Copyright

Caterina Sganga

As other intellectual property entitlements, copyright provides a bundle of exclusive rights which are limited in time and in scope, and are conceived to incentivize the production of new creations by granting to rightholders the possibility to gain, in a protected monopolistic setting, an appropriate remuneration from the exploitation of their works (Landes-Posner 1987, 344). Modern copyright statutes share several substantial traits, thanks to the international harmonization of the discipline, which originated from the Berne Convention in 1886, amended several times until 1971, and found its detailed completion in the WTO Agreement on Trade Related Aspects of Intellectual Property Rights (TRIPs) and the WIPO Internet Treaties (WIPO Copyright Treaty (WCT) and WIPO Performers and Phonograms Treaty (WPPT)) in 1996. While the two main models - the Anglo-Saxon copyright and the continental *droit d'auteur* – still diverge in some regulatory choices, such as the protection of moral rights and the regime of exceptions, most of the key features of the discipline, from exclusive rights to term of protection, notion of authorship, and definition of protected works, are now converging to a large extent (already before the last steps in the international standardization, see Strowel 1993, *passim*).

Copyright matters for data science on several accounts. Three are, however, the most pressing questions that are currently on the table of legislators and courts. First, it is not always clear whether raw data, their organization in databases and the algorithms used to structure and analyze data by artificial intelligence (AI) agents may or may not fall within the scope of copyright-protected works. Second, the enforcement of exclusive rights may hamper fundamental research activities on big data, such as text and data mining (TDM), which require the adaptation of existing exceptions to copyright in order to be lawfully carried out without the authorization of rightholders. Third, AI agents are now able to generate protected works, raising the question of whether non-human creators may be considered authors for the purposes of copyright protection, and what the consequences of this may be in terms of attribution of copyright ownership. This chapter will focus on each of these interpretative problems, offering a brief comparative overview of the answers provided to date, and will conclude by providing snapshots of the current policy debate on the advisability to introduce a new exclusive right on raw data to data producers, and in which form.

1. The copyrightability of data

Every copyright system excludes, implicitly or explicitly, certain subject-matters from its scope of protection, usually in light of their role of building blocks for the production of creative works. Ideas, facts, colors, perspectives, basic words, themes or plots are either banned from the list of protectable works, or crossed out from protection for they do not meet the threshold of originality required by law (Gervais 2017, 77).

The first international treaty in the field of copyright – the Berne Convention – rules that "the protection of this convention shall not apply to the news of the day or to miscellaneous facts having the character of mere items of press information" (Article 2(8) BC), thus clarifying that copyright does not extend to facts. No reference is made to ideas, although the requirements of originality and

authorship set by Article 2 BC suggest their exclusion from the objective scope of copyright, as also made clear by the WIPO Guide to the Berne Convention (id., 12). More recent texts, such as the TRIPs Agreement and the WCT, state that copyright protection extends only to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such (Article 9(2) TRIPs; Article 2 WCT), thus introducing for the first time in international sources the so-called idea-expression dichotomy to define the scope of copyright (Gervais 2017, 78). A similar language recurs in the EC Software Directive (2009/24/EU, Article 1(2)), which affirms that the protection offered "shall apply to the expression in any form of a computer program. Ideas and principles which underlie any element of a computer program, including those which underlie its interfaces, are not protected by copyright under this Directive".

The idea-expression dichotomy and its implications are enshrined also in the US Copyright Act (§102(b)), which embeds a principle made famous by the quote of Justice Brandeis in his dissent in *International News Service v. Associated Press* (248 U.S. 215 (1918)), who emphasized how "[t]he general rule of law is, that the noblest of human productions - knowledge, truths ascertained, conceptions and ideas - after voluntary communication to others, are free as the air to common use". Along the same lines, it is common for national courts to deny protection to otherwise copyrightable works where the underlying idea can be expressed only in very limited ways, or it is forced by technical rules or the technical result a functional work has to achieve (also known in the US as, respectively, merger doctrine and *scenes à faire* doctrine. See Goldstein 2005, §2.3.2; Walker-Von Lewinski 2010, 5.4.27). Data are not mentioned by any of such provisions. However, scholars and courts generally agree on excluding them from the subject matter of copyright, either on the basis of the idea-expression dichotomy or on ground of lack of originality and authorship (van Erp 2017, 237). A confirmation of the validity of such a reading comes from Articles 10(2) TRIPS and 5 WCT, which emphasize that copyright on compilations of data "shall not extend to the data or material itself" – an approach also confirmed by the EC Database Directive.

2. The copyrightability of AI software

As any other computer program, the software used as an artificial intelligence agent to structure data and operate on them may be protected by copyright and, to the extent it can be considered a computer-implemented invention, by patent. Several countries started introducing provisions in their copyright act to protect computer programs as literary works from the early 1990s, until the TRIPs Agreement confirmed the validity of the approach in Article 10(1), ruling that "computer programs, whether in source or object code, shall be protected as literary works under the Berne Convention", and Article 4 WCT similarly stated that "computer programs are protected as literary works within the meaning of Article 2 of the Berne Convention. Such protection applies to computer programs, whatever may be the mode or form of their expression". Some national statutes provided an explicit definition of software, like the US Copyright Act, which conceptualizes it as "a set of statements or instructions to be used directly or indirectly in a computer in order to bring about a certain result" (§101). Others cover with the same protection also preparatory materials, as the EU Software Directive (Article 1), or computer-generated works, as the UK Copyright Designs and Patents Act (CDPA, §178).

What distinguishes the most all the national solutions implemented in the past decades are the scope of the protectable subject-matter and the standards or requirements of protection.

In the EU, for example, Article 1(2) of the Software Directive specifies that "ideas and principles which underlie any element of a computer program, including those which underlie its interfaces, are not protected by copyright". This poses the question of whether algorithms, which are general rules and instructions that are used to solve specific informational problems, and represent the description of a process on a high level of abstraction, could then be copyright-protected. There is general convergence around the fact that the answer depends on whether the algorithm itself represents an abstract idea for resolving an informational problem – and is that case it does not enjoy protection – or it offers a specific and structured solution to a specific problem, which would make it possible to consider it protectable via copyright, even if extremely abstract (see Bath 2002, 138; Derclaye 2000, 7, 13). This reading is also supported by the legislative history of the Directive. The initial proposal, in fact, fully excluded logic, algorithms and programming languages from its subject-matter, while the final version limited the list to ideas and principles (Walker-von Lewinski 2010, 5.1.11). By the same token, ideas and principles underlying user interfaces are not protected by copyright (as in case C-393/09 *Bezpečnostní softwarová asociace - Svaz softwarové ochrany v Ministerstvo kultury*, EU:C:2010:816).

Article 1(3) sets as requirement of protection the fact that the computer program is "original in the sense that it is the author's own intellectual creation", preempting Member States from introducing any other criteria, while Recital 8 specifies that "no tests as to the qualitative or aesthetic merits of the program should be applied". According to the Commission's Report on the implementation of the Software Directive (Implementation Report, 6), the harmonization of the level of originality has required twelve Member States to lower the threshold of protection, and the remaining three to raise it. The standard is now set at a moderate level of creative input, and excludes programs that are commonplace and banal in the industry (Walker-von Lewinski 2010, 5.1.16; Derclaye 2000, 15; see also Explanatory Memorandum, Part 1 n.2.5, 2.6, 3.6).

Systems belonging to the copyright tradition follow similar principles. In the US, the Copyright Act protects computer programs as literary works (§101) since 1980, when the statute was amended upon the proposal of the Commission on New Technological Uses of Copyrighted Works (CONTU) to tackle the uncertainties faced by the Copyright Office when issuing copyright certificates on computer programs under its "rule of doubt" (Goldstein 2005, 366). Federal case law has soon adopted the same approach (*Apple v. Franklin*, 714 F.2d 1240 (3d Cir. 1983)), further adapting general copyright principles to the new subject matter.

Exactly as for other works, abstract ideas, principles, methods are not copyrightable (US Copyright Act, §102(b)). As a consequence, only the exact copy of the code and graphic visualization may amount to infringement (*Williams Elecs .v. Artic*, 685 F.2d 879 (3rd Cir. 1982)). The protection covers both literal and non-literal elements of the software, such as its sequence, structure and organization (SSO). However, the idea-expression dichotomy requires these elements to "incorporate authorship in programmer's expression of original ideas, as distinguished from the ideas themselves" (*Computer Assocs. Int'l v. Altai, Inc.,* 982 F.2d 693 (2d Cir. 1992), which introduced the so-called "abstraction-filtration-comparison" test to distinguish literal from non-literal elements and non-protectable ideas from protectable expressions). By the same token, user interfaces are protected only in their "look-and-feel" (*Apple Computer, Inc. v. Microsoft Corporation*, 35 F.3d 1435 (9th Cir. 1995)), and only to the extent that graphic choices are not dictated by functional needs or rules. The set of operations available

through it can be only subject to a utility patent (*Lotus Dev. Corp. v. Borland Int'l, Inc.*, 516 U.S. 233 (1996)).

3. The text and data mining (TDM) exception

Text and data mining (TDM) is a category of analytical tools that constitutes the backbone of the data science society. As per the definition provided by the EU legislator in the recent Directive on Copyright in the Digital Single Market (2019/720/EU, CDSMD), TDM is "any automated analytical technique aiming to analyse text and data in digital form in order to generate information such as patterns, trends and correlation" (Article 2(2) CDSMD), also framed as "the automated computational analysis of information in digital form, such as text, sounds, images or data" (Recital 8). TDM technologies enable the generation of new knowledge by inferring conclusions from the correlation of structured and unstructured data, no matter if generated for that or other purposes. Their strength lies in the capability to unveil patterns that could have been identified difficultly – if at all – with standard tools. This is made possible by machine-learning (ML) algorithms, which drastically decrease the time, money and risk connected to the analysis of data corpora (Geiger et al 2015, 686).

A research carried via TDM is made of several steps. First, it identifies works and data to be analyzed, either already collected or to be collected in a database. Then, it copies such materials by (i) turning them into a machine-readable format that could be processed by the TDM technology used and (ii) often uploading the materials so processed on a platform. Data are then extracted, and finally recombined to identify specific patterns (Han-Kamber-Pei, 144; Weiss-Indurkhya-Zhang, 15; Traille et al, 28). This chain of activities intersects with several uses covered by copyright or *sui generis* database right. Crawling and/or extraction may require, in fact, the reproduction of protected works or the substantial extraction of data from a database protected by *sui generis* right (see entry *Database protection*), which are exempted from liability for infringement only if qualitatively and quantitatively minimal. While there are extraction techniques that are capable of reproducing and processing only very small fragments or single data, thus falling below the threshold of infringement, they represent a minoritarian part of all TDM technologies (Traille et al., 29). Also pre-processing to turn the material into a machine-readable form usually entails reproduction, and the same can be said for the upload on a platform and the actual mining. In some instances, reporting the TDM results to verify them within the community of peers may infringe the right of communication to the public (Geiger-Frosio-Bulayenko 2019, 822).

To balance exclusive rights with conflicting rights and the public interest, copyright laws provide for exceptions and limitations, variously structured. Continental legal systems and EU copyright law opt for closed and exhaustive lists of exceptions, while Anglo-Saxon systems, based on the copyright model, use open-ended flexible clauses. The latter have obviously enjoyed more room to adapt their regulations to the needs engendered by TDM technologies, the most paradigmatic example coming from the US, where the fair use doctrine has been able to accommodate such necessities.[1] On the contrary, few of the exceptions provided by the InfoSoc Directive (2001/29/EC) could be used for the same goal, only one of them being mandatory and thus fully harmonized across the Union

[1] From *Baker v Selden*, 101 U.S. 99 (1880), US courts have qualified as fair the use of protected works which was a necessary incident to the use of unprotected materials. In *Authors Guild v Google*, 804 F.3d 202 (2nd Circ. 2015) e *Authors Guild v HathiTrust*, 755 F.3d 87 (2nd Cir. 2013), also known as the Google Books cases, a TDM activity has been considered a transformative use, thus fair under §107 of the US Copyright Act.

(ephemeral reproduction (Article 5(1)). This great legal uncertainty and fragmentation has hindered cross-border TDM activities, and the situation has not got better with the introduction of specific TDM exceptions in some Member States, such as the UK, France, Estonia, Germany (Rosati 2018a, 8; Benabou, 762; Geiger-Frosio-Bulayenko 2019, 836; Splinder, 1113). As a response, and after a very long and articulated debate, the CDSMD Directive has introduced two specific and mandatory TDM exceptions, the first purpose- and subject-specific (Article 3 CDSMD), the second of general nature and overridable by contract (Article 4 CDSMD).

Article 3 CDSMD requires Member States to provide an exception, not overridable by contract, to the copyright and *sui generis* rights over databases (Articles 5(a) and 7(1) Directive 96/9/EC), to the general reproduction right under Article 2 InfoSoc, and to the press publisher right (Article 15(1) CDSMD) for reproduction and extractions made by research organisations and cultural heritage institutions, in order to carry out TDM for scientific research on works and other subject matter to which they have lawful access. Copies of the latter should be stored with an appropriate level of security, and may be retained for similar research purposes, such as the verification of research results. Rightholders are free to apply proportionate measures directed to preserve the security and integrity of their networks and databases. Article 4 CDSMD introduces a similar exception, this time extended also to the right of reproduction and distribution of the Software Directive (2009/24/EC). The provision is not purpose-specific but refers to general TDM activities, and it applies only if rightholders have not expressly reserved these uses in an appropriate matter, such as machine-readable means.

While it is undeniable that the CDSMD reform had made a substantial step forward in creating a level playing field for European researchers to exercise TDM activities, particularly in cross-border settings, decreasing the burden of costs and uncertainty put on research institutions, the solution adopted by the EU legislator is still far from perfect. Commentators have underlined how the cautious approach to the list of beneficiaries and the limited scope of the mandatory, non-overridable exception of Article 3 CDSMD excludes key players in the EU innovation ecosystem, such as SMEs, which are still subject to rightholders' stronger bargaining powers and rent-seeking attitude in contracting in or out TDM uses and determining their price (Geiger-Frosio-Bulayenko 2019, 838; Rosati 2018b, 428). This may have a strong impact on the competitiveness of the EU data industry *vis-à-vis* top innovative economies such as the US, Canada and Israel (see Geiger et al, 24). At the same time, subordinating TDM to lawful access may still make such activities dependent on the payment of subscription fees, which rightholders may raise at any time to amortize the expected losses arising from the new exception, with obvious consequences in terms of budgetary constraints and impact on the scope and reach of research. The limitation of the exception to the right of reproduction, with the exclusion of distribution and communication to the public, constrains even more the potential of Articles 3 and 4 CDSMD, for it still leaves uncovered the possibility to share the results of TDM activities including parts of the original works, as in the case of natural language processing trained on a number of copyright-protected *corpora* (Margoni-Kretschmer 2018, 3). Last, the imposition of technological protection measures to protect rightholders' networks and databases, albeit within the boundaries of proportionality, may still negatively affect a range of users' rights and freedoms, which enjoy little or no tools to counterbalance the pervasiveness of rightholders' automated control over their works (Geiger-Frosio-Bulayenko, 2019, 840-844; Rosati 2018b, 429).

4. AI agents as authors

Works created by computer programs are generally protected under copyright. Common examples are musical compositions created by synthesizer, or drawings prepared via graphic design software. There is less agreement on whether the final result is not defined by the author but randomly generated by the software, with some authors accepting the possibility that also such works may be protected by copyright to the extent that the human creator gives the key input for the creation and impresses her creative choices (see the overview provided by Gaudamuz 2017; Ginsburg 2018; Ramalho 2017a; Ginsburg-Budiardjo 2019).

Few countries have a special regime for computer-generated works. In the UK, §9(3) CDPA rules that "in the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken". Subsequent cases (eg *Nova Productions Ltd v Mazooma Games Ltd* (2006) RPC 379) specified that the person responsible for the arrangements will depend on the factors weighed in, which can include *inter alia* the initiative to create the work, the proximity to the final act of creation (the closer to the final creation, the more likely to be in charge of the arrangements to create the work), or the extent to which the arrangements are responsible for the creation of the work – the latter factor putting more emphasis on the operation of the software). According to §78 CDPA, a work is computer-generated if it is created "by computer in circumstances such that there is no human author of the work". The duration of the right is 50 years from the calendar year of production, and no moral rights are attributed.

Absent a legislative intervention, the solution to the interpretative dilemma lies in the definition of authorship and the requirement of originality provided by national copyright statutes. In the US, for instance, the Copyright Act does not define who is the author, but a number of other provisions hint to the fact that the category is limited to natural persons, such as §101, which qualifies as anonymous works those where no "natural person" is identified as the author. Along these lines, cases like *Feist* (*Feist Publications, Inc. v. Rural Telephone Service Co.*, 499 U.S. 340 (1991)) and *Burrow-Giles* (*Burrow-Giles Lithographic Co. v. Sarony*, 111 U.S. 53 (1884) have emphasized the need for the author who claims infringement to prove the intention to create, in the form of intellectual production, thought and conception – actions that can difficultly be linked. On this basis, the Compendium of US Copyright Office practice stated that "to qualify as a work of 'authorship', a work must be created by a human being", denying registration for works created by animals, plants or machines and mechanical processes that operate randomly or automatically without any creative input from a human author (Compendium 2017, §306).

In Australia, *Acohs v Ucorp* (2010 FCA 577) excluded that a document produced automatically by a software from a database could be protected by copyright, since no single human author was involved, and no originality could be found. In the EU, only the Software, Database and Rental (2006/110/EC) Directives define the author as a natural person, a group of natural persons, or a legal person. In the Explanatory Memo to the Database Directive (p.5), the Commission reinstated that the key principle of the Berne Convention is that the human author creating the work is the first copyright owner. Along the same lines, in the original draft of the Software Directive only natural persons could be defined as authors, and despite the final version changed the approach, the Explanatory Memo (p.6) still clarified that "the human input as regards the creation of machine generated programs may be

relatively modest, and will be increasingly modest in the future. Nevertheless, a human 'author' in the widest sense is always present, and must have the right to claim 'authorship' of the program".

Another legislative obstacle to the protection of computer-generated works comes from the requirement of originality. In the EU, the CJEU has ruled that in order to be protected, a work should represent the author's own intellectual creation (case C-5/08, *Infopaq International A/S v Danske Dagblades Forening*, EU:C:2009:465). The principle was later developed to mean that a work should be the product of free and creative choices, not constrained within strict rules, and containing a personal touch (case C-604/10 *Football Dataco Ltd and Others v Yahoo! UK Ltd and Others* [2012], EU:C:2012:115; case C-393/09 *Bezpečnostní softwarová asociace - Svaz softwarové ochrany v Ministerstvo kultury*, EU:C:2010:816; case C-145/10 *Eva-Maria Painer v Standard Verlags GmbH and Others*, EU:C:2011:798; joined cases *Football Association Premier League et al v QC Leisure et al (C-403/08) and Karen Murphy v Media Protection Services Ltd (C-429/08),* EU:C:2011:631). In the US, the same concept has been expressed in *Burrow-Giles*, which defines a choice as creative if made independently by the author, and not dictated by the function of the work, the method or technique used, or by applicable standard or relevant good practice. In general, most of the jurisdictions have abandoned the so-called "sweat-of-the-brow" doctrine, which deemed enough to prove the investment of time and effort to generate the work in order to grant protection, in favor of the adoption of a more substantive originality requirement, differently graded (Ramalho 2017a, 16).

When confronted with the question of whether a computer-generated work may be protected, the notion of originality may act as an obstacle rather than as an enabler. In fact, creativity is closely linked with personality and individuality, and so is the notion of freedom of creative choice – all concepts quite far from the functionality of a machine and their artistic expression. The EU legislator seems to be fully aware of it, as proven by the Resolution on Civil Law Rules on Robotics (2017), where the EU Parliament has called the Commission to "support a horizontal and technologically neutral approach to intellectual property applicable to the various sectors in which robotics could be employed", also by engaging in "the elaboration of criteria for "own intellectual creation" for copyrightable works produced by computers or robots". Several policy considerations will need to be weighted, the most relevant ones being the compatibility of the system and rationale of copyright incentives with computer-generated works, and the higher suitability of other bodies of law to regulate the matter. The question is still on the table, and it will be most probably be subject to regulatory proposals in the near future.

5. Towards a data producer's right?

Against the steady growth of the big data economy and the inaptitude of the intellectual property system to provide an adequate solution to the need of protection of raw data, some industries – and particularly the automotive sector – started advocating for the legislative creation of a new property right in data. In its Communication on "Building a European Data Economy" of 2017, the European Commission proposed the introduction of a data producer's right, based on the German concept of *Dateneigentum* (data property) which had been long advocated for by several scholars (Hoeren 2014, 751; Lehmann 2015, 51 et seq.; Zech 2016a, 464) and had its strongest champion in Commissioner Oettinger, leading the Directorate General that was responsible for the Communication (DG CONNECT). The Communication was accompanied by a Staff Working Document on the free flow of data and emerging issues of the European Data Economy, followed by a public consultation,

focused on non-personal, machine generated data and their strategic market importance for the EU against its strongest competitors, and chiefly against the United States (SWD 2017, 3-4, 13).

At first, the Communication clarified that "the Database Directive did not intend to create a new right in the data. The CJEU thus held that neither the copyright protection provided for by the Directive nor the *sui generis* right aim at protecting the content of databases. Furthermore, the ECJ has specified that the investment in the creation of data should not be taken into account when deciding whether a database can receive protection under the *sui generis* right". Then, it focused on the need for market players to access to diversified and broad dataset, and thus for the legislator to incentivize the sharing and free flow of data, avoid lock-in effects, while at the same time protecting investments (Communication 2017, 4). It excluded personal data from the scope of the Communication, covering only non-personal and machine-generated data, in light of the fact that the exchange and access to such types of information is limited and dependent on the policies of companies owning the data generated by their products or services, which are generally keener not to share them (Communication 2017, 9). The Commission tabled a number of proposals to increase access and sharing of data. Among them, it proposed the development of technical solutions to identify and exchange data, the development of FRAND (fair, reasonable and non-discriminatory) principles to facilitate access to data against remuneration, or the introduction of a data producer's right (Zech 2016a, 461). The latter proposal was justified by the fact that no IP right had proven capable of protecting raw machine-generated data, forcing producers to regulate their exchange and use by contract, causing fragmentation and a higher risk of market failure (Communication 2017, 10).

The data producer's right envisioned in the Communication had the goal of "clarifying the legal situation and giving more choice to the data producer, by opening up the possibility for users to utilize their data and thereby contribute to unlocking machine generated data" (SWD 2018, 35). It included non-personal or anonymized machine-generated data, including metadata. In order to preserve the public domain on information, the Commission specified that the data covered by exclusivity should have been only at the syntactical (signs such as binary sequences) and not at the semantic (i.e. content) level (Zech 2016b, 53-54). The right was attributed to the owner or long-term user of the device producing the data, and consisted of "a set of rights enforceable against any party independent of contractual relations thus preventing further use of data by third parties who have no rights to use the data" (SWD 2017, 33-34). Exceptions were envisioned for public authorities, for research purposes, and for the lawful use of the manufacturer of the device, in order to balance their commercial interests with those of the data producers and to allow them monitoring the functioning of their products if so requested by law (SWD 2017, 36).

The proposal was heavily criticized by scholars (see, eg, Hugenholtz 2018; Ramalho 2017b; Drexl et al 2017), who pointed out that an all-encompassing property right in data would seriously compromise the European IP law system, contravene fundamental freedoms, distort freedom of service and competition, restrict scientific freedoms and undercut the promises of innovation carried by big data. Several industry sectors voiced strong concerns, being afraid of the prejudicial impact of a strong exclusivity over raw data on their R&D potentials, and of the excessive transaction costs that could have arisen because of the extreme market fragmentation that could have ensued from the introduction of a data producer's right such as the one envisioned by the Commission. As a

consequence, the proposal was abandoned in subsequent preparatory works, but the debate is yet to be over.

BIBLIOGRAPHY

Bath, U (2002), Access to Information v Intellectual Property Rights, 23 European Intellectual Property Review 138.

Benabou, V (2016), La loi pour une République numérique et la propriété intellectuelle, Dalloz IP/IT 533.

Berne Convention for the Protection of Literary and Artistic Works, Sept. 9, 1886, as revised at Paris on July 24, 1971 and amended in 1979, S. Treaty Doc. No. 99-27 (1986).

De Franceschi, A., Lehmann, M (2015) Data as tradable commodity and new measures for their protection, 1 Italian Law Journal 51.

Derclaye, E (2000), Software copyright protection: can Europe learn from American case law?, 22(1) European Intellectual Property Review, 7.

Drexl et al. (2017), Position Statement of the Max Planck Institute for Innovation and Competition of 26 April 2017 on the European Commission's "Public Consultation on Building the European Data Economy, available at https://pure.mpg.de/rest/items/item_2450915_5/component/file_2450913/content.

European Commission (2017), Communication 'Building A European Data Economy', 10 January 2017, COM(2017) 9 final.

European Commission (1988), Explanatory memorandum to the Proposal for a Council Directive on the legal protection of computer programs, 5 January 1989, COM(88) 816 final.

European Commission (1992), Explanatory memorandum to the Proposal for a Council Directive on the legal protection of databases, COM(92) 24 final.

European Commission, Staff Working Document on the free flow of data and emerging issues of the European data economy', 10 January 2017, SWD(2017) 2 final.

European Parliament, Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics, 2015/2103(INL)

Geiger, C, Frosio, G, Bulayenko, O (2019), Text and data mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?, 49(7) International Review of Intellectual Property and Competition Law 814.

Gaudamuz, A (2017), Do Androids Dream of Electric Copyright? Comparative Analysis of Originality in Artificial Intelligence Generated Works, 2 Intellectual Property Quarterly 169.

Geiger, C, Bulayenko, O, Hassler, T, Izyumenko, E, Shoenherr, F, Seuba, X (2015), The Resolution of the European Parliament of 9 July 2015: Paving the Way (finally) for a Copyright Reform in the European Union?, 37(11) European Intellectual Property Review 683.

Gervais, D (2017) The TRIPS Agreement: Drafting History and Analysis, 4th ed., Sweet and Maxwell, London.

Ginsburg J, Budiardjo L (2019) Authors and Machines, 34 Berkeley Technology Law Journal 343.

Ginsburg, JC (2018), People Not Machines: Authorship and What It Means in the Berne Convention, 49 IIC - International Review of Intellectual Property and Competition Law, 131-135

Goldstein, P (2005), Goldstein on Copyright, vol II, New York, Aspen.

Han, J., Kamber, M, Pei, J (2011), Data Mining: Concept and Techniques, 3rd ed., Waltham, Morgan Kaufmann.

Hoeren, T (2014) Big Data and the Ownership in Data: Recent Developments in Europe, 36(12) European Intellectual Property Review 751.

Hugenholtz, PB (2018), Against Data Property, in Ullrich, H, Drahos, P, Ghidini, G (ed), Kritika. Essays on Intellectual Property, vol.III, 48-71, Cheltenham-Northampton, Edward Elgar

Kerber, W (2016), A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis", 11 GRUR Int 989.

Landes, WM, Posner, MA (1989) An Economic Analysis of Copyright Law, 18 Journal of Legal Studies 325.

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, available at http://www.create.ac.uk/blog/2018/04/25/why-tdm-exception-copyright-directivedigital-single-market-not-what-eu-copyright-needs/

Masouye, C (1978), Guide to the Berne Convention for the Protection of Literary and Artistic Works, WIPO, Geneva.

Perry M., Margoni, T (2010), From Music Tracks to Google Maps: Who Owns Computer-generated Works?', 26 Computer Law & Security Review 621.

Ramalho A (2017a), Will Robots Rule the (Artistic) World? A Proposed Model for the Legal Status of Creations by Artificial Intelligence Systems, 21 Journal of Internet Law 12.

Ramalho A (2017b), Data producer's right: powers, perils and pitfalls, in Proceedings from the conference "Better Regulation for Copyright. Academics Meet Policy Makers", Brussels, 6 September 2017, 51-58, available at https://eprints.soton.ac.uk/414799/1/2017_09_06_Better_Regulation_for_Copyright_Academics_meet_Policy_Makers_Proceedings.pdf

Rosati, E (2018a), The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market - Technical Aspects, Briefing requested by the JURI committee of the European Parliament, Brussels, Publication Office of the EU Parliament.

Rosati, E. (2018b), An EU text and data mining exception for the few: would it make sense?, 13(6) Journal of Intellectual Property Law & Practice 429

Splinder, G. (2016), Text und Data Mining. Urheber- und datenschutzrechtliche Fragen, GRUR Int. 1112.

Strowel, A (1991) *Droit d'auteur* et *copyright*: divergences et convergences: étude de droit compare, Bruylant, Bruxelles

TRIPS: Agreement on Trade-Related Aspects of Intellectual Property Rights, Apr. 15, 1994, Marrakesh Agreement Establishing the World Trade Organization, Annex 1C, 1869 U.N.T.S. 299, 33 I.L.M. 1197 (1994).

Triallie, JP, De Meeus D'Argenteuil, J, De Francquen, A (2014), Study of the Legal Framework of Text and Data Mining (TDM), prepared for the European Commission.

U.S. Copyright Office (2017), Compendium of US Copyright Office Practices, 3rd ed., Washington D.C.

van Erp, S (2017), Ownership of Data: The Numerus Clausus of Legal Objects, 6 Brigham-Kanner Property Rights Conference Journal 235.

Walker, MM, W, von Lewinski, S (2010), European Copyright Law. A Commentary, Oxford, Oxford University Press.

Weiss, S, Indurkhya, N, Zhang, T (2010), Fundamentals of Predictive Text Mining, Texts in Computer Sciences, Amsterdam, Springer.

WIPO Copyright Treaty, 36 ILM 65 (1997).

Zech, H (2016a), A Legal Framework for a Data Economy in the European Digital Single Market: Rights to Use Data, 11 Journal of Intellectual Property Law & Practice 460.

Zech, H (2016b), Data as a Tradeable Commodity, in De Franceschi A (ed), European Contract Law and the Digital Single Market. The Implications of the Digital Revolution, Intersentia, 2016, 53-77.