9th Conference on Grey Literature and Repositories: proceedings [online]. Prague: National Library of Technology, 2016 [cit. 2016-12-5]. Available from: **http://nrgl.techlib.cz/conference/conference-proceedings/**. ISSN 2336-5021.

TEXT AND DATA MINING OF GREY LITERATURE FOR THE PURPOSE OF SCIENTIFIC RESEARCH

Matěj Myška[1]

matej.myska@law.muni.cz

Masaryk University, Faculty of Law, Institute of Law and Technology, Czech Republic

Abstract

This paper explores the legal possibilities of users to text and data mine repositories of grey literature for scientific research without the consent of the grey literature repository operator or the holders of rights to the content stored therein. In the first part of this short paper we briefly introduce the relevant intellectual property rights. In the second part, the current exceptions to these exclusive rights are discussed and evaluated. The third part critically analyses the suggested mandatory exception for text and data mining in the European Commission's proposal for a new directive on copyright in the Digital Single Market.

Keywords

Text and Data Mining, Exceptions, Sui Generis Database Rights, Copyright

The publication of this paper is supported by the Czech Science Foundation - project Legal Framework for Collecting, Processing, Storing and Utilizing of Research Data - registration no. GA15-20763S.

[1] I would like to thank Jakub Harašta for critical revision of this paper. However, all mistakes and omissions are mine.

Introduction

As observed by Floridi in 2014, the development of ICTs brought us into the zettabyte era, where tsunami of bytes submerge our environments (2014, p. 13). The amount of data and information produced is constantly growing, as is the number of scientific research papers (i.e. white literature).[2] Even though there are no similar exact data on the rising amount of grey literature, given the fact that GL is usually available online without the traditional constraints of white literature publishing (Banks 2006, p. 5), it can be reasonably assumed that the number is not declining. Jeffery and Asserson even go as far as proclaiming that the vast majority of research output is grey (2014, p. 223).

On the other hand, this technological development not only brought us the age of data deluge (Borgman 2012, p. 1059), but also offered researchers various new and innovative tools and techniques[3] to effectively process this vast amount of data and information (including GL)[4] in an automated way. These are commonly labelled "text and data mining".[5] In fact, the whole process of academic research has been reshaped with the technological development, in that TDM can be employed in all of the stages thereof. Technology can thus free up the time that would otherwise be spent on *"finding ideas for a research paper, literature review and formulation, data and methodology and analysis of results"* (Filippov, Hofheinz 2016, p. 5).

However, the legality of TDM remains unclear under European intellectual property laws[6] and subsequently the national law of Member States. Potential infringement of exclusive rights is presented in the first part of this paper. The second part discusses the current exceptions that are potentially applicable to TDM. The third part introduces and critically analyses the proposed mandatory exception for TDM in the proposal for a new directive on copyright in the Digital Single Market (hereafter the "Proposal").[7]

[2] The number of scientific papers published was estimated at around 50 million in 2010 (Larsen, von Ins 2010). The number continues to grow by 2.5 million a year and is steadily on the rise (Ware, Mabe 2015, p. 6), See also (Bornmann, Mutz 2015).

[3] For an overview of various types of data mining methods and applications from the technological point of view see (Colonna 2013). Colonna also critically notes that this "buzzword" is so semantically obfuscated that it is starting to lose its meaning (2013, p. 309). For further technical details of TDM see, e.g., (Larose 2014).

[4] Due to its variety, GL is an ideal raw material for TDM (Schöpfel, 2010, p. 29).

[5] However, automated processing might also include protected or unprotected content other than text and data and so "content mining" would probably be a more accurate term (especially in the context of GL) (Murray-Rust, Molloy, Cabell 2014, p. 11).

[6] In this paper, we deal mainly with the two most relevant directives, namely Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (hereafter "ISD") and Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases (hereafter "DD"). For further details on European copyright law see, e.g., (Walter, Lewinski 2010; Stamatoudi, Torremans 2014). For treatment of TDM outside Europe see, e.g., (European Commission, Directorate-General for Research and Innovation 2014, pp. 44–48).

[7] Proposal for DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on copyright in the Digital Single Market of the European Commission. Brussels, 14.9.2016, COM(2016) 593 final, 2016/0280 (COD).

TDM as possible infringement of intellectual property rights

The principle of legal license, i.e. that *"everything which is not forbidden is allowed"*[8], forms the basis for further inquiries. Consequently, text and data mining is only relevant from the legal point of view when it encroaches upon protected subject-matter. We further focus only[9] on 1) copyright protection[10] (and specifically the right of reproduction) and 2) the sui generis database right (specifically extraction right).[11] Such choice is justified by the fact that any processing of digital content[12], and implicitly TDM, includes making (at least a temporary) copy[13] of the subject-matter being processed. (Triaille et al., 2014, p. 28). Such copying would involve copyrighted works[14] (contained in, e.g., the GL repository). Moreover, when the structure/arrangement of the database is copied, the copyright of the database creator is infringed. Finally, reproduction of whole/substantial parts of (otherwise unprotected) content of a database[15] is prohibited by the sui generis right of extraction. In order to avoid liability for copyright/sui generis database rights infringement, the miner must rely either on the consent of the holder of rights or on existing exceptions to these exclusive rights.

Potentially applicable exceptions

The first potentially applicable copyright exception is the one that allows for a **temporary copy to be made (Art. 5(1) ISD).** Such a copy must be transient and incidental (not permanent) and for the duration of TDM only (i.e. the protected subject matter being mined must be automatically deleted afterwards). Furthermore, it must be an integral and essential part of the technological process, which should not pose a problem as TDM is essentially a technology built on making copies. Next, the copy must enable lawful use – lawfulness might be based on the application of another exception[16] or "using" the works in a way that is lawful, e.g. mining for new information. Finally, the allowed copy must be without independent economic significance.[17] This exception, with its extensive pre-requisites, is applicable in only a few

[8] As expressed, e.g., in Art. 2(4) of the Constitution of the Czech Republic (Constitutional Act No. 1/1993 Sb., as amended).

[9] For a detailed discussion of legal aspects of TDM see (Triaille et al. 2014; Truyens, Van Eecke 2014); for a concise analysis of EU and Czech Law see (Myška, Harašta 2015).

[10] Including copyright protection for the selection and arrangement of the contents and structure of the database.

[11] TDM techniques could also infringe the rights to privacy and protection of personal data of a natural person. However, these issues are not discussed in this paper. For discussion of this topic see, e.g., (Rubinstein 2013).

[12] As noted by (Truyens, Van Eecke 2014, p. 158).

[13] This copy is usually done in the process of creating corpora for subsequent analysis (Truyens, Van Eecke 2014, pp. 154–155).

[14] Provided that they are protectable subject-matter, i.e. that they are original as stated by the CJEU, e.g., in Infopaq I, C-5/08, ECLI:EU:C:2009:465. For details on originality in EU copyright law see (Rosati 2013).

[15] A database is protected by the sui generis rights if the database owner made substantial investment in obtaining, verifying and presenting the contents of the database (Art. 7 DD).

[16] E.g. the "scientific research" exception specified hereunder.

[17] CJEU stated in Infopaq II, C-301/10, ECLI:EU:C:2012:16, para. 54, that this condition precludes modification of the used protected subject-matter under this exception.

cases (especially due to the non-permanent character of the copies) and does not provide enough legal certainty.

The exception for the purpose of **scientific research** (Art. 5(3)(a) ISD) allows for the reproduction of protected subject-matter solely for that purpose, to the extent justified by the non-commercial purpose to be achieved and under the condition of indicating the source (including author's name), if it is not impossible.[18] This exception provides prima facie solid foundations for performing TDM. However, the non-commercial criterion significantly reduces the scope of applicability, especially in borderline cases where commercial entities participate in the research. Further, TDM relies on the quantity of the mined content – the scientific research exception, however, suggests a rather restrictive approach as reproductions should be made only to the above-mentioned *"extent justified by the non-commercial purpose".*

Apart from the temporary copy exception, the other potentially applicable copyright exceptions in ISD are only facultative "blueprints" and may not be transposed into the respective national law. However, the assessment of the legality of TDM (including TDM of GL) will be dependent on the applicable national law.[19]

Regarding the protection of **databases,** DD foresees a facultative exception for scientific research (Art. 6(2) DD) allowing for the reproduction of copyrightable structure/arrangement thereof. As these pre-requirements do not differ in further details, the short remarks made to the copyright exception for scientific research are also applicable here. Moreover, DD also includes a counterpart of the "temporary copy" exception in Art. 6(1) DD which is also mandatory. This exception allows the lawful user to reproduce the arrangement/structure of the database *"which is necessary for the purposes of access to the contents of the databases and normal use"* thereof. The sui generis right of extraction might also be limited for scientific research under same preconditions (Art. 9(b) DD), but only for the "lawful user".[20] Such subject might also extract unsubstantial parts of the content from the database made available to the public for any purpose (Art. 8 DD). However, the user may not do so in a repeated and systematic manner in a way that would imply *"acts which conflict with a normal exploitation of that database or which unreasonably prejudice the legitimate interests of the maker of the database"* (Art. 7(5) DD).[21] Such an exception is thus practically irrelevant to TDM as it relies on extraction of all or at least substantial parts of the content. Similarly, as far as copyright exceptions are concerned, the national transpositions of the facultative exceptions (including the one for scientific research) differ, which consequently leads to undesired legal uncertainty (Triaille et al., 2014, pp. 80–81).

[18] As the number of works used in TDM is usually fairly high, this safeguard clause could be used.

[19] For an overview of different implementations of ISD see in detail, e.g., (Hugenholtz, Eechoud, Gompel, Helberger 2006; Westkamp, Guibault, Rieber-Mohn 2007; Eechoud, Hugenholtz, van Gompel, Guibault, Helberger 2009; Triaille, Dusollier, Depreeuw, Hubin, Coppens, Francquen 2013). The transpositions differ, e.g., in scope. In BE, LU and IT, only parts of the works might be used, which poses a practical problem for TDM. (Triaille et al. 2014, p. 56).

[20] This concept itself is rather opaque. For a detailed discussion see, e.g., (Derclaye 2008, pp. 120–126). However, in the context of GL mining this condition will be fulfilled rather easily as the GL repositories are usually made available online to all users without restrictions.

[21] For the complicated interplay between Art. 8 and 7(5) DD, see in detail (Triaille et al. 2014, pp. 78–79) and sources and case law of CJEU cited therein.

It could thus be concluded that the current system of exceptions does not provide enough legal certainty for TDM,[22] which could, as Renda et al. suggest, be consequently *"detrimental to the development of new offers and services, which in turn limits benefits to society through a direct negative impact on so-called 'dynamic efficiency' (e.g. innovation and the development of new welfare-enhancing products and services)."* (2015, p. 131). Handke et al. (2015, p. 21) even showed that as far as data mining research is concerned, copyright seems to have a negative net effect on innovation.

The proposed text and data mining exception

To eliminate the problems described above, a specific TDM exception was proposed by the European Commission in the recently introduced reform of European copyright law.[23]

TDM is defined in the Proposal 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".* According to Art. 3 of the Proposal, the beneficiary of this exception should be, without the consent of the respective right holder, allowed to 1) directly or indirectly, temporarily or permanently reproduce by any means and in any form, in whole or in part the copyrighted works; [24] (2) temporarily or permanently reproduce the structure of the database by any means and in any form, in whole or in part and (3) extract and/or re-utilize of the whole or of a substantial part of the contents of the database.[25] These activities could in the end-effect lead to commercialisation of the results, as the exception does not restrict the nature thereof. The pre-requisite invariably required for application of the exception is lawful access to the protected subject-matter that is to be mined and the scientific research purpose to be achieved. In accordance with the suggestions of various expert groups,[26] such an exception shall not be overridden by contract. Contractual arrangements to the contrary shall be deemed unenforceable (Art. 3(2) Proposal). In order to avoid potential overloading and security/integrity breaches of the networks and databases, the right holders shall be allowed to implement proportionate measures (Art 3(3) Proposal). Lastly, the overprotective application of such preventive measures shall be avoided by defining *"commonly-agreed best practices"* between the right holders and researchers (Art. 3(4) Proposal). No fair compensation for the use allowed under this exception that is comparable to, e.g., the private copying levies, is foreseen for the TDM exception. It is also subject to the three-step test, which again limits its scope of application (Art. 6 Proposal). The complicated provision of Art. 6(4) ISD, regulating the relation between technological measures of protection (DRM) and exceptions to exclusive rights, shall also apply. Consequently, if the works/repository is

[22] This view was also expressed by the respondents to the European Commission consultation on review of EU copyright rules (*Report on the responses to the Public Consultation on the Review of the EU Copyright Rules* 2014, p. 63).

[23] Article 3(1) of the Proposal: "Member States shall provide for an exception to the rights provided for in Article 2 of Directive 2001/29/EC, Articles 5(a) and 7(1) of Directive 96/9/EC and Article 11(1) of this Directive for reproductions and extractions made by research organisations in order to carry out text and data mining of works or other subject- matter to which they have lawful access for the purposes of scientific research."

[24] Art. 2 ISD.

[25] Furthermore, the "press publishers right" according to Art. 11(1) of the Proposal shall not be infringed by acts under the TDM exception.

[26] See, e.g., (European Commission, Directorate-General for Research and Innovation 2014, p. 52).

protected by technological measures, the right holders should voluntarily allow the beneficiaries to make use of the TDM exception. In the absence thereof, Member States should provide for appropriate measures so that the beneficiaries may actually profit from the TDM exception.

A fundamental flaw of the proposed TDM exception lies in the restricted personal scope. The beneficiary must be a public research institution (Art. 2(1), i.e. "*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".* Furthermore, these activities must be done *"(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"* and 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;"*

This relatively complicated precondition should eliminate business as beneficiaries, but should enable Public-Private Partnerships.[27] As the League of European Research Universities remarks, this limits significantly the practical value of TDM (League of European Research Universities 2016). It could be argued that such an exception is discriminatory in its nature. The NGO Communia Association even goes as far as to say that this regulation creates a *"privileged class of data miners"* by pointing out that the proposed article could be interpreted in such a way that any other subject (i.e. not research organisations) must specifically ask for the consent of the right holders to mine (Vollmer 2016). Cocoru and Boehm see this condition as *"just another way of reflecting the commercial vs. non-commercial debate"* (2016, p. 8). The uncertainties raised would, according to them, only benefit those *"who can afford to capitalise on legal uncertainty"* (Cocoru, Boehm, 2016, p. 8). A broader approach to this exception would lead to *"more competition, broader services, and more creation and dissemination of knowledge"* (Cocoru, Boehm, 2016, p. 18). A logical solution to this problem is the elimination of this restrictive criterion.

In the context of mining **GL repositories**, the proposed solution would seem to be favourable for miners. Unlike the traditional databases of white literature, GL and its repositories are usually not hidden behind a paywall and are made available online without further contractual restrictions. Consequently, the prerequisite of lawful access should be fulfilled as standard. By contrast, for the operators of GL repositories, the exception, when adopted in its current format, would potentially lead to heightened traffic and usage of their repositories. They could, however, employ the required adequate measures to prevent the collapse of these. Logically, criticism regarding the scope of beneficiaries, as provided above, remains the same and is even more substantiated due to the value and importance of GL.[28]

[27] COMMISSION STAFF WORKING DOCUMENT IMPACT ASSESSMENT on the modernisation of EU copyright rules. Accompanying the document Proposal for Directive of the European Parliament and of the Council on copyright in the Digital Single Market and Proposal for a Regulation of the European Parliament and of the Council laying down rules on the exercise of copyright and related rights applicable to certain online transmissions of broadcasting organisations and retransmissions of television and radio programmes. Brussels, 14.9.2016 SWD(2016) 301 final. Part 1/3, p. 109.

[28] See more on this: e.g., (Myška, Šavelka 2013, para. 3). GL also comprises the full spectrum of content to be mined, not only text and data. What is more, GL comprises detailed data and information not available in standard distribution channels (Castro, Salinetti 2004, p. 4).

Conclusion

Due to the legal uncertainty and technical complexity of TDM, the status quo in terms of the possibility of mining content without the consent of the respective right holders is far from optimal. The users of TDM must rely on the national implementations of ISD exceptions, which are relatively restrictive, do not actually enable the use of TDM to its full potential and, in the case of the scientific research exception (Art. Art. 5(3)(a) ISD and 9(b) DD), may differ at a national level. Consequently, the content miner could well end up infringing both the rights of the GL repository operator (extraction, copyright) and of the right holders of the works contained therein (copyright).

A possible solution to the status quo, which must, however, be developed and investigated in further research, could be found in even broader opening of the repository by licensing it and the works contained therein openly (e.g. under the Creative Commons 4.0 Attribution International license,[29] which allows for TDM). The "all-in" solution of broadest possible license/waiver (where legally possible) of rights without any further obligations imposed on the potential miner is even more advisable. However, this might not always be so easily achieved because of, e.g., the obligations of the GL repository operator or the author of the mined works related to receiving public funding.[30]

This less-than-ideal state of affairs should be improved by the newly introduced TDM exception in the Proposal of the Commission, which seems to be *prima facie* beneficial to the content miner. Due to its limited personal scope, however, it does not fulfil its promising potential. These critical remarks also remain valid for the mining of GL.

De lege ferenda, an amendment to the current wording of the proposed TDM exception, as eliminating the remaining ambiguities of the Proposal, is advisable. Content mining should be allowed for *"anybody that has lawful access"* (European Commission 2016, p. 109) to the mined subject-matter without further conditions.[31] Furthermore, the delicate issues of technical protection of repositories/works should be regulated in more detail – ideally in such a way that the right holders cannot technically block activities that would be legally permitted under the exception. Yet again, this rather simplistic claim must be developed and investigated in further research. On the other hand, an attribution (information) obligation could be imposed on the beneficiary of the exception. Naming the works which have been mined (including their source, i.e. the repository from which they have been acquired) is not technically unfeasible, e.g. in the form of a link to a dedicated website where the sources would be mentioned.[32]

[29] Creative Commons 4.0 Attribution International License [online]. [Accessed 28 September 2016]. Available from: **https://creativecommons.org/licenses/by/4.0/**.

[30] As is the case in H2020 Framework Programme for Research and Innovation – Art. 43(4) Regulation (EU) No 1290/2013 of the European Parliament and of the Council of 11 December 2013 laying down the rules for the participation and dissemination in Horizon 2020 – the Framework Programme for Research and Innovation (2014-2020).

[31] Such wording of the TDM exception was initially considered as one possible solution, but was ultimately not chosen as it would allegedly have *"a considerable negative effect on publisher's TDM licensing market"* (European Commission 2016, p. 109).

[32] The author would like to express his thanks to prof. Christian Handke for suggesting this idea.

The abovementioned revisions would ultimately lead to achieving the Open Mining Manifesto,[33] would support the deployment of TDM and would consequently help to *"drive science, competitiveness and innovation"* in the EU (Association of European Research Libraries, 2015, p. 1) by increasing legal certainty.

References

ASSOCIATION OF EUROPEAN RESEARCH LIBRARIES, 2015. *A Copyright Exception for Text and Data Mining* [online]. 9 December 2015, [Accessed 28 September 2016]. Available from: **http://libereurope.eu/wp-content/uploads/2015/11/TDM-Copyright-Exception.pdf**.

BANKS, Marcus, 2006. Towards a Continuum of Scholarship: The Eventual Collapse of the Distinction between Grey and Non-Grey Literature. *Publishing Research Quarterly*. Spring 2006, **22**(1), 4–11.

BORGMAN, Christine L., 2012. The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology*. 2012, **63**(6), 1059–1078. DOI 10.1002/asi.22634.

BORNMANN, Lutz and Rüdiger MUTZ, 2015. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*. 1 November 2015, **66**(11), 2215–2222. DOI 10.1002/asi.23329.

CASTRO, Paola De and Sandra SALINETTI, 2004. Quality of grey literature in the open access era: Privilege and responsibility. *Publishing Research Quarterly*. **20**(1), 4–12. DOI 10.1007/BF02910856.

COCORU, Diane and Mirko BOEHM, 2016. *An analytical review of text and data mining practices and approaches in Europe* [online]. 1 May 2016. OpenForum Europe. [Accessed 28 September 2016]. Available from: **http://www.openforumeurope.org/wp-content/uploads/2016/05/TDM-Paper-Diana-Cocoru-and-Mirko-Boehm.pdf**.

COLONNA, Liane, 2013, A Taxonomy and Classification of Data Mining. *SMU Science & Technology Law Review*. 1 October 2013, **16**, p. 309.

DERCLAYE, Estelle, 2008. *The Legal Protection of Databases A Comparative Analysis*. Cheltenham, UK; Northampton, MA: Edward Elgar. ISBN 978-1-84720-133-1.

EECHOUD, Mireille M. M. van, HUGENHOLTZ, P. Bernt, VAN GOMPEL, Stef, GUIBAULT, Lucie M. C. R. and Natali HELBERGER, 2009. *Harmonizing European Copyright Law: the Challenges of Better Lawmaking*. Alphen aan den Rijn: Kluwer Law International. Information law series, Vol. 19. ISBN 978-90-411-3130-0.

EUROPEAN COMMISSION and DIRECTORATE-GENERAL FOR RESEARCH AND INNOVATION, 2014. *Standardisation in the area of innovation and technological*

[33] "The right to read is the right to mine"; "Users and providers should encourage machine processing"; "Facts don't belong to anyone" (Murray-Rust et. al 2014, pp. 26-29).

development, notably in the field of text and data mining: report from the Expert Group. [online]. Luxembourg: Publications Office [Accessed 28 September 2016]. ISBN 978-92-79-36743-4. Available from: **http://bookshop.europa.eu/uri?target=EUB:NOTICE:KI0114289:EN:HTML**.

EUROPEAN COMMISSION, 2016, COMMISSION STAFF WORKING DOCUMENT IMPACT ASSESSMENT on the modernisation of EU copyright rules. Accompanying the document Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market and Proposal for a Regulation of the European Parliament and of the Council laying down rules on the exercise of copyright and related rights applicable to certain online transmissions of broadcasting organisations and retransmissions of television and radio programmes. Brussels, 14.9.2016, SWD(2016) 301 final. Part 1/3.

FILIPPOV, Sergey and Paul HOFHEINZ, 2016. *Text and Data Mining for Research and Innovation What Europe Must Do Next. Interactive Policy Brief*. No. 20.

FLORIDI, Luciano, 2014. *The 4th revolution: how the infosphere is reshaping human reality*. First edition. New York ; Oxford: Oxford University Press. ISBN 978-0-19-960672-6.

HANDKE, Christian, GUIBAULT, Lucie and Joan-Josep VALLBÉ, 2015. *Is Europe Falling Behind in Data Mining? Copyright's Impact on Data Mining in Academic Research* [online]. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, [Accessed 28 September 2016]. Available from: **http://papers.ssrn.com/abstract=2608513**.

HUGENHOLTZ, P. Bernt, EECHOUD, Mireille M. M. van, GOMPEL, Stef Van and HELBERGER, Natali, 2006. *EUROPEAN COMMISSION DG INTERNAL MARKET STUDY CONTRACT NO. ETD/2005/IM/D1/95: The Recasting of Copyright & Related Rights for the Knowledge Economy.* Amsterdam: Institute for Information Law, University of Amsterdam.

JEFFERY, Keith G. and Anne ASSERSON, 2014. Data Intensive Science: Shades of Grey. Procedia Computer Science. *Procedia Computer Science* [online]. **33**, 223–230. DOI 10.1016/j.procs.2014.06.036. ISSN 18770509.

LAROSE, Daniel T., 2014. *Discovering Knowledge in Data: An Introduction to Data Mining*. Second edition. Hoboken: Wiley. Wiley Series on Methods and Applications in Data Mining. ISBN 978-0-470-90874-7.

LARSEN, Peder Olesen and Markus VON INS, 2010. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*. **84**(3), 575–603. DOI 10.1007/s11192-010-0202-z. ISSN 0138-9130.

LEAGUE OF EUROPEAN RESEARCH UNIVERSITIES, 2016, EU copyright reform and TDM : potentially good for research but certainly not (yet) for innovation! [online]. LERU : League of European Research Universities. LERU [online]. 14 September 2016. [Accessed 28 September 2016]. Available from: **http://www.leru.org/index.php/public/news/eu-copyright-reform-and-tdm-potentially-good-for-research-but-certainly-not-yet-for-innovation-/**.

MURRAY-RUST, Peter, MOLLOY, Jennifer C. and Diane CABELL, 2014. Open Content Mining. In: *Issues in Open Research Data* [online]. London: Ubiquity Press. p. 11–30. ISBN 1-909188-30-1. Available from: **http://dx.doi.org/10.5334/ban**.

MYŠKA, Matěj and Jakub HARAŠTA, 2015. Omezení autorského práva a zvláštních práv pořizovatele databáze v případě datové analýzy. *Časopis pro právní vědu a praxi.* **23**(4), 375–384.

MYŠKA, Matěj and Jaromír ŠAVELKA, 2013. A Model Framework for publishing Grey Literature in Open Access. *jipitec* [online]. 25 August 2013, **4**(2), 104-115 [Accessed 28 September 2016]. Available from: **http://www.jipitec.eu/issues/jipitec-4-2-2013/3744**.

RENDA, Andrea, SIMONELLI, Felice, MAZZIOTTI, Giuseppe, BOLOGNINI, Alberto and Giacomo LUCHETTA, 2015. *The implementation, application and effects of the EU Directive on Copyright in the information society* [online]. Brussels: Centre for European Policy Studies, [Accessed 28 September 2016]. CEPS Special Report, 120/2015. ISBN 978-94-6138-487-4. Available from: **https://www.ceps.eu/system/files/SR120_0.pdf**.

Report on the responses to the Public Consultation on the Review of the EU Copyright Rules [online], 2014. Brussels: European Commission Directorate General Internal Market and Services, [Accessed 28 September 2016]. Available from: **http://ec.europa.eu/internal_market/consultations/2013/copyright-rules/docs/contributions/consultation-report_en.pdf**.

ROSATI, Eleonora, 2013. *Originality In EU Copyright: Full Harmonization through Case Law*. Cheltenham, UK ; Northampton, MA: Edward Elgar. ISBN 978-1-78254-893-5.

RUBINSTEIN, Ira S., 2013. Big Data: The End of Privacy or a New Beginning? *International Data Privacy Law*. **3**(2), 74–87. DOI 10.1093/idpl/ips036.

SCHÖPFEL, Joachim, 2010. Access to European Grey Literature. In: *Grey Literature Repositories*. Zlín: VeRBuM. p. 20–33. ISBN 978-80-904273-6-5.

STAMATOUDI, Irini A. and Paul TORREMANS (eds.), 2014. *EU Copyright Law: A Commentary*. Cheltenham: Edward Elgar. Elgar Commentaries. ISBN 978-1-78195-242-9.

TRIAILLE, Jean-Paul, DUSOLLIER, Séverine, DEPREEUW, Sari, HUBIN, Jean-Benoit, COPPENS, François and Amélie de FRANCQUEN, 2013. *Study on the Application of Directive 2001/29/EC on Copyright and Related Rights in the Information Society* [online]. Brussels: European Commission, [Accessed 28 September 2016]. ISBN 978-92-79-29918-6. DOI DOI:10.2780/90141.

TRIAILLE, Jean-Paul, MEEÛS D'ARGENTEUIL, Jérôme de and Amélie de FRANCQUEN, 2014. *Study on the legal framework of text and data mining (TDM)* [online]. Luxembourg: Publications Office, [Accessed 28 September 2016]. ISBN 978-92-79-31976-1. Available from: **http://bookshop.europa.eu/uri?target=EUB:NOTICE:KM0313426:EN:HTML**.

TRUYENS, Maarten and Patrick VAN EECKE, 2014. Legal aspects of text mining. *Computer Law & Security Review*. April 2014, **30**(2), 153–170. DOI 10.1016/j.clsr.2014.01.009.

VOLLMER, Timothy, 2016. *Commission proposes to limit text and data mining in Europe. International Communia Association* [online]. 6 September 2016 [Accessed 28 September 2016]. Available from: **https://www.communia-association.org/2016/09/06/commission-proposes-limit-text-data-mining-europe/**.

WALTER, Michel M. and Silke von LEWINSKI (eds.), 2010. *European Copyright Law: A Commentary*. Oxford ; New York: Oxford University Press. ISBN 978-0-19-922732-7.

WARE, Mark and Michael MABE, 2015. *The STM report An overview of scientific and scholarly journal publishing* [online]. STM: International Association of Scientific, Technical and Medical Publishers, [Accessed 28 September 2016]. Available from: **http://www.stm-assoc.org/2015_02_20_STM_Report_2015.pdf**.

WESTKAMP, Guido, GUIBAULT, Lucie and Thomas RIEBER-MOHN, 2007. *MARKT/2005/07/D: 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* [online]. Amsterdam: Institute for Information Law, University of Amsterdam, [Accessed 28 September 2016]. Available from: **http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2006358**.