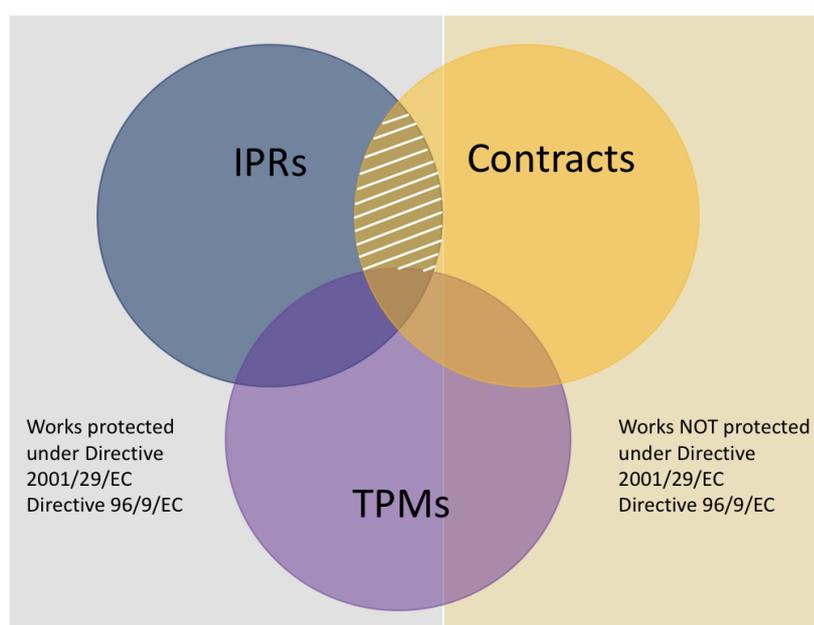


Fig. 1. Representation of limits to TDM in the context of smart disclosure systems.

¹² Art. 4(1), General Data Protection Regulation.

3. Before the (IP) Law: limits and counter-limits to TDM

IP is one of the first barriers to TDM. The latter may, in principle, clash with a bundle of exclusive rights if the work or the database qualify for protection under Directive 2001/29/EC (“InfoSoc Directive”) or Directive 96/9/EC (“Database Directive”).

As known, copyright protects “databases which, by reason of the selection or arrangement of their contents, constitute the author’s own intellectual creation”.¹³ A database is defined as “a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means”.¹⁴ A website, like that through which online platforms offer their service (for example, the AirBnB website), can certainly fall into this notion.¹⁵ Copyright covers the database’s peculiar “expression”, i.e. the originality of its systematic organization, which is exteriorised through the “free and creative choices”¹⁶ that show the “personal touch”¹⁷ of the author. *A contrario*, the requirement of originality is not satisfied “when the setting up of the database is dictated by technical considerations, rules or constraints which leave no room for creative freedom”.¹⁸ Therefore, copyright does not apply when the selection and arrangement follow a chronological or alphabetical ordering.¹⁹

Most websites of the sharing economy platforms that can be considered as databases will probably not reach the threshold of originality required for copyright to protect the database as such (to be distinguished from the content of its pages). First of all, because the selection and arrangement of the data is essentially shaped by the kind of service that the platform provides. A carpooling platform will allow users to search for the closest car available, displaying, for example, the distance, the location of the car, the level of fuel, the license plate. In addition, to be user-friendly and easy to retrieve, some information has to be organised according to “trivial” criteria: e.g. T&C, privacy policy, copyright notices, FAQs, norms of the community are generally included in the “Legal conditions” section. Therefore, in the given scenario, copyright is unlikely to apply to the platform’s database.

However, the database can be protected under the *sui generis* right regime. The latter is an exclusive protection granted to the maker of the database “which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification

¹³ Art. 3, Database Directive. For a comprehensive overview, see, Beunen 2007; Derclaye 2008, 2014.

¹⁴ Art. 1(2), Database Directive.

¹⁵ For the classification of a website as a database, see Strowel and Derclaye 2001, p. 311-312.

¹⁶ ECJ, Case C-604/10, *Football Dataco Ltd and Others v Yahoo! UK Ltd and Others* [2012], ECLI:EU:C:2012:115, para. 38. Derclaye 2012; Rosati 2013b.

¹⁷ ECJ, Case C-604/10, *Football Dataco Ltd*, para. 38.

¹⁸ *Ibid.*, para. 39.

¹⁹ In line with the US leading case *Feist Publications Inc. v. Rural Telephone Service Co.*, 499 U.S. 340 (1991). See, Waelde et al. 2013, p. 65.

or presentation of the contents”.²⁰ We could debate whether there is substantial investment in obtaining data, if the latter are constantly produced by the web-users (e.g., personal information, localisation data, pictures, preferences). Equally, one could wonder whether there is substantial investment in the verification of the content if the platforms or sharing economy websites explicitly state and follow the policy that they do not check the accuracy of the data provided by users. At the same time, there are rooms to allege that a substantial investment could lie in the presentation of the content: after all, the platform arranges the website to facilitate the sharing of information between the users, adjusts the filters and search tools to customise and make appealing the experience for the end-user.

If we assume for the sake of argument that the threshold of substantial investment is met, the way the sui generis right is conceived and designed could affect the behaviour of the legitimate user. Like copyright, the sui generis right does not need any formality to exist.²¹ At the same time, the substantial investment in either one of the three activities listed in Art. 7 (Directive 96/9/EC), as interpreted by the ECJ²², is a determination that is likely to be verified in court only and, considering the requested proof, the evidence is in the hands of the maker of the database. This means that, unlike copyright, it will be difficult for a user to know, by simply consulting the database, whether the latter is protected by the sui generis right.²³ Hence, the sui generis right framework can contribute to legal uncertainty, reducing in practice users’ faculties that would be totally legitimate: a user who wants to avoid the risk of private sanctions (the suspension of the account or the ban from the platform) or legal actions, will adopt a precautionary approach. She will act as if the sui generis right protects the database (by the way, the standard T&C will confirm that the provider of the online platform intends to protect the website by IPRs).

Another issue is whether copyright protection applies to the T&C and/or the privacy policy, which will be primarily investigated in the case of automated analysis of contractual documents. There is some discussion as to the copyrightability of such documents. In a way, they are “standard” by definition and their formulas are usually based on templates. Furthermore, in the case of privacy policy, its structure is mainly determined by law (see, for instance the list of mandated disclosures in Arts. 13 and 14 of the General Data Protection Regulation, “GDPR”). However, the EU criterion of originality does not require novelty or a high level of creativity. It

²⁰ Art. 7, Database Directive.

²¹ Derclaye 2008, p. 107.

²² ECJ, C-46/02, *Fixtures Marketing Ltd v. Oy Veikkaus Ab* [2004] ECLI: ECLI:EU:C:2004:694; C-338/02, *Fixtures Marketing Ltd v. Svenska Spel AB* [2004] ECLI:EU:C:2004:696; C-444/02, *Fixtures Marketing Ltd v. Organismoa Prognostikon Agnon Podosfairou AE (OPAP)* [2004] ECLI:EU:C:2004:697; C-203/02, *The British Horseracing Board Ltd and Others/William Hill Organization Ltd* [2004] ECLI:EU:C:2004:695.

²³ The maker of the database could state the existence of the sui generis right in the T&C, but such a circumstance is quite rare. Usually, T&C contain general formula such as “All right reserved”, leaving ample room for interpretation to the end-user, which in the majority of the cases is nor a lawyer, much less an IP expert.

suffices to demonstrate the “author’s own intellectual creation”²⁴, which can be rather modest.²⁵ Often there is still room for choices and adjustments in the presentation and drafting of those documents, which means they could be protected by copyright.

There is no case law concerning the standard of originality applied to contractual texts. Therefore, it is necessary to refer to the general framework and to reason in *consimili casu*. For instance, in *Infopaq I*, the Court of Justice stated that words, as such, are not protectable, but “through the choice, sequence and combination of those words [...] the author may express his creativity in an original manner and achieve a result which is an intellectual creation”²⁶. Thus, the ECJ concluded that even eleven consecutive words can potentially “express the author’s own intellectual creation”²⁷, leaving such a determination to national courts. The latter have ruled, episodically, over similar issues, i.e. the creativity of legal or technical works. In the Italian case law, for example, a technical article (a sort of machine’s instruction manual) describing the functions of a monoscope has been considered original.²⁸ In another case dealing with legal guidelines, the District Court of Venice found that “a regulation against counterfeiting” written by a lawyer presented the degree of originality required by copyright.²⁹ In Spain, the Madrid Provincial Court concluded the same in relation to an exercise book of mathematical problems in statistics.³⁰ Under Belgian and French copyright laws, instruction manuals and other informational documents have been recognized as protected by copyright.³¹ For instance, the text of a patent before its official publication³² as well as the wording of various contracts³³ have been

²⁴ As known, the originality requirement is not harmonised in the InfoSoc Directive (it is mentioned in the Software, Term and Database directives). However, the concept has been interpreted as “the author’s own intellectual creation” and applied to works in several decisions of the European Court of Justice. See, ECJ, C-5/08, *Infopaq International* [2009] ECLI:EU:C:2009:465, para. 37; C-403/08, *Football Association Premier League and Others* [2011] ECLI:EU:C:2011:631, para. 97; C-393/09, *BSA* [2010], ECLI:EU:C:2010:816, paras. 44-45; C-355/12, *Nintendo and Others* [2014], ECLI:EU:C:2014:25, paras. 21-22. For an overview of the “Europeanization” of originality, Strowel 2012.

²⁵ Rosati 2013a. See also ECJ, Case C-161/17, *Land Nordrhein-Westfalen v Dirk Renckhoff* [2017] ECLI:EU:C:2018:634 (on a photograph of the city of Cordoba).

²⁶ ECJ, Case C-5/08 *Infopaq International*, para. 45.

²⁷ *Ibid.*, para. 48.

²⁸ District Court of Milan, specialised section, decision no. 6057/2014, available online here: https://www.giurisprudenzadelleimprese.it/wordpress/wp-content/uploads/2014/08/20140512_RG79952-20111.pdf.

²⁹ District Court of Venice, specialised section, decision of 17 December 2014, RG 1522/2011 (not published).

³⁰ Madrid Provincial Court (s. 12) of 3 march 2004, cited in Vallés 2009, pp. 114-115.

³¹ See Bernault et al. 2017, p. 116, footnote 94.

³² Paris Criminal Court, 17 January 1968, Gaz. Pal. 1968, I, p. 197.

³³ Paris Commercial Court, 67ème ch., 4 Sept. 1989, Expertises, 1991, p. 273, obs. Gross (contract proposed by a credit provider to traders); Paris Court of Appeal, 4th ch., 27 Nov. 2002, Expertises, 2003, p. 190 (software licence); Douai Court of Appeal, 27 March 2013, Propr. Intell., 2013, p. 285, obs. Brugière (terms of reference for public procurement).

protected in France. In Belgium, the instructions for using IT equipment³⁴ have been granted copyright protection. In the UK, a wide range of subject matter has been protected as compilations in the past, including a leaflet conferring information about herbicides.³⁵ Applying a low threshold of originality, courts have accepted as original railway tables and exam papers.³⁶ In Germany, various decisions of the Federal Supreme Court (BGH) have considered as protected by copyright the following technical documents: user guidelines for the usage of a technical apparatus³⁷, technical rules to be applied in the construction of roads³⁸, technical drawings.³⁹

Therefore, the outcome of the intellectual effort of the author in drafting the various clauses of some online T&C, the choice of words or the structuring of the document, could qualify as protected work. Who has ever written a complex contract knows that it can be a creative task, and that many free choices have to be made for the organisation of the clauses and the drafting of each. In addition, thanks to some legislative interventions and a few recent scandals in the field of consumer and data protection⁴⁰, there is a growing trend towards the encouragement of user-friendliness in legal documents: e-commerce sites, social networks and online platforms in general are starting to provide T&C and privacy policy in novel formats, using different fonts, layout and icons to increase the understandability, transforming the “legalese” in a plainer language, etc. The drafters of the text of the *creative commons* licenses claim they are protected by copyright, and made them available under the CC0 Public Domain Dedication.⁴¹ In exceptional cases, designers are even involved in the drafting, this circumstance making it hard to contest the originality of the legal documents.⁴²

To sum up, to retrieve the relevant information, a SDS can automatically analyse contracts and privacy policies, considered as original works, and the various legal sections of a website could be considered as part of a database, potentially protected by the sui generis right.

³⁴ Brussels Court of Appeal, 28 January 1997, cited in de Visscher and Michaux 2000, p. 31, footnote 125.

³⁵ Supreme Court of Judicator – Court of Appeal, *Elanco v. Mandops* [1980] RPC 213.

³⁶ Bently and Sherman 2009, p. 64.

³⁷ Federal Supreme Court, 10 October 1991 - I ZR 147/89 (“Bedienungsanweisung”).

³⁸ Federal Supreme Court, 11 April 2002 - I ZR 231/99 (“Technische Lieferbedingungen”).

³⁹ Federal Supreme Court, 22 September 1999 - I ZR 48/97 (“Planungsmappe”).

⁴⁰ The speech of Senator Kennedy at the Congress hearing of Mark Zuckerberg for the Cambridge Analytica affair has become viral: “Here's what everybody's been trying to tell you today, and – and I say this gently. Your user agreement sucks [...] The purpose of that user agreement is to cover Facebook's rear end. It's not to inform your users about their rights. Now, you know that and I know that. I'm going to suggest to you that you go back home and rewrite it. And tell your \$1,200 an hour lawyers, no disrespect. They're good. But – but tell them you want it written in English and non-Swahili, so the average American can understand it. That would be a start”.

⁴¹ See, point 5 of the Creative Commons Terms, <https://creativecommons.org/terms/>.

⁴² See, for instance, the privacy policy elaborated by the designer Stefania Passera: <https://juro.com/policy.html>.

In case of IP protection, the TDM activity might come into conflict with the various exclusive rights included in each IP bundle. Depending on the technique used, TDM may involve:

- the reproduction and/or communication to the public of a content (the text of T&C/privacy policy);⁴³
- the extraction and/or reuse of a substantial part of the database.⁴⁴

A communication to the public or a reuse do not always occur in the case of TDM: the latter usually elaborates the information and publishes the results of the analysis in the form of aggregate data, statistics, reports, etc. Therefore, unless the output of the TDM shows the whole or the excerpts of the protected work or the database, there will be no communication to the public or reuse.⁴⁵

The most problematic issue with TDM is the broadness of the notions of reproduction (for copyright) and extraction (for the database right): when the tool has to run the analysis, it copies all or part of the work, it transfers all or a substantial part of the contents of a database to another medium or it technically adapts or translates the content (e.g. conversion from a PDF to another format).⁴⁶ So, these operations that are necessary steps in the TDM process in principle fall under copyright or under the database right.

The right of reproduction belongs to the copyright owner, while the right of extraction (even a temporary one, like the visualisation on a computer's screen) of the whole or substantial part of a database is granted to its maker. This means that the user cannot perform TDM without the authorisation of the right holder or in the absence of a copyright exception.

The interplay between IPRs, contracts and TPMs is visible in the field of copyright exceptions and limitations. Under the current EU framework (mainly defined by the InfoSoc Directive), some exceptions could apply to TDM: for works protected under copyright, for example, Art. 5(1) of the InfoSoc Directive allows for the temporary reproduction of works, "which are transient or incidental [and] an integral and essential part of a technological process and whose sole purpose is to enable: (a) a transmission in a network between third parties by an intermediary, or (b) a lawful use of a work or other subject-matter to be made, and which have no independent economic significance"; Art. 5(2)(b) admits the reproduction for personal use;

⁴³ We do not consider the distribution right here, because it applies when the work is incorporated in a tangible article only (see, recital 28, Directive 2001/29/EC), which is not our case.

⁴⁴ TDM may potentially perform also the reproduction and/or adaptation and/or communication and/or distribution of the database itself. However, we do not analyse this case here.

⁴⁵ In the same sense, Caspers and Guibault 2016a; Triaille et al. 2014.

⁴⁶ Triaille et al. 2014, p. 32. According to the authors, this is not the case if the tool simply spots one or two words through the text without making a copy of the work (e.g., spotting and counting the occurrences of the word "malaria"). In particular, p. 31. Similarly, Stamatoudi 2016, p. 1261; Montagnani and Aime 2017, pp. 379 ff.

and Art. 5(3)(a) includes a specific exception for teaching and scientific research. A similar exception for research is included in the Database Directive.⁴⁷ The maker of the database cannot prevent the legitimate user from extracting insubstantial parts (qualitatively or quantitatively considered), as long as she does not “perform acts which conflict with normal exploitation of the database or unreasonably prejudice the legitimate interests of the maker of the database”⁴⁸ and “cause prejudice to the holder of a copyright or related right in respect of the works or subject matter contained in the database”.⁴⁹

However, such exceptions may not confer effective rights to the consumers: they are narrow, do not cover all the spectrum of TDM technologies, and are differently implemented in Member States. Besides, the framework of rights and exceptions under the InfoSoc Directive and the Database Directive is not completely homogeneous: what would be allowed under copyright is not necessarily allowed under the sui generis right (and vice versa).

The research exceptions, for example, are limited for the (sole)⁵⁰ purpose of illustration for teaching or scientific research.⁵¹ TDM would be permitted under these exceptions in very few cases, e.g. in the context of non-profit scientific projects dealing with SDSs or for educational purposes in order to demonstrate the functioning of the tool. Such exception arguably could also cover the “training” of machine learning systems, on the model of human teaching, but such use would nevertheless require to be justified by the non-commercial purpose to be achieved through TDM and the indication of the source.⁵²

The exception for private copying will allow even more limited uses: as reported in Caspers and Guibault (2016), national laws have been quite restrictive in the implementation of

⁴⁷ We do not take into consideration Art. 9 (a) Database Directive, since the exception for personal use applies only to non-electronic database, which do not permit TDM in any case.

⁴⁸ Art. 8(2), Database Directive.

⁴⁹ Art. 8(3), Database Directive.

⁵⁰ As required by Art. 5(3)(a) InfoSoc Directive and Art. 6(1)(a) Database Directive, but not by Art. 9 (b) Database Directive. See, Derclaye 2008, p. 131.

⁵¹ For a complete analysis of the activities that could fall within the exception, Triaille et al. 2013, p. 359 ff.; Guibault et al. 2012, p. 49 ff.; Montagnani and Aime 2017, pp. 385 ff.

⁵² See Arts. 5(3)(a) InfoSoc Directive and Arts. 6(2)(b) and 9(1)(b) Database Directive. Another inconsistency between the two Directives must be noted with regard to the research exception. Under the Database Directive, the reference to the source appears to be a mandatory requirement, while the InfoSoc admits the possibility of not indicating the author's names if “this turns out to be impossible” (Art. 5.3.a. InfoSoc Directive). According to some authors, this difference is more a declamation than a substantial matter, considering the general principle of “ad impossibilia nemo tenetur”, which will be applicable to the exception for database in any case (see Montagnani and Aime 2017, p. 387, citing Walter and Von Lewinski 2010). Other authors interpret literally the provisions (cf. Triaille et al. 2014, p. 70, as reported in Montagnani and Aime 2017, p. 387).

such a copyright limitation, focussing on “personal use, study or (small scale) research”⁵³, and “sometimes the scope is limited to a few copies”⁵⁴. Considering for example Italian law, the reproduction for private use is legitimate with reference to printed works and as long as made manually or with means that do not allow the distribution of the work (Art. 68, Law 633/1941).⁵⁵ According to the majority opinion in the literature, the exception refers to the reproduction used in the family circle.⁵⁶ The exception could be extended to online works if the three steps test is respected (Art. 71-nonies, Law 633/1941), but it has received a narrow interpretation by national courts so far.⁵⁷ Therefore, the Italian private copying exception is rather limited in the digital environment and may not likely permit TDM performed by an individual or, *a fortiori*, by a consumer association.

Furthermore, to rely on those exceptions of the InfoSoc Directive is no guarantee as both the research exception and the private copying exception are optional and not imperative: Member States are free to adopt them and, where they exist, the exceptions can be overridden by contracts or TPMs.⁵⁸

A stronger defence for TDM could, in principle, come from Art. 5(1) of the InfoSoc Directive. In fact, the temporary reproduction is a mandatory exception that must be implemented by all Member States. However, there are some drawbacks. First, mandatory does not necessarily mean that the exception is imperative (non-overridable by contracts). When the European Legislator wanted to limit the freedom of contracts, it expressly did so (see, for instance, Art. 15, Database Directive, and Art. 8, Directive 2009/24/EEC, “Software Directive”).⁵⁹ There is nothing similar in the InfoSoc Directive that would prevent the

⁵³ Caspers and Guibault 2016a, p. 34. See, also Helberger and Hugenholtz 2007. The private copying exception has traditionally received little attention in the literature and the case law, apart from the issues related to the copyright levies and fair compensation. See, Strowel 2015.

⁵⁴ Caspers and Guibault 2016a, p. 34.

⁵⁵ Valenti 2007b, p. 195.

⁵⁶ Ibid. p. 202, para. III and the bibliography thereby cited.

⁵⁷ Despite the potential “open-ended” nature of the three steps test. As reported by Margoni 2012. See also Hilty et al. 2008. While, for phonograms and videograms, there is a specific provision: the reproduction is permitted if done by a physical person solely for personal use and for non-commercial purposes, in compliance with the applicable TPMs (Art. 71-sexies (1), Law 633/1941). The exception will not apply if the reproduction is done by a third party (Art. 71-sexies (2), Law 633/1941) and if the works are available on-demand and protected by TPMs or contracts (Art. 71-sexies (3), Law 633/1941). On the limits of the private copying exception for the digital context and the interplay with TPMs, cf. Caso 2004; Montagnani 2007; Mazziotti 2008.

⁵⁸ See Derclaye and Favale 2010; Helberger et al. 2013; Triaille et al. 2013; Caspers and Guibault 2016a. Regarding TPMs, Art. 6(4) does not require that the Member States take appropriate measures to ensure that the beneficiaries of the private copying exception do in practice benefit of the exception.

⁵⁹ Derclaye and Favale 2010, p. 90.

prohibition of such use by contract.⁶⁰ Nevertheless, some countries, like Belgium, Ireland and Portugal, have expressly excluded the possibility to overcome the temporary reproduction exception via contractual means.⁶¹ The situation is not crystal clear in other Member States.⁶²

A second (and more preclusive) problem concerns the content of the exception: Art. 5(1) is crafted for caching and browsing activity⁶³, but the cumulative conditions set out in that provision, as restrictively interpreted by the ECJ⁶⁴, hardly apply in TDM activities. In particular, the copies made through TDM are not necessarily temporary, transient or accessory.⁶⁵ It is also far from clear that the “sole purpose” of this reproduction is “to enable a transmission in a network between third parties”. Furthermore, a smart disclosure system does not necessarily meet the independent economic significance element. The latter not only requires that the temporary reproduction does not have to generate “an additional profit, going beyond that derived from lawful use of the protected”⁶⁶ item, but the reproduction also does not have to “lead to a modification of that work”⁶⁷. Literally interpreted, this last requirement will exclude most of TDM activities, since the data analysis process usually implies a transformation of the original work for making it processable by the machine (e.g., the conversion from a format to another one).⁶⁸ Therefore, there are several limitations of the temporary copy exception that make it irrelevant for exempting TDM.

Lastly, one has to consider whether TDM could fall in the lawful uses recognised by the Database Directive at Art. 6(1) and 8. These rights are expressly protected against conflicting contractual provisions (Art. 15, Database Directive). Art. 6(1) allows the lawful user to make a copy of the database in order to access the contents or to allow a “normal use” of the same. It would be coherent with the rationale of the Directive to consider a SDS aiming to analyse the terms or the privacy policy as performing a normal use of the database.⁶⁹ Meanwhile, with

⁶⁰ The conflict between freedom of contracts and copyright limitations is extensively investigated in Guibault 2002. Even if with reference to the proposal for the InfoSoc Directive, the issue was already pointed out by Hugenholtz 2000.

⁶¹ Kretschmer et al. 2010, p. 13.

⁶² In Italy, see Art. 68-bis, Law 633/41. Cf. Valenti 2007a.

⁶³ Recital 33, InfoSoc Directive.

⁶⁴ ECJ, C-5/08, *Infopaq International A/S v Danske Dagblades Forening* [2009], ECLI:EU:C:2009:465; C-403/08, *Football Association Premier League Ltd and others v QC Leisure* [2011], ECLI:EU:C:2011:631; case C-302/10, *Infopaq International A/S v Danske Dagblades Forening* [2012], ECLI:EU:C:2012:16; case C-360/13, *Public Relations Consultants Association Ltd v Newspaper Licensing Agency Ltd and Others* [2014], ECLI:EU:C:2014:1195.

⁶⁵ Triaille et al. 2014.

⁶⁶ ECJ, C-302/10, *Infopaq International A/S v Danske Dagblades Forening*, para. 54.

⁶⁷ Ibidem.

⁶⁸ Triaille et al. 2014, pp. 31-32.

⁶⁹ The “normal use” of the content of the database has been established, for example, in the national proceedings of the Ryanair case (ECJ, Case C-30/14, *Ryanair Ltd v PR Aviation BV* [2015], ECLI:EU:C:2015:10). In that case, the Netherlands court stated that the activity of an online intermediary

reference to Art. 8 Database Directive, the TDM tool would likely extract and mine information from an insubstantial part of the database (the “Legal conditions” section), without conflicting with the normal exploitation of the database or unreasonably prejudicing the legitimate interests of the maker or the author of the works or subject matter contained in the database. At the same time, in some cases it will be difficult to allege this, as an online platform like Uber has 48 different legal documents (between T&C and a variety of “contractual” policies) that will not qualify as insubstantial content.

Therefore, the Database Directive leaves some room to perform TDM, especially because it expressly protects statutory permitted uses from contrary contractual provisions. However, this balance of interests is ensured as long as there is a protection by the Database Directive. As the *Ryanair* decision established, if the database is not protected either under copyright or the sui generis right, the database owner can set down contractual limitations to its use.⁷⁰

4. TDM exception and the draft Copyright in the DSM Directive: a new hope?

If the current system does not fit for the purpose of TDM, shall a specific exception – like the one contained in the currently debated proposal for a Directive on Copyright in the DSM – be able to restore the balance between IPRs, on the one hand, and the access to information and encouragement of AI innovation, on the other?

The current proposal arrives at the end of a process initiated by Commission with the Communication for a Digital Single Market Strategy for Europe (2015)⁷¹, followed by the Communication “Towards a modern, more European copyright framework” (2015)⁷², and confirmed in the Communication “Promoting a fair, efficient and competitive European copyright-based economy in the Digital Single Market”.⁷³ In those preparatory documents, the

comparing the prices of flights, by performing also the extraction of information from the Ryanair website, was a “normal use” of that database, thus any contrary contractual provision considered unenforceable. However, the ECJ noted that the existence of the sui generis right was not proven in the case: as a consequence, the prohibition of contractual overriding did not apply.

⁷⁰ ECJ, Case C-30/14, *Ryanair Ltd v PR Aviation BV*, paras. 44-45. Cf. Borghi and Karapapa 2015; Myska and Harasta 2016.

⁷¹ Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, “A Digital Single Market Strategy for Europe”, COM/2015/0192 final.

⁷² 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/2015/0626 final.

⁷³ Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, “Promoting a fair, efficient and competitive European copyright-based economy in the Digital Single Market”, COM/2016/592.

focus of the Commission has always been on the need to promote innovation in research, threatened by an uncertain legal framework for TDM and national differences.

In the September 2016 proposal for a Directive on Copyright for the Digital Single Market (“Draft Copyright in the DSM Directive” or “Draft Directive”), the Commission therefore introduced a specific exception for TDM. The exception expressly applies to the acts of reproduction and extraction contemplated in Art. 2, InfoSoc Directive (right of reproduction), Art. 5(a), Database Directive (temporary or permanent reproduction of the database), Art. 7(1), Database Directive (extraction of the whole or a substantial part of it), Art. 11(1) Draft Directive (right of reproduction and making available to the public, recognised to publishers of press publications).

The proposal essentially allows research organisations to text and data mine works to which they have lawful access for the purposes of scientific research. Research organisations are defined as “a university, a research institute or any other organisation the primary goal of which is to conduct scientific research or to conduct scientific research and provide educational services: (a) on a non-for-profit basis or by reinvesting all the profits in its scientific research; or (b) pursuant to a public interest mission recognised by a Member State; in such a way that the access to the results generated by the scientific research cannot be enjoyed on a preferential basis by an undertaking exercising a decisive influence upon such organisation”. Notably, Member States should not provide for compensation for rightholders as regards uses under the TDM exception (see Recital 13) and any contractual provision limiting TDM shall be unenforceable. The norm is therefore both mandatory (for the Member States) and imperative (for private parties), and there is no way to circumvent it through private ordering relying on contracts.

During the legislative procedure, the Council has specified the scope of the TDM and its relationship with the proposal.⁷⁴ First of all, by making clear that TDM is not always a copyright relevant activity: “in relation to mere facts or data which are not protected by copyright [...] no authorisation is required under copyright law” (Recital 8a).⁷⁵ Secondly, by reaffirming that acts already covered by temporary reproduction exception under the InfoSoc Directive will continue to benefit from that provision (Recital 8a).

Notably, the Council extends the TDM exception so as to include cultural heritage institutions (“*cultural heritage institution* means a publicly accessible library or museum, an archive

⁷⁴ See the version of the text dated 25 May 2018 (Council of the EU, Interinstitutional File: 2016/0280(COD), doc. 9134/18, available here: <https://www.consilium.europa.eu/media/35373/st09134-en18.pdf>).

⁷⁵ Here the Council tries to fix the wording of Recital 8 (Commission text). The latter has been criticised in the literature for being a potential source of confusion, since it “wrongly suggests that carrying out TDM is per se of relevance to copyright. The explanations given in Recital 8, according to which an authorisation to undertake such acts must be obtained from rightholders if no exception or limitation applies, are too sweeping”. Hilty and Richter 2017, p. 3. However, the clarification offered by the Council could not be sufficient: the temporary reproduction exception is only one of the possible legitimate activities that can be lawfully performed by users without authorisation nor a specific TDM exception.

or a film or audio heritage institution”).⁷⁶ It added a security requirement in a new Art. 3(1a), specifying also that the copies of the works and other subject-matter generated through TDM shall not be retained for longer than necessary for achieving the purposes of scientific research (see also Recital 11c).⁷⁷

The text of the Council adds a further exception for TDM (new Art. 3a): Member States are free to allow the “temporary reproduction and extraction of lawfully accessible works and other subject-matter that form an integral part of the process of text and data mining”, if such use “has not been expressly reserved” by the rightholder “including by technical means” (see also Recital 13a). This seems to make the exception dependent on the rightholder’s willingness to accept it, and the application of technical means would be enough to express it. Contrary than the TDM exception provided in the initial text (Art. 3), it is an optional exception for the Member States. This is not welcome as it will reinforce the risk of possible divergences between Member States on an issue that should be dealt with seamlessly across the EU internal borders. Furthermore, the additional TDM exception, if implemented, would be overridable (as Art. 6(1) of the Commission’s draft has not been extended to cover the newly proposed Art. 3a). Anyway, the possibility for rightholders to make some reservation risks to subordinate the legislative exception to some private will. On the positive side, the scope of such additional exception is much broader than the one proposed by the Commission as its beneficiaries go well beyond research institutions and, eventually, cultural heritage institutions. Furthermore, it is true that Art. 3a mirrors the wording and the content of the temporary reproduction exception at Art. 5(1) InfoSoc Directive (“temporary reproduction”, “form an integral part of the process”), but with one important difference: the economic independent significance is out of the picture.

The discussion within the European Parliament (EP) showed the genuine concerns by some EP members about the narrow scope of the TDM exception.⁷⁸ However, many of the most

⁷⁶ Defined at Art. 2(3), Draft Directive.

⁷⁷ The wording of the Recital is far from being clear. On the one hand, by recognising the importance of peer review and verification, the proposal seems to allow the retention of the copies made under the exception “in certain cases” (not specified). Such copies must “be stored in a secure environment and not be retained for longer than is necessary for the scientific research activities”. The text leaves to the Member State the task to determine the concrete modalities for retaining the copies. However, the hard issue to determine is: when will the copies no longer be necessary? It is interesting to note a parallelism: the GDPR has ensured through several provisions that data should not be stored longer than necessary with respect to the purpose of the processing. However, scientific research is one of the cases that justifies the possibility to make an exception to that rule (see, Art. 5(1)(e), GDPR). Interestingly, the scientific research can trump storage limitations when the data subjects’ interests are at stake, but not in the IP domain when data aggregators prerogatives are involved.

⁷⁸ See, for instance, amendments proposed to the text of Art. 3 of the Commission proposal: amendment 538 by Julia Reda, Nessa Childers, Max Andersson, Michel Reimon, Brando Benifei (deleting any reference to research organisation, research purposes and lawful access), amendment 539 by Jytte Guteland (extending TDM to cultural heritage institutions), amendments 546 and 547 (respectively encouraging and obliging Members States to allow research organisations, without lawful access to works and other subject-matter, to perform TDM), amendment 548 (protecting the mandatory TDM exception against TPMs) and amendments 551-555 (limiting the scope of the measures that the rightholder can adopt to ensure the security and integrity of the networks and databases where the works or other subject-matter are hosted), amendment 564

favourable amendments to TDM were not tabled in the version the Committee on Legal Affairs (JURI) presented to the EP plenary for the vote of July, 5.⁷⁹

The JURI report reflected the dual approach followed by the Council in the text adopted in May 2018, proposing a mandatory TDM exception in favour of research organisations for the purpose of scientific research, provided that they have lawful access to the works or other subject-matter (Art. 3), and an optional exception available to anyone as long as the rightholder has not reserved the use of works and other subject-matter in a machine readable format (Art. 3a). Query what those reservations in a machine readable format are and how they could be implemented. The main reservation we have with this part of the EP resolution on the copyright reform is that the TDM exceptions could still be overridden by contracts (Art. 3a) and/or technical means (Arts. 3 and 3a).

Furthermore, the JURI version added two twin provisions at both Art. 3 and Art. 3(a), allowing Member States to introduce respectively mandatory or optional TDM exceptions in accordance with Art. 5(3)(a) InfoSoc Directive, which refers to the teaching and scientific research exception for original works and other subject-matter. These provisions have been probably designed for preserving the TDM exceptions already adopted by some Members States (for instance, UK, Estonia, France, Germany). However, such legislative choice is likely to fragment the European legal framework for TDM, allowing the blossoming of diverse national solutions and increasing the divide between TDM acts over works and databases (since the possibility refers uniquely to the research exception under the InfoSoc Directive).

Finally, with reference to the boundaries of the copyright relevant activity in TDM processes, the EP version introduced a new specific provision at Recital 8a. Dealing with the technical aspects of TDM, such amendment states that TDM as mere “reading and analysis of digitally stored, normalised information” is not a copyright relevant act. Copyright may come into play only in case of reproductions or extractions linked to the access and process of “information normalisation”. The latter is not expressly defined, but as it emerges from the same recital, it refers to the preparatory activities which enable the automated computational analysis, such as the change of the format of information or the extraction from a database into another one that will be subjected to TDM. The exceptions provided in the Draft Directive, therefore, will cover only these activities. Apparently, this recital seems to restrict the scope of the exception. On the contrary, it reaffirms that not all TDM’s stages need a specific exception to be valid: many

(mandating the adoption of open formats for publicly-funded research and data in order to enable TDM). The text of the amendments presented by the Member of the EP is available here: <https://euractiv.eu/wp-content/uploads/sites/2/2017/05/JURI-copyright-amendments.pdf>.

⁷⁹ Committee on Legal Affairs, Report on the proposal for a directive of the European Parliament and of the Council on copyright in the Digital Single Market (COM(2016)0593 - C8-0383/2016 - 2016/0280(COD)), 29 June 2018, available here: <http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&mode=XML&reference=A8-2018-0245&language=EN>

of them (such as the analysis, the creation of patterns and the subsequent publication) can be done freely and without infringing any IP entitlement.

The European Parliament voted against the JURI version during the plenary of July 2018, postponing any further decision to the September's meeting. However, in September, the EP approved the TDM amendments that essentially reproduce the July's version.⁸⁰ The Copyright reform is now in the trilogue's phase of discussions between the Commission, the Council and the EP.

It thus appears that the various texts on TDM available in September 2018 provide for the following exceptions:

	Commission text	Council text	EP text
TDM for research organisations	Mandatory and not overridable	Mandatory and not overridable	Mandatory and not overridable + Member States may continue to provide TDM exceptions in accordance with Art. 5(3)(a) of InfoSoc Directive
TDM for other beneficiaries (including in a commercial context)	/	Optional and overridable	Optional and overridable + Member States may continue to provide TDM exceptions in accordance with Art. 5(3)(a) of InfoSoc Directive

The Commission, the Council and the Parliament texts are unsatisfactory to promote the development of Big Data analytics in Europe. The conditions on the objectives of TDM (not-for-profit research or research with a public interest goal) and the beneficiaries (research organisations and cultural heritage institutions) lead to a narrow exception.⁸¹ First, because TDM

⁸⁰ Amendments adopted by the European Parliament on 12 September 2018 on the proposal for a directive of the European Parliament and of the Council on copyright in the Digital Single Market (COM(2016)0593 - C8-0383/2016 - 2016/0280(COD)), available here: http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P8-TA-2018-0337+0+DOC+XML+V0//EN#ref_1_1

⁸¹ In this sense, see the second sentence added at para. 1 of Art. 3, EP text.

(and research in general!) is not exclusive of the academic circles.⁸² TDM could be used by policy makers to test draft policies and regulation, by journalists or private individuals for fact checking, by consumers and lawyers to automatically compare the terms of service of different platforms, just to give a few examples. Such entities and individuals will not be able to acquire additional knowledge and/or provide additional services and tools relying on the TDM insights.

Second, in many instances (even in academic research which is partly supported by private funding), the boundary between commercial and non-commercial research is not always easy to trace. Such a limitation is likely to undermine also the spirit of the initial goal of the Commission, anticipated in the *Communication for a Digital Single Market Strategy*, whose intention was to promote research innovation for both non-commercial and commercial purposes.⁸³

Third, even if contractual limitations are pushed out of the door, private ordering can come back through the window, via TPMs. The proposal, in fact, does not grant any effective protection against TPMs as it is not clear there is a possibility to legally circumvent those that would unlawfully limit TDM.⁸⁴

Finally, even if the copyright reform will expand TDM boundaries, the main obstacle remains the ECJ *Ryanair* holding and the power of contracts in the absence of copyright or database protection. Of course, it does not make much sense to address this issue in an instrument, as the Draft Directive, dealing with intellectual property – as the problem is to secure some uses when there is no IP protection – but such issue risks never being addressed in another context either.

⁸² Many scholars have argued that the exception should be broadened. For instance, Margoni and Kretschmer 2018; Caspers and Guibault 2016b; Margoni and Dore 2016; Hilty and Richter 2017; Geiger et al. 2018.

⁸³ See above Communication, “A Digital Single Market Strategy for Europe”, COM/2015/0192 final, p. 7: “Innovation in research for both non-commercial and commercial purposes, based on the use of text and data mining (e.g. copying of text and datasets in search of significant correlations or occurrences) may be hampered because of an unclear legal framework and divergent approaches at national level. The need for greater legal certainty to enable researchers and educational institutions to make wider use of copyright-protected material, including across borders, so that they can benefit from the potential of these technologies and from cross-border collaboration will be assessed, as with all parts of the copyright proposals in the light of its impact on all interested parties”.

⁸⁴ Margoni and Kretschmer 2018; Caspers and Guibault 2016a. As acutely pointed out by Margoni and Kretschmer: “The EU legislature is fully aware of this contradiction but failed to address it properly. In fact, Art. 6 of the Proposal (“common provisions”) clarifies that the provisions of the first, third and fifth subparagraph of Art. 6(4) InfoSoc directive apply. In plain English, this means that if a user qualifies for an exception to copyright (e.g. TDM) but a Technological Protection Measure prevents them from doing it, Member States have an obligation to take appropriate measures to ensure that right holders make available to the beneficiary an exception or limitation. In the almost 20 years since when the InfoSoc directive was enacted, the UKIPO, which has correctly put in place a specific procedure for this type of situations, has received less than a handful of requests”.

5. Reconstructing the reproduction right: back to the roots

The TDM exception, if it is well designed at the end of the legislative process, will somewhat improve the situation but it is not likely to be the *panacea*. Some issues, like the one of SDSs that prompted our paper, will remain if the provision is overridable and not broad. It is probably too little considering the downside of the exercise: by searching for some legal certainty, guarded by a statutory exception, the EU legislator will confirm in black letter that TDM, even with the clarifications made by the EP in Recital 8a, is a copyright-relevant activity. The risk we are running is to give up on a crucial point, instead of embracing the fatigue of a serious discussion about the boundaries of the right of reproduction and copyright foundations.⁸⁵

That copyright might extend to the TDM process does not appear legitimate: no author of literary or artistic production in the past has conceived her copyright as a way to limit the use of her work as a source of useful information, for instance for discerning fluctuations in interest in a particular subject or for determining fashionable expressions. No author, while producing a work, has seriously relied on the possibility of earning revenues from the derivative use related to searching and indexing a corpus that includes her work.⁸⁶ Such use is far removed from the core exploitation field of most works. In addition, when it appeared more than two centuries ago, copyright was not only intended to remunerate authors (and publishers) for creating (and disseminating) works, but was also promoted to expand public learning. This is not only true for the US copyright system, as the same rationale is at the origin and core of the continental *droit d'auteur* system.⁸⁷

Even without an express exception, we can rely on the implicit requirement that the reproduction involves a use as a work. Such use as a work does not exist in the case of TDM, nor in other cases involving copying for deriving information or checking conduct (e.g., to identify plagiarism). As put by the Court in *Authors Guild v. Google, Inc.*, ‘the purpose of Google’s copying of the original copyrighted books is to make available significant information *about those books*, permitting a searcher to identify those that contain a word or term of interest’ (emphasis added).⁸⁸ This purposive analysis of the act of copying strongly weighs in favour of fair, because highly transformative, use. In the EU, the requirement of use as a work can help to reach the same outcome. Indeed, when acts of reproduction are carried out for the purpose of search and TDM, the work is not used as a work, it only serves as a tool or data for deriving other relevant information. The expressive features of the work are not used, and there is no public to enjoy

⁸⁵ These thoughts were firstly elaborated in Strowel 2018.

⁸⁶ See also Poort 2018.

⁸⁷ See the review of the history and principles of copyright in both legal traditions: Strowel 1993. More recently, Strowel 2014, pp. 701–703.

⁸⁸ *Authors Guild v. Google, Inc.* No. 13-4829-cv (2d Cir. Oct. 16, 2015). On April 18, 2016, the Supreme Court denied the petition for a writ of certiorari, leaving the Second Circuit ruling in Google's favour intact. To make available parts of the corpus of books, Google has scanned the digital copies and established a publicly available search function, the ngrams tool.

the work, as the work is only an input in a process for searching a corpus and identifying occurrences and possible trends or patterns.

Even if the EU Parliament and Council decide to include an exception for TDM that is broad enough (this would require some commercial research to be exempted), the analysis based on ‘the use of the work as a work’ condition for copyright infringement remains necessary to address other types of copying for the purpose of providing information, such as copies for checking mistakes and plagiarism, copies to use or repair a protected work with a utilitarian function, non-transitory copies made on proxy servers, smart disclosure systems and many other uses that cannot be anticipated.⁸⁹

As said, the main reason behind the introduction of the TDM exception was the reduction of legal uncertainties and the diverging national implementations regarding the research exceptions. If we look at the texts to be discussed during the trilogue, such uncertainties remain, and more divergences are likely because of the optional nature of the TDM exception provided in favour of beneficiaries other than research organisations.

6. Prohibition of TDM in T&C

As already mentioned, contracts also can affect TDM. In Sections 3 and 4, we have pointed out that, in some cases, copyright exceptions cannot be limited via contractual means. However, the issue is not harmonised under the InfoSoc Directive and even when the exceptions are protected against the power of contract, the free room they provide might be annihilated by TPMs. Furthermore, according to the *Ryanair* ruling, the copyright (or IP) balance cannot be exported so as to apply when there is no copyright (IP), but only contracts between parties.

To give an idea of the scope of the issue, we analyse in this Section the occurrences of contractual clauses prohibiting TDM or otherwise restricting copyright exceptions. As explained in Section 3, we looked at the copyright and database protection provisions, since TDM usually involves a “reproduction” and/or an “extraction” of the subject-matter to mine.⁹⁰

⁸⁹ The following US decisions quoted in *Authors Guild v. Google, Inc.* have, for instance, exempted several uses that, without the application of the work use requirement, cannot escape copyright’s exclusivity in the EU: *A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630, 638–640 (4th Cir. 2009) (justifying as transformative fair use purpose the complete digital copying of a manuscript to determine whether the original included matter plagiarized from other works); *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1165 (9th Cir. 2007) (justifying as transformative fair use purpose the use of a digital, thumbnail copy of the original to provide an Internet pathway to the original); *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 818–819 (9th Cir. 2003) (same); *Bond v. Blum*, 317 F.3d 385 (4th Cir. 2003) (justifying as fair use purpose the copying of author’s original unpublished autobiographical manuscript for the purpose of showing that he murdered his father and was an unfit custodian of his children).

⁹⁰ Caspers and Guibault 2016a; Stamatoudi 2016; Triaille et al. 2014.

The analysis focused on the terms and conditions (T&C) of twenty-one online platforms, equally distributed among three sectors: mobility (carpooling and car sharing), accommodation (including services for sharing office space), food (i.e., initiatives for the recuperation of unsold or unused food, sharing or delivery of home-cooked meals, etc.). In the selected sample, we tried to ensure a good balance between platforms operating globally and locally and between large capitalistic and cooperative initiatives.⁹¹ If different contractual versions were available, we have consulted the T&C for the Belgian market.

As shown in Table 1, 20 out of 21 platforms published the T&C on their website and 14 of them contained specific intellectual property clauses, directly or indirectly, related to TDM activities. More specifically:

- four platforms expressly prohibit TDM on the website content⁹²;
- Three others do not allow the use of any kind of bot, crawler or scraper (i.e., the automated software agents that search through the content of webpages. These are necessary tools for TDM, e.g. when there is no available application programming interface⁹³);
- in four cases, the reproduction or copy of website materials – which is usually a preliminary step of TDM process – is forbidden; and
- in three occasions, the formulation was vague or broad enough to exclude TDM.⁹⁴

This shows a trend toward a general contractual ban of TDM. The prohibition is broad and refers to all the website’s contents and services, thus including the informative pages containing the legal conditions. Such provisions were inserted with a view to protect the data (in some ways valuable) related to the service (e.g. the timetable of flights, the prices, the list of accommodation and contact details of the owners, etc.) from free exploitation: their effect on T&C and privacy policy would be an involuntary side effect. However, this is not the case: the contractual provision is usually confirmed and embedded into the technical instructions of the website.

To check this, we used the robots.txt file. It is an exclusion protocol that content providers can insert into the root directory to prevent crawling or indexing activities on certain

⁹¹ On the theoretical foundation of platform cooperativism, see Scholz 2016.

⁹² Blablacar does not prohibit TDM on the whole content of the website, but only on a substantial part of it.

⁹³ On crawling and scraping see Caspers and Guibault 2016a, p. 8.

⁹⁴ According to T&C of Bar d’Office, users cannot obtain (or attempt) to obtain any material or information through any means not intentionally made available by the platform. We shall conclude that third party applications aiming at analysing Bar d’office legal documents do not fall within the permitted uses. Wibee does not allow to “exploit in any way the content” but it has to be combined with the contractual provision that allows the use of the platform for non-commercial purposes only. Finally, Menu Next Door contained the broad, but vague, formulation “All rights reserved”.

pages of their website.⁹⁵ Adding the extension “robots.txt” at the end of the address of the website is possible to see the underlying instructions.⁹⁶ The latter can consist of two main commands:

- User-agent: it shows at what robots the instructions are directed. If there is an asterisk (“User-agent: *”), this means that the section applies to all robots.
- Disallow: it indicates what pages cannot be visited by the robots.

We found that three platforms are actually using the exclusion protocol to keep robots away from the whole server, one from a directory which contains amongst others the legal documents, and two specifically from the page of T&C and privacy policy (see the annexed Table 1, “Robots.txt” column). In all cases where the robots were disallowed, the T&C provisions prohibited TDM as well. Only in one case there was no TDM-related provision in the contract, but nevertheless the code instructions did not allow the indexing of the website content by a series of robots. However, it should be noticed that the robots excluded are those usually mentioned in the “black lists” of malicious programs.

The nature of robots.txt as a TPM is controversial. Looking at the definition of “technological measures” provided at Art. 6 of the InfoSoc Directive, the robots.txt could fall under it. Indeed, the protocol is a kind of technology that “is designed to prevent or restrict acts, in respect of works or other subject-matter, which are not authorised by the rightholder of any copyright or any right related to copyright”⁹⁷. However, the effectiveness of such technological measures is debatable. According to the Directive, it implies that “the use of a protected work or other subject-matter is controlled by the rightholders through application of an access control or protection process, such as encryption, scrambling or other transformation of the work or other subject-matter or a copy control mechanism, which achieves the protection objective”.⁹⁸ Some authors have argued that the protocol contains instructions that do not qualify as a technical barrier: any software agent can simply ignore the “Disallow” command without actively forcing any digital fence.⁹⁹ The content provider has just to rely on the voluntary compliance of the user, hoping that the visiting agent has been designed to follow the ASCII syntax. In any case, even if the robots.txt is not considered as an effective technological measure or not, its use by the content providers and online platforms confirms their willingness to limit TDM, as stated in their T&C.¹⁰⁰

⁹⁵ On the origin and functioning of this file, see <http://www.robotstxt.org/robotstxt.html>.

⁹⁶ Rotenberg and Compañó 2009.

⁹⁷ Art. 6(3), InfoSoc Directive.

⁹⁸ Ibidem.

⁹⁹ Sire 2015. *Contra*, Groom 2004, according to which a visiting agent programmed to systematically ignore the robots.txt can be seen as a strategy to circumvent a technological protection measure.

¹⁰⁰ In Europe, the robots.txt file has been questioned with reference to the issue of implied license only. See, for instance, the *Copiepresse v. Google* saga, commented in Strowel 2007, 2011. Doubts about the classification of robot.txt as a TPM in the US context are expressed by Jasiewicz 2012. In the US case *Healthcare Advocates, Inc. v. Harding* (497 F. Supp. 2d 627, 643 (E.D. Pa. 2007)), the Court for the Eastern District of Pennsylvania

How legitimate is such a general and absolute ban imposed in many T&C? The prohibition of TDM is likely to undermine the legitimate activity of consumers that could use smart disclosures mechanisms and instruments for the automatic analysis of contractual documents to better understand the terms of the user agreement. If a sort of “bionic eye” is available in order to scan a document and extract the relevant pre-contractual information, is it justified to prohibit its use?

As shown in the previous Sections, neither the current framework of copyright exceptions nor the new TDM exceptions under consideration by the Council and EP could apply to SDSs and be able to protect the rightful interests of the users. However, the principle of transparency, embedded in consumer and data protection legislations, could offer a last line of defence against an absolute prohibition of TDM.

7. Transparency 2.0: a right to machine legibility

Transparency is a cardinal principle in EU Law.¹⁰¹ In consumer and data protection legislations, information transparency is designed as a necessary feature of mandated disclosures, i.e. the obligation to provide one party (traditionally the weak one) with the information concerning the transaction. If the information is accessible, clear and understandable, mandated disclosures can effectively inform the consumer or the data subject about the essential content of the agreement and allow her to make an optimal decision or express a meaningful consent, where requested. In this sense, the substantive requirement (the duty to provide certain information) is complemented by formal requirements (the provision of the information in advance and the use of plain and intelligible language).

In consumer law, the principle of transparency has been traditionally interpreted as comprising two components:¹⁰² 1) the consumer has to be able to have knowledge of the terms before entering into contract;¹⁰³ 2) the information has to be provided in a way that the average

incidentally discussed the nature of the robots.txt file. The judge recognised the protocol as a TPM under the DMCA in that specific case. However, the Court expressly affirmed that robots.txt is not “analogous to digital password protection or encryption” and its nature must be assessed case-by-case (“This finding should not be interpreted as a finding that a robots.txt file universally qualifies as a technological measure that controls access to copyrighted works under the DMCA”).

¹⁰¹ Buijze 2013. The analysis presented in this paragraph has been further developed in Ducato forthcoming.

¹⁰² Loos 2015; Kästle-Lamparter 2018, pp. 429-430, 474 and 481. See also, Micklitz et al. 2009, pp. 135 ff.

¹⁰³ The transparency principle, as a duty to provide information before the conclusion of the contract, is envisaged in the Annex of the Unfair Terms Directive (UTD), which includes, among the list of potential unfair terms, the contractual provision which: “irrevocably binds the consumer to terms with which he had no real opportunity of becoming acquainted before the conclusion of the contract” (Annex, 1.i, UTD). It can also be derived by Recital 20, UTD. It is further recalled at Art. 6(1) of the Consumer Rights Directive (CRD).

consumer can understand without a legal advice. The latter means that information must be legible and given in a plain and intelligible language.¹⁰⁴

The principle of legibility, in particular, requires to take into consideration the font size, the layout and the accessibility of pre-contractual information. As pointed out by Micklitz et al., the possibility to actually read the text of the contract, i.e. the design of conditions “plainly both from an editing and optical point of view”¹⁰⁵ (no “small-print” for instance), must be seen as a corollary of intelligibility enshrined in the 93/13/EEC Unfair Terms in consumer contracts Directive (“UTD”).

The principle of legibility has been expressly codified in the 2011/83/EU Consumer Rights Directive (“CRD”) at Art. 7(1) for off-premises contracts and Art. 8 (1) for distance contracts, which both state that information shall be legible if provided on a durable medium. Interestingly, Kästle-Lamparter points out that: “Taken at a face value, the requirement of legibility excludes mere audio tapes or files: information must be provided in a human-readable format. But perhaps this was not intended and the provision should be rather read ‘legible if text-based’, or possibly ‘legible or audible’ ”.¹⁰⁶ Furthermore, it should be noted that if we assume the additional application of Directive 2000/31/EC (“e-commerce Directive”) when the contract is concluded with an online platform, Art. 10(3) establishes that: “Contract terms and general conditions provided to the recipient must be made available in a way that allows him to store and *reproduce* them” [emphasis added]. The latter could therefore constitute an additional argument to support the mining of T&C’s text.

The rationale of consumer protection, the ECJ’s interpretation of the principle of transparency but also the formal requirements set out in the e-commerce Directive are likely to

¹⁰⁴ The principle of transparency, *sub specie* of understandability of the information provided to the consumer, is specifically mentioned in several legislative instruments. The duty to provide the consumer with information in a clear and comprehensible manner is recalled at Art. 5, UTD: “In the case of contracts where all or certain terms offered to the consumer are in writing, these terms must always be drafted in plain, intelligible language”. Moreover, it is expressed at Arts 5(1), 6(1) CRD. and further expanded at Art. 8 CRD. In addition, when the contract is concluded “through a means of distance communication which allows limited space or time to display the information” (Art. 8.4, CRD), like the screen of a mobile phone, the trader will have to provide at least a set of pre-contractual information, such as the main characteristics of the goods or services, the identity of the trader, the total price, the right of withdrawal, the duration of the contract and, if the contract is of indeterminate duration, the conditions for terminating the contract. Among the appropriate means to display information to the consumer, the Commission suggested the adoption of a set of icons, making also available a model. However, such a measure does not seem to have taken hold (EC Commission, DG Justice Guidance document concerning Directive 2011/83/EU of the European Parliament and of the Council of 25 October 2011 on consumer rights, amending Council Directive 93/13/EEC and Directive 1999/44/EC of the European Parliament and of the Council and repealing Council Directive 85/577/EEC and Directive 97/7/EC of the European Parliament and of the Council, June 2014, available here: https://ec.europa.eu/info/sites/info/files/crd_guidance_en_0.pdf).

¹⁰⁵ Micklitz et al. 2009, p. 136.

¹⁰⁶ Kästle-Lamparter 2018, p. 474.

accommodate a broad understanding of legibility.¹⁰⁷ In particular, to ensure the balance of interests embedded in the 1993 UTD and 2011 CRD, it will be coherent to frame the legibility requirements, adopting a technologically neutral approach. This functional interpretation is further confirmed in the GDPR.

The analogy with the transparency in the data protection framework, despite the distinct area of application and the slightly different terminology used, is justified as there is a *koinè*, if not an open dialogue, between these two legal branches.¹⁰⁸

The GDPR has recently codified transparency as a pillar of data protection along the principles of lawfulness and fairness of the processing (Art. 5.1.a, GDPR). As specified by the European Data Protection Board (EDPB): “Transparency is an overarching obligation under the GDPR applying to three central areas: (1) the provision of information to data subjects related to fair processing; (2) how data controllers communicate with data subjects in relation to their rights under the GDPR; and (3) how data controllers facilitate the exercise by data subjects of their rights”¹⁰⁹. The principle of transparency echoes in many aspects the articulation shown in the consumer legislation.

The GDPR explicitly refers to legibility when it allows the use of standardised icons to complement the privacy notice. Icons shall give “in an easily visible, intelligible and clearly legible manner a meaningful overview of the intended processing”¹¹⁰. The GDPR further specifies that “where the icons are presented electronically they shall be machine-readable” (Art. 12.7, GDPR).

¹⁰⁷ When it has interpreted the plainness and intelligibility requirements, the ECJ has always excluded formalistic readings. In *Kásler*, for instance, the Court held that the requirement of transparency of terms, under the UTD, cannot “be reduced merely to being formally and grammatically intelligible” (ECJ, C-26/13, *Árpád Kásler and Hajnalka Káslerné Rábai v OTP Jelzálogbank Zrt.* [2014] ECLI:EU:C:2014:282, para. 71). Principle confirmed in the subsequent jurisprudence. See, ECJ, *Bogdan Matei and Ioana Ofelia Matei v SC Volksbank România SA* [2015] ECLI:EU:C:2015:127. See also, ECJ, C-191/15, *Verein für Konsumenteninformation v Amazon EU Sà* [2016] ECLI:EU:C: 2016:612). Terms must be transparent “so that the consumer can foresee, on the basis of clear, intelligible criteria, the economic consequences for him which derive from it” (ECJ, C-26/13, *Árpád Kásler and Hajnalka Káslerné Rábai v OTP Jelzálogbank Zrt.*, para. 73). As later specified in *Gutiérrez Naranjo*, under the transparency principle, the consumer has to be able to understand not only the economic consequences but also the legal ones (ECJ, C-154/15, *Francisco Gutiérrez Naranjo v Cajasur Banco SAU, Ana María Palacios Martínez v Banco Bilbao Vizcaya Argentaria SA (BBVA), Banco Popular Español SA v Emilio Irlés López and Teresa Torres Andreu* [2016] ECLI:EU:C:2016:980).

¹⁰⁸ About the interplay between consumer protection and data protection, Helberger et al. 2017.

¹⁰⁹ EDPB, *Guidelines on Transparency under Regulation 2016/679*, 13 April 2018.

¹¹⁰ Art. 12.7 GDPR. See also Recital 60. The link between transparency, information and visualisation is further stressed at Recital 58, GDPR: “The principle of transparency requires that any information addressed to the public or to the data subject be concise, easily accessible and easy to understand, and that clear and plain language and, additionally, where appropriate, visualisation be used. Such information could be provided in electronic form, for example, when addressed to the public, through a website. This is of particular relevance in situations where the proliferation of actors and the technological complexity of practice make it difficult for the data subject to know and understand whether, by whom and for what purpose personal data relating to him or her are being collected, such as in the case of online advertising”.

This is a highly behaviourally-informed legal innovation, considering that more often online users derive information from icons and pictures.¹¹¹

However, if the GDPR is systematically interpreted, the principle of transparency emerges in several other *loci*. First of all, data controllers, i.e. the person or entity which determines the purposes and means of the processing of personal data, shall take appropriate measures to provide the information required by law (at Arts. 13 and 14 GDPR) and any communication regarding the right of access, the use of automated individual decision-making, and personal data breach¹¹²:

- a) in a concise, transparent, intelligible and easily accessible form;
- b) using clear and plain language;
- c) provided in writing or by other means, including, where appropriate, by electronic means;
- d) provided orally, if requested so by the data subject.

Furthermore, where the processing is based on the data subject's consent, the request for it has to be presented in "a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language"¹¹³.

The *Guidelines on Transparency* (2018) drafted by the EDPB offer a first reading of such requirements. If according to "concision and transparency", information has to be presented "efficiently and succinctly, in order to avoid information fatigue"¹¹⁴, the "easy accessibility" imposes to make information clearly visible, e.g. on the website. This means that data controllers have to actively furnish the information (or the way to find it), while it is not an obligation of the data subject to start a quest for retrieving information. For example, data controllers should provide information by giving it directly to data subjects, "by linking them to it, by clearly signposting it or as an answer to a natural language question"¹¹⁵. In addition, the duty to provide information in "writing or by other means" specifically recalls the CRD obligation to make information available to the consumer "in a way appropriate to the means of distance

¹¹¹ OECD 2016, p. 17. A version for the GDPR set of icons was released in the Annex of the first reading of the European Parliament (European Parliament legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) ([COM\(2012\)0011- C7-0025/2012 -2012/0011\(COD\)](#)) (Ordinary legislative procedure: first reading). But no other official initiatives have been taken in this direction. An interesting methodology to answer the challenges of GDPR's icons has been developed within the research project run by the Cirsfid group at the University of Bologna: <http://gdprbydesign.cirsfid.unibo.it/>

¹¹² Art. 12.1, GDPR. See also Recitals 39 and 42, GDPR.

¹¹³ Art. 7.2, GDPR.

¹¹⁴ EDPB, *Guidelines on Transparency under Regulation 2016/679*, p. 8.

¹¹⁵ *Ibid.*, p. 11.

communication used” (Art. 8.1, CRD). The EDPB emphasises that such a provision has to be interpreted broadly, allowing to choose the most appropriate means and format to reach the informative goal: for a privacy policy on websites, “digital layered privacy, but also ‘just in time’ contextual pop-up notices, 3D touch or hover-over notices, and privacy dashboards”¹¹⁶ can be used.

The Board further underlines the importance to consider the specific circumstances where the provision/communication occurs or how the interactions between the controller and the data subject happen. For instance, the EDPB warns that the electronic privacy notice on a website is likely to be ineffective for screenless IoT or smart devices: it could be preferable, for example, to include the privacy policy in the instruction manual or make it easier to access through a QR code printed on the device.¹¹⁷

To sum up, through the transparency principle, *sub specie* of obligation to provide concise and transparent, easy accessible information, in writing or other means, the GDPR recalls the conceptual background already seen in consumer law. Furthermore, the EDPB encourages the use of technological advancements to inform users in a more meaningful way.

If we look at these provisions systematically and at their interpretation by EDPB and the ECJ, they can operate as functional equivalents: both of them pursue the same rationale, i.e., protecting the weak party from information asymmetries, by requiring that information must be visible, accessible and readable.

Therefore, if this is the policy goal, the principle of transparency enshrined in consumer and data protection laws is technologically neutral and can accommodate a right to “machine legibility”. By machine legibility we mean the possibility for a SDS to have access to the pre-contractual information (T&C) and the information related to the processing (privacy policy) in a format processable by the smart system. If thanks to AI there are now instruments enhancing the human ability to read and understand contractual terms and privacy settings, not only information should be legible to a human eye, but also to the tool that a user can take advantage of. Therefore, either depending on IPRs, contractual or TPM limitations, an absolute prohibition of TDM on pre-contractual and privacy information available online on the website of the platform will unreasonably restrict a legitimate prerogative of the consumer or the data subject.

Furthermore, the solution will also be costless to the data controller/trader: the platform could allow TDM, by enabling the indexing of the corresponding page on their website through the robots protocol and lowering TPMs barriers, if any.

¹¹⁶ Ibid., p. 12.

¹¹⁷ Ibidem.

8. Conclusion

SDSs offer clear benefits for the consumers as they reduce information asymmetries and improve the access to data needed to take informed decisions. The possibility to deploy SDSs might however be prevented as there are technical and legal obstacles, including the prohibitions inserted in many T&C and other online documents of interest for the consumers. In this context, it is not clear whether TDM tools needed for the use of SDSs can be used. The existing copyright and database exceptions do not adequately tackle the TDM issue, and, as recognized by the European Commission, there is a need for introducing a TDM exception. However, neither the initial proposal by the European Commission focusing on the research context, nor the amendments discussed within the Council and the European Parliament appear sufficient to facilitate the use of TDM for improved smart disclosure and, more broadly, for AI applications.

Regarding SDSs, the transparency principle embedded in consumer and data protection rules offers some legal ground to justify the use of TDM for consumer empowerment. But the requirement of machine legibility that appears necessary in a society where the automatic treatment of information becomes central could be further promoted by a well-designed TDM exception. We are not yet there.

References

- Ayres I, Schwartz A (2014) The No-Reading Problem in Consumer Contract Law. *Stanford Law Review*, 66(3): 545-610
- Bakos Y, et al. (2014) Does Anyone Read the Fine Print? Consumer Attention to Standard Form Contracts. *The Journal of Legal Studies* 43(1): 1-35
- Bar-Gill O (2015) Defending (Smart) Disclosure: A Comment on More Than You Wanted to Know. *Jerusalem Review of Legal Studies* 11(1): 75-82
- Ben-Shahar O (2009) The Myth of the 'Opportunity to Read' in Contract Law. *European Review of Contract Law* 5(1): 1-28
- Ben-Shahar O (2013) Regulation Through Boilerplate: an Apologia. *Mich. L. Rev.* 112: 883-903
- Bently L, Sherman B (2009) *Intellectual property law*. Oxford University Press, Oxford
- Bernault C, et al. (2017) *Traité de la propriété littéraire et artistique*. Litec, Paris
- Bernhardt B, et al. (2015) Revolutionizing Scholarship: a Panel Discussion on Text and Data Mining. *Serials Review*, 41(3): 184-186
- Beunen AC (2007) Protection for databases: The European database directive and its effects in the Netherlands, France and the United Kingdom. Wolf Legal Publishers, Nijmegen
- Borghi M, Karapapa S (2015) Contractual restrictions on lawful use of information: sole-source databases protected by the back door?. *European Intellectual Property Review* 37(8): 505-514
- Braun D, et al. (2018) Customer-centered LegalTech: Automated Analysis of Standard Form. *Internationales Rechtsinformatik Symposium (IRIS)*: 627-634
- Buijze A (2013) The six faces of transparency. *Utrecht L. Rev.* 9(3): 3-25
- Busch C (2016) The future of pre-contractual information duties: from behavioural insights to big data. In: Twigg-Flesner C (ed), *Research Handbook on EU Consumer and Contract Law*. Edward Elgar, Cheltenham, UK, Northampton, MA, USA, 221-240
- Busch C (forthcoming) Implementing Personalized Law: Personalized Disclosures in Consumer Law and Privacy Law. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3181913. Accessed 18 August 2018
- Caso R (2004) Digital rights management: il commercio delle informazioni digitali tra contratto e diritto d'autore. CEDAM, Padua

- Caspers M, Guibault L (2016a) Baseline report of policies and barriers of TDM in Europe. https://www.futuretdm.eu/wp-content/uploads/FutureTDM_D3.3-Baseline-Report-of-Policies-and-Barriers-of-TDM-in-Europe-1.pdf. Accessed 18 August 2018
- Caspers M, Guibault L (2016b). A right to 'read' for machines: Assessing a black-box analysis exception for data mining. *Computer Sciences* 53(1): 1-5
- Contissa G, et al. (2018) CLAUDETTE meets GDPR. Automating the Evaluation of Privacy Policies using Artificial Intelligence. https://www.beuc.eu/publications/beuc-x-2018-066_claudette_meets_gdpr_report.pdf. Accessed 18 October 2018
- de Visscher F, Michaux B (2000). *Précis du droit d'auteur et des droits voisins*. Bruylant, Brussels
- Derclaye E (2008) *The Legal Protection of Databases: A Comparative Analysis*. Edward Elgar, Cheltenham, UK, Northampton, MA, USA
- Derclaye E (2012). Football Dataco: skill and labour is dead. <http://copyrightblog.kluweriplaw.com/2012/03/01/football-dataco-skill-and-labour-is-dead/>. Accessed 18 August 2018
- Derclaye E (2014) The Database Directive. In: Stamatoudi I, Torremans P (eds.), *EU copyright law: A commentary*. Edward Elgar, Cheltenham, UK, Northampton, MA, USA, 298-354
- Derclaye E, Favale M (2010) Paper 3: User Contracts. In: Kretschmer M, et al. (eds.) *The relationship between copyright and contract law*. http://eprints.bournemouth.ac.uk/16091/1/_contractlaw-report.pdf. Accessed 18 August 2018
- Ducato R (forthcoming) Transparency by (Legal) Design. Paper presented at the Private Law Consortium, Harvard Law School, 15 May 2018 and at the Younger Scholar Informal Symposium, organised within the General Congress of the International Academy of Comparative Law, Fukuoka, 25 July 2018
- Dussolier S (2007) Sharing Access to Intellectual Property through Private Ordering. *Chicago-Kent Law Review* 82(3):1391-1438
- Feldman R, Sanger J (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press, Cambridge
- Geiger C, et al. (2018) The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Legal Aspects. Centre for International Intellectual Property Studies (CEIPI) Research Paper No. 2018-02: 1-34
- Groom J (2004) Are Agent Exclusion Clauses a Legitimate Application of the EU Database Directive. *SCRIPTed* 1: 83-118

- Guibault L (2002) Copyright limitations and contracts. Kluwer Law International, The Hague
- Guibault L, et al. (2012) Study on the implementation and effect in Member States' laws of Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society. Report to the European Commission, DG Internal Market, February 2007. Amsterdam Law School Research Paper No. 2012-28. Institute for Information Law Research Paper No. 2012-23. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2006358. Accessed 18 August 2018
- Hall W, Pesenti J (2017) Growing the Artificial Intelligence Industry in the UK. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf. Accessed 18 August 2018
- Harkous H, et al. (2018) Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning. arXiv preprint arXiv:1802.02561. Accessed 18 August 2018
- Hearst MA (2003) Text Data Mining. In: Mitkov R (ed.), *The Oxford Handbook of Computational Linguistics*. Oxford University Press, Oxford, New York, 616-62
- Helberger N (2013) Forms matter: Informing consumers effectively (Study commissioned by BEUC). https://www.ivir.nl/publicaties/download/Form_matters.pdf. Accessed 18 August 2018
- Helberger N, et al. (2017) The perfect match? a closer look at the relationship between eu consumer law and data protection law. *Common Market Law Review* 54(5): 1427-1465
- Helberger N, Hugenholtz PB (2007) No place like home for making a copy: Private copying in European copyright law and consumer law. *Berkeley Tech. LJ*, 22(3): 1061-1093
- Helberger N, et al. (2013) Digital content contracts for consumers. *Journal of consumer policy* 36(1): 37-57
- Helleringer G, Sibony A-L (2017) European Consumer Protection through the Behavioral Lens. *Columbia Journal of European Law* 23(3): 607-649
- Hillman RA (2006) Online Boilerplate: Would Mandatory Website Disclosure of E-Standard Terms Backfire? *Michigan Law Review* 104(5): 837-856
- Hilty R et al. (2008) Declaration on a balanced interpretation of the “three-step test” in copyright law. *International Review of Intellectual Property and Competition Law* 39(6): 707-712
- Hilty R, Richter H (2017) Position Statement of the Max Planck Institute for Innovation and Competition on the Proposed Modernisation of European Copyright Rules Part B

- Exceptions and Limitations (Art. 3-Text and Data Mining). https://www.ip.mpg.de/fileadmin/ipmpg/content/stellungnahmen/MPI_Position_Statement_Part_B_Chapter_1_Update23022017.pdf. Accessed 18 August 2018
- Hugenholtz PB (2000) Copyright, contract and code: what will remain of the public domain. *Brook. J. Int'l L.* 26: 77-90
- Jasiewicz MI (2012) Copyright Protection in an Opt-Out World: Implied License Doctrine and News Aggregators. *The Yale Law Journal* 122(3): 837-850
- Kästle-Lamparter D (2018) Pre-contractual information duties. In: Jansen N, Zimmermann R (eds.), *Commentaries on European Contract Laws*. Oxford University Press, Oxford, 384-504
- Kretschmer M, et al. (2010) The relationship between copyright and contract law. http://eprints.bournemouth.ac.uk/16091/1/_contractlaw-report.pdf. Accessed 18 August 2018
- Lippi M, et al. (2018) CLAUDETTE: an Automated Detector of Potentially Unfair Clauses in Online Terms of Service. arXiv preprint arXiv:1805.01217. Accessed 18 October 2018
- Loos M (2015) Transparency of standard terms under the Unfair Contract Terms Directive and the proposal for a common European sales law. *European Review of Private Law* 23(2): 179-193
- Margoni T (2012) Eccezioni e limitazioni al diritto d'autore in Internet= Exceptions and Limitations to Copyright Law in the Internet. https://www.ivir.nl/publicaties/download/Giurisprudenza_Italiana_2011_8_9.pdf. Accessed 18 August 2018
- Margoni T, Dore G (2016) Why We Need a Text and Data Mining Exception (But it is Not Enough). <https://interop2016.github.io/pdf/INTEROP-13.pdf>. Accessed 18 August 2018
- Margoni T, Kretschmer M (2018) The Text and Data Mining exception in the Proposal for a Directive on Copyright in the Digital Single Market: Why it is not what EU copyright law needs. <https://www.create.ac.uk/blog/2018/04/25/why-tdm-exception-copyright-directive-digital-single-market-not-what-eu-copyright-needs/>. Accessed 18 August 2018
- Mazziotti G (2008) *EU Digital Copyright Law and the End-User*. Springer, Verlag, Berlin, Heidelberg
- Micklitz H-W, et al. (2009) *Understanding EU consumer law*. Intersentia, Antwerp

- Montagnani ML (2007) Dal peer-to-peer ai sistemi di Digital Rights Management: primi appunti sul melting pot della distribuzione online. *Il diritto d'autore* (1):1-57
- Montagnani ML, Aime G (2017) Il text and data mining e il diritto d'autore. *AIDA*: 376-394
- Myska M, Harasta J (2016) Less is More: Protecting Databases in the EU after Ryanair. *Masaryk UJL & Tech.* 10: 170-198
- OECD (2016) Protecting Consumers In Peer Platform Markets: Exploring The Issues. https://unctad.org/meetings/en/Contribution/dtl-eWeek2017c05-oecd_en.pdf. Accessed 18 August 2018
- Poort J (2018) Borderlines of Copyright Protection: An Economic Analysis. In: Hugenholtz PB (ed.), *Copyright Reconstructed: Rethinking Copyright's Economic Rights in a Time of Highly Dynamic Technological and Economic Change*. Wolters Kluwer, Alphen aan den Rijn, 283-338
- Porat A, Strahilevitz LJ (2013) Personalizing Default Rules and Disclosure with Big Data. *Mich. L. Rev.* 112(8): 1417-1478
- Radin MJ (2013). *Boilerplate: the Fine Print, Vanishing Rights, and the Rule of Law*. Princeton University Press, Princeton, Oxford
- Rosati E (2018) National and EU text and data mining exceptions: room for coexistence? <http://ipkitten.blogspot.com/2018/03/national-and-eu-text-and-data-mining.html>. Accessed 18 August 2018
- Rosati E (2013a). *Originality in EU Copyright*. Edward Elgar, Cheltenham, UK, Northampton, MA, USA
- Rosati E (2013b) Towards an EU-wide Copyright?(Judicial) pride and (legislative) prejudice. *Intellectual Property Quarterly* 1:47-68
- Rotenberg B, Compañó R (2009) Search Engines for audio-visual content: Copyright law and its policy relevance *Telecommunication Markets*. In: Curwen P, et al. (eds.) *Telecommunication Markets*. Springer, Heidelberg and New York, 113-139
- Sadeh N, et al. (2013) The usable privacy policy project. <http://reports-archive.adm.cs.cmu.edu/anon/isr2013/CMU-ISR-13-119.pdf>. Accessed 18 august 2018
- Scholz T (2016) *Platform Cooperativism. Challenging the Corporate Sharing Economy*. Rosa Luxemburg Stiftung, New York
- Sire G (2015) Inclusion exclue : le code est un contrat léonin. *Enquête sur la valeur technique et juridique du protocole robots.txt*. *Réseaux* 189(1): 187-214

- Stamatoudi IA (2016) Text and data mining. In: Stamatoudi IA (ed.), *New Developments in EU and International Copyright Law*. Wolters Kluwer, Alphen aan den Rijn, 251-282
- Strowel A (1993) *Droit d'auteur et copyright, Divergences et convergences*. Bruylant, Paris, Bruxelles
- Strowel A (2007) Google et les nouveaux services en ligne: quels effets sur l'économie des contenus, quels défis pour la propriété intellectuelle. *Journal des tribunaux* 22: 589-598
- Strowel A (2011). *Quand Google défie le droit*. Larcier, Bruxelles
- Strowel A (2012) European Copyright: Beyond the Additions Made by the European Court of Justice, Some Pieces Are Still Missing. In: Janssens M-C, Overwalle GV (eds.), *Harmonisation of European IP Law, In Honor of Fr. Gotzen*. De Boeck, Bruxelles, 73-98
- Strowel A (2014) *Droit d'auteur et copyright. Convergences des droits, régulation différente des contrats* Mélanges en l'honneur du Professeur André Lucas. LexisNexis, Paris, 699-717
- Strowel A (2015) *Fair Compensation for Private Copying Copyright and the Digital Agenda for Europe: Current Regulations and Challenges for the Future*. Sakkoulas Publications, Athens, 189-195
- Strowel A (2018) *Reconstructing the Reproduction and Communication to the Public Rights: How to Align Copyright with Its Fundamentals*. In: Hugenholtz PB (ed.), *Copyright Reconstructed Rethinking Copyright's Economic Rights in a Time of Highly Dynamic Technological and Economic Change*. Wolters Kluwer, Alphen aan den Rijn, 203-240
- Strowel A, Derclaye E (2001) *Droit d'auteur et numérique: logiciels, bases de données, multimédia*. Bruylant, Bruxelles
- Sunstein C (2012) *Informing Consumers through Smart Disclosure*. <https://obamawhitehouse.archives.gov/blog/2012/03/30/informing-consumers-through-smart-disclosure>. Accessed 28 August 2018
- Triaille JP, et al. (2013) *Study on the application of Directive 2001/29/EC on copyright and related rights in the information society*. <https://publications.europa.eu/en/publication-detail/-/publication/9ebb5084-ea89-4b3e-bda2-33816f11425b>. Accessed 18 October 2018
- Triaille JP, et al. (2014) *Study on the legal framework of text and data mining (TDM)*. <https://publications.europa.eu/en/publication-detail/-/publication/074ddf78-01e9-4a1d-9895-65290705e2a5/language-en>. Accessed 18 October 2018

- Valenti R (2007a) Art. 68-bis. In: Ubertazzi LC (ed.), *Diritto d'autore*. Estratto da L. C. Ubertazzi "Commentario breve alle leggi su proprietà intellettuale e concorrenza". CEDAM, Padua, 203-204
- Valenti R (2007b) Introduzione agli artt. 65-71-quinquies. In: Ubertazzi LC (ed.), *Diritto d'autore*. Estratto da L.C. Ubertazzi "Commentario breve alle leggi su proprietà intellettuale e concorrenza". CEDAM, Padua, 190-196
- Vallés RC (2009) *Research Handbook on the Future of EU Copyright The Requirement of Originality*. Edward Elgar, Cheltenham, UK, Northampton, MA, USA
- Vila T, et al. (2003) Why we can't be bothered to read privacy policies models of privacy economics as a lemons market. ICEC '03 Proceedings of the 5th international conference on Electronic commerce, 403-407
- Waelde C, et al. (2013) *Contemporary intellectual property: law and policy*. Oxford University Press, Oxford
- Walter M, Von Lewinski S (2010) *European Copyright Law. A Commentary*. Oxford University Press, Oxford
- Wilhelmsson T (2004) The Abuse of the "Confident Consumer" as a Justification for EC Consumer Law. *Journal of Consumer Policy* 27(3): 317-337

Table 1. Overview of limitations and prohibitions in Sharing Economy Platforms (IoP project)¹

PLATFORM		T&C	IP-related CLAUSE		Robots.txt	
	Name	Sector	Date	What users can do	What users cannot do	Does the platform use an exclusion protocol to keep robots away from T&C and privacy policy webpages?
1	Uber www.uber.com	Mobility	04/12/2017	Access and use the content only for personal and non-commercial use	- reproduce, modify, prepare derivative works, etc. - scraping, data mining	NO
2	Blablacar www.fr.blablacar.be	Mobility	24/04/2017	Access and use the content only for personal and non-commercial use	- reproduce, modify, prepare derivative works, etc. - data mine a substantial part of the platform	YES
3	CarAmigo https://www.caramigo.eu/be/en	Mobility	N/A	- print or download fragments of the materials appearing on the Website, for non-commercial purposes, informative and personal - copy the material appearing on the website with the intention to disseminate and communicate them within the family circle	- copy, send, distribute, sell, publish, transmit, circulate, arrange or modify any Website material (unless otherwise provided by law)	NO
4	Cambio ² www.cambio.be	Mobility	N/A	N/A	No provisions TDM-related	The following robots are excluded from the entire server: BLEXBot, MauiBot, DomainCrawler, AhrefsBot,

¹ This analysis has been primarily conducted within the framework of the project “The Internet of Platforms: an empirical study on private ordering and consumer protection in the sharing economy” (<http://www.rosels.eu/research/research-project-iop/>). All links have been accessed on 10/08/2018.

² Terms available in French and Dutch only. There are three different contracts: one for Brussels, one for Wallonie, and one for Flanders. The version analysed here is the contract available for users in Brussels.

						OmniExplorer_Bot/1.09, SemrushBot, SemrushBot-SA, Mechanics, jobs.de-Robot, test_robot@gmx-topmail.de, XoviBot, SEOkicks-Robot, MJ12bot
5	Zipcar www.zipcar.be	Mobility	31/05/2017	N/A	No provisions TDM-related	NO
6	Wibee ³ www.wibee.be	Mobility	01/08/2018	-use the platform contents for non-commercial purposes	- not exploit in any way the content	NO
7	Zencar www.zencar.eu	Mobility	02/01/2017	N/A	No provisions TDM-related	NO
8	Airbnb www.airbnb.com	Accommodation	19/06/2017	Access and view the content only for personal and non-commercial use	- copy, adapt, modify, prepare derivative works, etc. - use any robots, spider, crawler, scraper or other automated means or processes to access, collect data or other content from or otherwise interact with the Airbnb Platform for any purpose	YES. The “Terms” page is specifically disallowed

³ Terms available in French and Dutch only.

Limitations to text and data mining and consumer empowerment

9	Couchsurfing www.couchsurfing.com	Accommodation	19/07/2017	Access and use the contents and the services only for their intended purposes	<ul style="list-style-type: none"> - reproduce, modify, prepare derivative works, etc. (unless otherwise provided by law) - data mining robots or similar data gathering or extraction methods - download (other than the page caching) of any portion of the Service 	The following robots is excluded from the entire server: ADMantX
10	Homeaway www.abritel.fr	Accommodation	14/11/2017	<ul style="list-style-type: none"> - download, display or print individual pages of the Site to evidence any agreement with HomeAway and to retain a copy of their bookings. - The relevant file or the relevant printout must clearly bear the text “© 2017 HomeAway - All Rights Reserved” 	<ul style="list-style-type: none"> - reproduce, copy in whole or in part - use robots, spiders or similar data gathering and extraction tools 	YES. It disallows a directory that contains also Terms and Privacy Policy
11	Sharedesk www.sharedesk.net	Accommodation/Office space	22/02/2016	<ul style="list-style-type: none"> - store temporary copies of the website materials in RAM (incidental to the surfing) - store files automatically cached by the Web browser for display enhancement purposes - print or download one copy of a reasonable number of pages of the Site for your use in connection with the Services. - download a single copy of the application to the personal computer or mobile device solely for personal, non-commercial use 	<ul style="list-style-type: none"> - reproduce, distribute, modify, create derivative works of, publicly display, publicly perform, republish, download, store or transmit any of the material on the Website, with the exception of the permitted uses - use manual or automated software, devices, scripts robots, other means or processes to access, “scrape,” “crawl” or “spider” any web pages or other services contained in the Site, Application, Services or Content 	NO

12	Warmshowers www.warshomwes.org	Accommodation	N/A	N/A	No provisions TDM-related	NO
13	Bar d'Office ⁴	Accommodation/Office space	N/A	N/A	- obtain or attempt to obtain any materials or information through any means not intentionally made available by the platform	NO
14	Flipkey www.rentals.tripadvisor.com	Accommodation	16/10/2017	- download and print material from the Site solely for the purposes of keeping a record of the Booking and the Property booked.	- not copy, transmit, modify, republish, save, pass off, or link to any content or material on the Site without the platform prior written consent	YES. The "Termsandconditions" and "privacy" pages are specifically disallowed
15	GASAP www.gasap.be	Food	N/A	N/A	N/A	NO
16	Topino.be ⁵ www.topino.be	Food	N/A	N/A	-reproduce, decompile, reverse engineers, etc.	NO
17	Meal Sharing www.mealsharing.com	Food	07/11/2012	- Access, view, and print Meal Sharing and User Content for non-commercial, at-home use	- reproduce, duplicate, copy, reverse engineer, sell, resell or exploit any portion of the Service, use of the Service, or access to the Service without the express written permission by the platform	NO

⁴ Terms and conditions are named "Disclaimer" and it's a miscellaneous: <http://bardoffice.com/disclaimer/>

⁵ The IP provisions are listed in their Privacy policy [sic!]

Limitations to text and data mining and consumer empowerment

18	Vlaams www.thuisafgehaald.be	Food	05/11/2012	N/A	- copy, multiply, spread or provide the design and contents, including trademarks, logos, images, photos and texts on the website to others in any way without written permission	NO
19	Recup'Kitchen http://www.recupkitchen.be	Food	N/A	No terms available	No terms available	/
20	Menu next door ⁶ www.menunextdoor.com	Food	N/A	N/A	All rights reserved	N/A
21	Flavr ⁷ www.flavr.be	Food	03/03/2017	- use the platform for personal and non-commercial purposes	- Not copy, modify, alter, adapt, make available, translate, port, reverse engineer, decompile, or disassemble any portion of the Platform in any way - manually or automatically with the use of any robot, spider, crawler, any search or retrieval application, or use other automatic device, process or method to access the Platform and retrieve, index and/or data-mine any information	N/A

⁶ The platform closed during the performance of the study

⁷ The platform closed during the performance of the study.