2015

Is Europe Falling Behind in Data Mining? Copyright's Impact on Data Mining in Academic Research

Christian Handke

Lucie Guibault

Joan-Josep Vallbé

New Avenues for Electronic Publishing in the Age of Infinite Collections and Citizen Science
B. Schmidt and M. Dobreva (Eds.)

doi:10.3233/978-1-61499-562-3-120

Is Europe Falling Behind in Data Mining? Copyright's Impact on Data Mining in Academic Research

Christian HANDKE[a,1], Lucie GUIBAULT[b] and Joan-Josep VALLBÉ[c]
[a]*Erasmus University Rotterdam*
[b]*University of Amsterdam*
[c]*University of Barcelona*

Abstract. With the diffusion of digital information technology, data mining (DM) is widely expected to increase the productivity of all kinds of research activities. Based on bibliometric data, we demonstrate that the share of DM-related research articles in all published academic papers has increased substantially over the last two decades. We develop an ordinal categorization of countries according to essential aspects of the copyright system affecting the costs and benefits of DM research. We demonstrate that countries in which data mining for academic research requires the express consent of rights holders, data mining makes up a significantly smaller share of total research output. To our knowledge, this is the first time that an empirical study identified a significant negative association between copyright protection and innovation. We also show that within countries where DM requires express consent by rights holders, there is an inverse relationship between rule of law indicators and the share of DM related articles in all research articles.

Keywords. Copyright, data mining, research output

1. Introduction

This paper discusses the effect of different copyright arrangements on data mining (DM) by academic researchers.[2] Hand et al. [1] broadly define DM as "the discovery of interesting, unexpected or valuable structures in large datasets." Digital information and communication technology (ICT) reduces the costs of collecting, accessing, combining and jointly analysing large amounts of data. DM is widely expected to increase the productivity of many types of research activities and to produce many valuable new insights. As we will show, conceptual, methodological and applied empirical DM research has accounted for an increasing share of total academic research output over the last two decades.

In particular, this paper is concerned with copyright arrangements that affect academic researchers' costs and benefits when accessing and jointly analysing data from databases or publications produced by others. The familiar expression "standing on the shoulders of giants" alludes to the cumulative nature of much academic research.

[1] Corresponding author. E-mail: handke@eshcc.eur.nl.
[2] This is an abbreviated and preliminary version, as of May 22, 2015.

The principle applies in cases where researchers acquire and analyse data collected and made available by others, say other researchers and publishers, private firms or public institutions.

Where unrestricted IP rights like copyright vest in 'input works' – databases or their content – data mining activities often require the express consent of rights owners to avoid the risk of litigation or other punitive measures. Subject to the transaction costs of clearing rights and any price charged by rights holders, effective copyright protection can thus increase the full economic costs of data mining. On the other hand, copyright could also encourage the supply of valuable input works and thus foster the benefits of data mining. We present empirical evidence regarding the net effect on the output of data mining-related academic research.

The evidence presented in this paper relates to a current policy debate in particular in the European Union (EU). Under current EU legislation (Directive 2001/29/EC on copyright in the information society and Directive 1996/9/EC on the protection of databases), DM requires prior authorization of rights holders even if the potential user has lawful access to the research articles and databases in question. The European Commission is currently considering copyright reforms to allow for data mining without express consent of the rights owner, so that the right to read would be the right to mine (cf. [2]). The USA have generally followed a more permissive attitude towards DM. Other countries like the United Kingdom and Japan have introduced more permissive legislation over recent years. Uncertainty complicates the matter, as neither the law nor the rights holders follow a clear line with respect to data mining. Another important variation is probably the extent to which rights holders or public authorities enforce copyright legislation in practice, and countries differ widely in this respect as well.

This paper exploits the variations in relevant copyright policy to develop evidence on the effect of different copyright regimes relevant for DM. We analyse bibliometric data to establish whether copyright policy and its enforcement affect the application of DM in academic research. We demonstrate that countries in which data mining for academic research requires the express consent of rights holders, data mining makes up a significantly smaller share of total research output.

2. The Empirical Literature

The empirical literature regarding copyright, academic research and DM in particular is limited to descriptive analyses. Regarding the supply of academic work, the paper by Tsai [3] contains recent bibliometric data on DM. Tsai uses information from the Social Science Citation Index, a section of a database called Thomson Reuter's Web of Science (WoS). He finds 1,181 academic publications between 1998 and 2009 with the topic "data mining". The vast majority of these articles, over 97%, were in English. Relevant articles are spread over a great number of academic fields.

The data presented by Tsai [3] illustrates rapid, approximately exponential growth in the number of DM-related publications and their citation counts between 1992 and 2009. For all the difficulties in predicting technological change, this makes rapid further growth likely. [3] also contains data on the share of various countries in DM related, academic publications. The U.S.A. accounts for almost 47% of all publications featuring DM in subject headers. Over 11% of the DM publications came from the U.K. and the other five largest EU economies accounted for just below 10% (Germany,

France, Italy, Spain, and the Netherlands).[3] Tsai [3] does not control for the size of countries and their domestic research output nor does he relate these findings to copyright policy. Filippov [4] contains an update of Tsai, confirming continued growth in the number of articles with "data mining" in the title up to 2013.

3. Empirical Strategy in this Paper

This paper makes greater use of the rich data available on academic research output than the preceding literature. Main factors driving the output of academic publications of any type should be (1) the means available for academic research, in particular labour and capital, and (2) the productivity of researchers, as measured by the number and quality of research output relatively to the resources used. To control for the size and productivity of academic research, we use the ratio between DM-related research output and total research output per country as the dependent variable.

The main independent variables of interest derive from a categorization of countries according to relevant copyright law and practice in each jurisdiction. Copyright protection has ambiguous effects according to economic theory. On the one hand, it should increase the number and quality of potential input works for DM applications made available. On the other hand, holding other things equal, stronger copyright protection increases the price of such works and the transaction costs compared to a situation where input works that researchers can acquire are available without an explicit contract with any rights holders.

There is often a gap between the provisions of IP law and social practice, since IP is hard to enforce. In our analysis, we thus consider relevant indicators of the rule of law within countries. The share of DM-related research output should also be affected by a number of further factors, for instance: (1) the supply of potential input works relevant to domestic academic researchers independent of copyright policy; (2) inter-country differences in academic cultures and incentive schemes that would affect the propensity to publish DM-related articles relative to other research output; (3) differences in the age structure of academic researchers that could affect adoption of DM, assuming that younger researchers may be more likely to adopt novel data collection and analysis methods; (4) targeted funding for DM; (5) learning curves as researchers improve their DM related skills with practice. However, no valid measures on these factors are available from a sufficient number of countries. We estimate a multilevel model to account for unobserved, constant country differences.

4. Data

4.1. Dependent Variable: Data Mining Research Output

One important measure of research output is the number of academic journal articles published. We collected data from Thomson Reuter's Web of Science (WoS). This is a relatively comprehensive database of academic publications, which features items from thousands of journals with a strong international reputation. We used the entire WoS

[3] The figure for the UK is the sum of England, Scotland and Wales reported separately on SSCI and in Tsai [3].

Core Collection Database including the so-called Science Citation Index Expanded, Social Science Citation Index and Art & Humanities Citation Index.

To identify the research output of interest, we extracted the number of all published research articles from a number of countries that contained the exact expression "data mining". Our panel includes the 15 largest EU member states, as well as the 27 largest other economies based on national GDP in 2013 according to the World Bank. The data covers the years 1992 to 2014. WoS includes articles published since 1975. It contains no articles on DM published before 1992. We thus have 966 country-year observations. In the data analysis, we exclude some countries for years on which no data on control variables are available.

The Boolean searches on the WoS database were defined by three simultaneous restrictions: (1) "data mining" entered in inverted commas in the field 'Topic'; (2) a country name according to the format used on WoS in the field author's 'Address', which relates to the country of residence of the first or main author; (3) a year of publication in the field 'Year Published'. Search results were further restricted by ticking the option 'Articles' in the user interface of WoS, so that results only contain academic journal articles rather than conference proceedings, book reviews and the like. For each country and year, we recorded the number of different items in the WoS database that fulfill these search criteria.

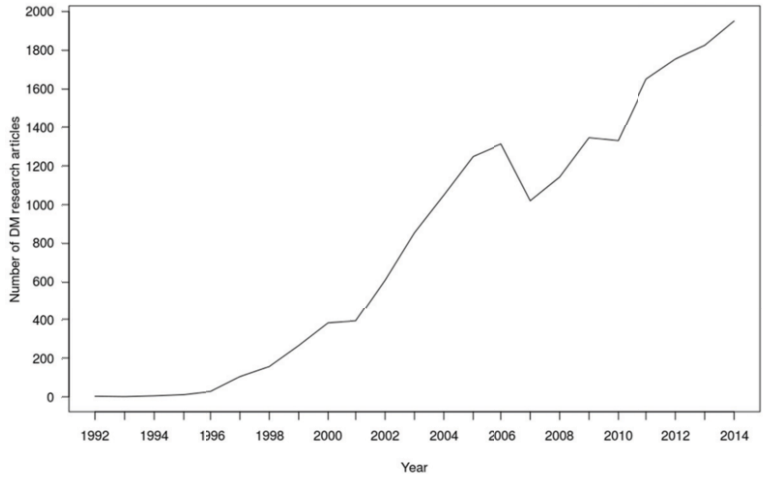

Figure 1. Absolute number of DM research articles published per year (42 countries, 1992 to 2014). *Source*: Own calculations based on search results on the WoS database.

The articles featured in the search results contain DM applications as well as related conceptual and methodological work. Among the 40 countries covered, searches on WoS brought up 18,441 DM-related articles between 1993 and 2014. We also collected data on the total number of research articles published for the same set of countries and years. Search parameters were the same as reported above, except that no 'Topic' was specified. This brought up 23,802,650 articles for the entire panel. That is for all countries and entire time period covered, 0.7‰ had DM as a topic. Starting from a low

base, there has been quite a rapid expansion of the DM share in total research output. See Figure 1 for an illustration.

In our empirical analysis, for each country and year we used the ratio of the number of DM-related research and total number of articles as the dependent variable. This variable is referred to as 'DM share' below.

4.2. Main Independent Variable: Copyright

Despite international and regional harmonization efforts, copyright protection is determined at the national level [5]. Different elements of the copyright regime may have an impact on the lawfulness of DM activities. The two features of the copyright system that bear the most on DM are the scope of rights granted on compilations of articles and other data, and the exceptions on these rights recognized in the various jurisdictions. Worldwide copyright laws can roughly be divided in two main traditions: first, the author's rights tradition, existing in countries of Continental Europe and countries that were inspired at some point in their history by the laws of one of these countries; second, the copyright tradition, reflected in the legislation following the Anglo-American legal system. Because the theoretical foundations of both regimes diverge, they are considered to follow a different approach with respect to the scope of rights and exceptions. In some countries, exceptions to copyright expressly allow DM activities to take place for research purposes without the authorization of the rights holder, while in others such activities are only lawful with a specific permission of the rights holder.

The laws of the countries examined in this paper are classified according to the possibility for academic researchers to engage in DM activities for research purposes without the need to obtain prior permission from the rights holder. Our assessment of the state of the copyright rules in each jurisdiction is based on a reading of the current legislative provisions, as well as the scholarly commentaries and the judicial interpretation, when available. In the following, we classify a number of other countries according to whether DM by academic researchers, who have lawful access to data, is either definitely 'not allowed', 'probably not allowed', 'probably allowed', or definitely 'allowed'.

Among the countries examined, sixteen belong to the European Union (EU) or the European Economic Area (EEA). Directive 2001/29/EC confers rights owners with the exclusive right to reproduce, communicate to the public and distribute their works. Directive 1996/9/EC grants protection with respect to non-original databases if they show a substantial investment in the obtaining, verification or presentation of the data. The rights granted under Directive 2001/29/EC and Directive 1996/9/EC have traditionally received a broad interpretation from the European Court of Justice (ECJ) [6, 7]. Directive 2001/29/EC contains a list of exceptions on these exclusive rights, the most relevant of which in the context of data mining allows Member States to provide for exceptions in the case of 'use for the sole purpose of illustration for teaching or scientific research. This exception is optional; Member States may decide whether to implement it or not [8]. As a result the research exception is generally vague and unevenly implemented at national level [7]. The same holds for the database right, where Member States are free to adopt an exception allowing the substantial extraction of the content of a database for research purposes, but not the re-utilization. Most EU and EEA Member states are thus 6 classified as academic DM 'not allowed' without express consent. The UK differs from the rest of the EU/EEA, as the legislator adopted

a specific copyright exception in the course of 2013, allowing DM activities for non-commercial research purposes to take place without the need to obtain prior authorization from the rights holders. Since 2014, the UK is thus classified as 'allowed'. Before, the situation in the UK was similar to that of the rest of Europe ('not allowed').

Switzerland, not being an EU/EEA country, is not bound by the European legal framework. The Swiss Copyright Act grants authors of original works a number of exclusive rights including that of reproduction, retransmission and making available. Among the several exceptions listed in the act, none would seem to cover acts of mining for purposes of research, beyond the right to make a copy for private use. Switzerland does not protect databases separately. Nevertheless, DM activities are most likely 'not allowed'. The same holds for Russia and Turkey. In these countries, DM activities are most likely 'not allowed'.

The laws of most countries with a colonial past have been influenced by those of the colonial power, most often a European country. The copyright laws of Latin American countries[4] find their root in the legislation of Continental Europe, where the rights and exceptions recognised show similar features to their European counterparts. As a result, acts of DM are most likely 'not allowed' in these countries.

While the modern Japanese Copyright Act was strongly influenced by the German Copyright Act of 1965, the recent development of the Japanese Act pursued its own course, under a greater influence from the USA. Up to 2009, DM activities were 'probably not allowed' in Japan. In that year, Japan is reported to have been the first country in the world to adopt a specific copyright exception allowing the 'analysis of in copyright works using computers in order to extract statistics and information', and come up with new ideas [9]. At the current state of our information, Japan is classified as 'allowed' since 2010.[5]

In the countries adhering to the Anglo-American copyright system, there are two different approaches with respect to exceptions on copyright: some countries recognize a 'fair dealing' exception, while others recognize a 'fair use' defense. The countries of the British Commonwealth[6] commonly recognize 'fair dealing' exceptions for different purposes, including for criticism and comment, private use and research. To fall under the fair dealing exception, the purpose of the dealing must qualify as one of the allowable purposes under the copyright act, and the dealing must be fair. The fair dealing exception generally receives a restrictive interpretation, which lets us conclude that DM activities are 'probably not allowed' in most 'fair dealing' countries. Compared to other 'fair dealing' countries, Canada has followed in recent years a more flexible approach: not only has the Supreme Court ruled twice in favour of fair dealing for research purposes, but the Copyright Act was amended in 2012, to expand the allowable fair dealing purposes. Since 2012, DM activities in Canada are 'probably allowed'. The same development can be observed for Singapore where acts of DM are today 'probably allowed'.

The 'fair dealing' exception differs from the 'fair use' defense primarily in the fact that the latter is characterized by an open-ended list of purposes for which the use of a work may be regarded as fair, marked by the words 'such as'. The fair use defense was first developed at the beginning of the 20th Century in the USA as a judicial doctrine before being codified in § 107 of the Copyright Act 1976. The assessment of whether a

[4] Argentina, Brazil, Colombia, Mexico, Venezuela.

[5] The classification of Japan is not straightforward, since this provision excludes its application to databases that are precisely made for data analysis.

[6] Australia, Canada, India, Ireland, Malaysia, Nigeria, Singapore, South Africa, United Kingdom.

particular use is fair is done by the judge according to four factors: the purpose and character of your use; the nature of the copyrighted work; the amount and substantiality of the portion taken, and the effect of the use upon the potential market. The USA is classified as 'probably allowed'.[7]

For a long time, the fair use doctrine was a unique feature of the American copyright regime. The copyright acts of countries, like Israel and the Republic of Korea, contained a list of specific exceptions, which were too narrow to cover acts of DM. However, following the conclusion of a bilateral trade agreement with the U.S.A., both Israel (2008) and the Republic of Korea (2012) introduced a fair use defense in their copyright legislation in addition to a list of specific exceptions. Since the legislative amendment introducing the fair use defense, acts of DM possibly shifted from 'probably not allowed' to 'probably allowed'.

The People's Republic of China only adopted Berne Convention compliant copyright norms in 2007, upon its accession to the TRIPS Agreement. Before that time, copyright protection on the Chinese territory was below the Berne standard, meaning that before 2007 the existence and enforcement of copyright rules was not a priority. The Chinese Copyright Act of 2007 lists the permissible exceptions, including for use of a published work for the purposes of the user's own private study, research or self-entertainment. Literal interpretation of this provision would not permit acts of data mining. However, Geller and Nimmer [10] report that the Supreme People's Court of China issued a policy document at the end of 2011, according to which in circumstances necessary to stimulate technical innovation and commercial development, an act that would neither conflict with the normal use of the work nor unreasonably prejudice the legitimate interest of the author could be deemed "fair use". This policy document was followed in a 2014 case. Since that time, acts of DM are 'probably allowed' in China while they were 'probably not allowed' between 2007 and 2012.

Taiwan Copyright Law has a long history, having first been enacted in 1928. Taiwan's modern Copyright Act was adopted in 1992 and contained a list of exceptions, none of which was broad enough to encompass DM activities. In recent years, Taiwan copyright law has been marked by American influence. In 2003 the list of exceptions was complemented by a fair use provision, which must be applied in conjunction with the specific exceptions. Since that time, acts of DM are 'probably allowed' in Taiwan, provided that they meet the four factors used to evaluate fair use in any of its many enumerated circumstances. By contrast, the law of Thailand contains a more restrictive provision according to which DM is 'probably not allowed'.

Although most Muslim countries included in the sample are members of the Berne Convention, finding specific information on the scope of the exceptions in the laws of Iran, Indonesia, Saudi Arabia and the United Arab Emirates proves difficult. They are thus excluded in the data analysis. See Table 1 for the average DM share among the four main copyright categories. By far the largest number of observations is available for the category 'not allowed', and the average DM share for this category is lower than for all other categories. The category 'allowed' only contains six observations, so that it is hardly suited for a statistical analysis. The average DM share for this category is

[7] A very recent ruling in 2014 may result in a status change of the USA to 'allowed'. In the Authors' Guild of America vs. Hathitrust case, the Court of Appeal for the Second Circuit ruled in 2014 that the digitization of books held by the Libraries for the purpose of allowing full-text searches is permissible under all four fair use factors. United States Court of Appeal for the Second Circuit, June 10, 2014 (Authors' Guild of America vs. Hathitrust), No. 12-4547-cv.

relatively low. These observations come from very recent changes, however. Chances are that the full effect of changing to 'allowed' transpire over a longer period than covered by our data.

Table 1. Categorization of countries according to their level of copyright restriction to data mining research.

Country	Not allowed	Probably not allowed	Probably allowed
Not allowed	0.54	0.54	528
Probably not allowed	0.67	0.66	162
Probably allowed	1.64	1.37	71
Allowed	0.60	0.18	6

4.3. Other Control Variables

Besides the total research output of countries, we use several control variables: (1) GDP per capita as reported by the World Bank World Development Indicators [11], with complete data for the 1992-2013 period; (2) country population size, also from official World Bank data [11], also complete from 1992 until 2013; (3) and the level rule of law as reported by the Worldwide Governance Indicators Project [12]. The level of rule of law is captured by one of the six dimensions of governance of the WGI indicators, and is defined as "the extent to which agents have confidence in and abide by the rules of society" [12], including the quality of contract enforcement and property rights. We use it as a proxy to measure the level of enforcement of property rights. Data availability for this indicator begins in 1996 and last estimates are from 2013.

5. Main Empirical Results

Due to the panel structure of our data and the low temporal variation of copyright legal arrangements within countries, we fit a multilevel linear regression model with varying intercepts by group (i.e. country), also known as a random effects model. The dependent variable is the share of articles on DM in the total number of articles published (DM share) per country and year. The main predictor is each country's copyright category, with 'not allowed' as the reference category. Table 2 presents the results of four different specifications of the model.

Model 1 only contains the main predictor. As expected, the 'allowed' category does not yield significant coefficients: it contains only six observations and we only report it for completeness.[8] There is a significant positive coefficient for the category 'probably allowed', which suggests that a more permissive copyright framework is associated with more DM research. The specification in Model 2 tests the effect of copyright categories controlling for GDP per capita, population size, and the rule of law. In Model 3 we also control for the total number of research articles published to test whether changes in DM share are confounded by changes in total research output. (We discuss Model 4 with interaction terms separately below.) The number of observations is reduced in models with control variables, since no data are available on the 'rule of law' before 1996. The control variables improve model fit considerably

[8] Furthermore, the effects of introducing permissive copyright regulations on DM share should be gradual, so that the full effect of recent changes in Japan and the UK may not have transpired.

compared to Model 1. In all specifications, we find significant positive coefficients for the category 'probably allowed' (p<.01). For the category 'probably not allowed', results are less stable. Coefficients for the category 'allowed' are generally positive but not significant with a very low number of observations. The coefficient for 'probably allowed' is consistently larger than for 'probably not allowed' in all specifications. This is in line with our ordinal categorization: there is a stronger and more reliably significant coefficient for the category that differs more from the reference category 'not allowed'. Overall, there is extensive evidence that DM share is greater in countries with more permissive DM-related copyright than in the 'not allowed' category of countries.

Table 2. Results of the multilevel regressions (varying intercept, random effects) with DM share as dependent variable.

	Model 1	**Model 2**	**Model 3**	**Model 4**
(Intercept)	0.542*** (0.075)	0.559 (1.072)	3.550*** (1.120)	4.388*** (1.009)
Copyright [Ref. *Not allowed*]				
Allowed	0.245 (0.283)	0.440 (0.355)	0.362 (0.331)	9.383 (26.470)
Probably allowed	1.407*** (0.160)	1.653*** (0.191)	1.465*** (0.188)	1.419*** (0.327)
Probably not allowed	-0.005 (0.137)	0.403*** (0.164)	0.264 (0.164)	0.503** (0.229)
GDP/capita ($1,000)		0.046*** (0.006)	0.013*** (0.007)	0.007 (0.006)
Population (log)		-0.028 (0.059)	-0.478*** (0.076)	-0.512*** (0.072)
Rule of Law		-0.761*** (0.126)	-0.701*** (0.120)	-0.639*** (0.116)
Total research output (log)			0.604*** (0.064)	0.600*** (0.065)
*Definitely allowed**Rule of Law				-6.922 (20.131)
*Probably allowed**Rule of Law				0.036 (0.266)
*Probably not allowed**Rule of Law				-0.258 (0.187)
R^2	0.144	0.233	0.351	0.333
F	42.820***	28.208***	42.948***	27.564***
N	767	564	564	564

*p<0.1; **p<0.05; ***p<0.01

In Models 3 and 4, 'total research output' has a positive and significant coefficient. Countries with a high share of data mining articles in total research output also tend to have larger total research output. There is thus no indication that DM would reduce incentives for other types of research within the same country. With the control for total research output in Model 3, the coefficient for 'probably allowed' changes little compared to Model 2.

For the entire panel, the rule of law consistently has a significant, negative coefficient. The rule of law should make the enforcement of copyright law more effective, and thus have a stronger negative effect on DM share in countries with

stronger copyright. To test for this, Model 4 includes a multiplicative interaction between copyright categories and the rule of law indicator. In this model, the coefficients of the variables that constitute the interaction (the categories of copyright regulation and rule of law in Model 4) are no longer to be interpreted as unconditional marginal effects. For instance, the coefficient of the 'probably allowed' constitutive term (1.419) represents the effect of this type of copyright regulation only when the rule of law is zero.[9] The coefficients for the multiplicative interaction terms (copyright*rule of law) are not significant, suggesting that for less restrictive countries, different levels of rule of law do not affect DM share. However, the coefficient for the constitutive term of the rule of law – that represents the effect of the rule of law for countries in the category 'not allowed' – is negative and significant. This suggests that in particular the combination of strong copyright law and strong enforcement (and/or a cultural propensity to adhere to legal norms) reduces academic researchers' data mining performance.

6. Conclusions

In most EU/EEA member states, DM-related copyright protection is comparatively strong. Our results suggest that the net effect is a weaker performance of domestic academic researchers in this increasingly important type of research. To our knowledge, this is the first time that an empirical study identified a significant negative association between copyright protection and the supply of new copyright works of any type. The academic culture and incentive scheme (based on public subsidies and the ideal of producing public goods) sits uncomfortably with academic publishing, which is operated by for-profit firms. One of the battle lines regarding DM is between for-profit academic publishers and representatives of some academics, universities and libraries. At least in Europe, publishers tend to favour restrictions on DM so that there is greater potential to sell rights to DM, whereas many representatives of the "academic community" favour a situation in which academic researchers, who have lawful access to input works, are generally allowed to conduct DM on these works. Our results suggest that in the case of academic research and DM, the adverse consequences of copyright protection on the creation of new information goods are greater than the benefits. As a rule, DM research draws heavily on input works to which others may hold copyrights. Copyright exemptions or limitations could promote this type of research, at least to enable DM of input works that have been publicly financed.

References

[1] D. J. Hand, H. Mannilla, P. Smyth, *Principles of Data Mining*, The MIT Press, Cambridge, 2001.
[2] P. Murray-Rust, *Open content mining*, Working Paper, Cambridge University, 2012. Online: http://www.dspace.cam.ac.uk/handle/1810/243749
[3] H.-H. Tsai, Global data mining: An empirical study of current trends, future forecasts and technology diffusions, *Expert Systems with Applications* **39** (2012), 8172–8181.
[4] S. Filippov, *Mapping text and data mining in academic and research communities in Europe*, The Lisbon Council, Brussels, 2014.

[9] In our panel, there are only four observations in the data with rule of law between -0.01 and 0.01 (there are no exact zero matches), which are South Africa in 1996, Argentina in 1997, and Brazil in 2010 and 2011.

[5] P. Goldstein, P. B. Hugenholtz, *International copyright*, Oxford University Press, Oxford, 2012.
[6] I. Hargreaves, L. Guibault, C. Handke, P. Valcke, B. Martens, *Report from the expert group on standardisation in the area of innovation and technological development, notably in the field of text and data mining*, Publications Office of the European Union, Luxembourg, 2014.
[7] J.-P. Triaille, S. Dusollier, S. Depreeuw, J.-B. Hubin, A. De Francquen, *Study on the application of Directive 2001/29/EC on copyright and related rights in the information society*, European Commission, Brussels, 2013.
[8] L. Guibault, Why cherry picking never leads to harmonisation: The case of the limitations on copyright under Directive 2001/29/EC, *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* **1** (2010), 55-66.
[9] J.-P. Triaille, *Study of the legal framework of text and data mining (TDM)*, European Commission, Brussels, 2014, p. 10.
[10] P. E. Geller, M. B. Nimmer, *International copyright law and practice*, Matthew Bender, Los Angeles, CHI-72, 2015.
[11] World Bank, *World Development Indicators*, Data, 2015.
[12] World Bank, *Worldwide Governance Indicators (WGI) Project*, Data, 2015.