

Text and Data Mining in the Proposed Directive: Where Do We Stand?

Kluwer Copyright Blog
March 23, 2018

Bengi Zeybek (VIR)

Please refer to this post as: Bengi Zeybek, 'Text and Data Mining in the Proposed Directive: Where Do We Stand?', Kluwer Copyright Blog, March 23 2018, <http://copyrightblog.kluweriplaw.com/2018/03/23/text-data-mining-proposed-directive-stand/>

The global research community generates over 1.5 million new scholarly articles per year. Text and data mining (TDM) enables individuals to analyse such large amounts of data, to categorise that data, and to unravel the underlying patterns in order to attain new knowledge, and to create new databases. That being said, utilisation of TDM in research and innovation is possible only if the applicable legal framework delivers precise rules that promote adoption of TDM for researchers, businesses and other beneficiaries.



The Proposal for a Directive on Copyright in the Digital Single Market, published on 14 September 2017, addresses legal uncertainty as regards TDM practices within the European Union by introducing a mandatory exception for TDM in Article 3. The proposed exception applies only to non-commercial research organisations that mine content to which they have lawful access for scientific research purposes. Article 3(2) of the Proposal further stipulates that the exception shall not be overridden by contracts.

This blog post briefly analyses the proposed exception and its legislative trail, including the amendments proposed by several committees of the European Parliament.

The Origins of the Proposal

Within the EU copyright framework, challenges to the use of TDM techniques have been identified by end users, consumers, and institutional users as twofold: first, the legal uncertainty on whether and how copyright may apply to TDM; and second, the problems with existing licensing mechanisms. Realisation of TDM involves the obtaining of the sources, transformation of the data, loading of the data, analysis of the data, and drafting of a report. These acts may be restricted by copyright, related rights, or the *sui generis* database right, principally in relation to the obtaining of the sources and their transformation (for example from PDF to HTML or HML). The proposed exception would target these acts and do so more adequately than available alternatives, such as licensing or relying on existing exceptions.

Licensing often causes barriers to successful mining, as researchers and research institutions face unreasonable terms and additional costs. These include limitations on the number of articles or prohibitions against doing so in the first place. Licensing also fails to solve the problem of legal uncertainty, as it requires dealing with a wide variety of complex contractual terms and conditions.

Beyond licensing, TDM could benefit from either the temporary reproduction exception in Article 5(1) of the InfoSoc Directive, or the scientific research exception in Article 5(3)(a) of the same Directive. However, these exceptions do not sufficiently address the copyright-related restrictions faced by scientists using TDM techniques, as some of their activities do not fall within the scope of those provisions. Because some mining involves permanent reproductions, Article 5(1) of the InfoSoc Directive does not apply. Other exceptions in Article 5 are not mandatory and do not refer to TDM, rendering their application to these activities difficult.

In contrast to countries where 'fair use' (or similar doctrines) can be invoked against copyright infringement claims for using TDM techniques (like the U.S., Israel, Republic of Korea, Singapore and Taiwan), in Europe, the use of such techniques will for the most part require the permission of the rights holder. Because of this complex, and at times vague, legal framework, European scholars occasionally have to outsource their text and data mining needs. Against this background, there appears to be a need for a clear legal framework for TDM in order to promote European scientific progress, technological innovation, and economic growth.

Pursuant to the Orientation Debate on Content in the Digital Economy of 28 November 2012, the Commission put forward the stakeholder dialogue 'Licences for Europe' in early 2013, aiming to unlock the full economic potential of TDM. The dialogue was not successful, as its limited focus on licensing led to the withdrawal of representatives of various stakeholders. This was followed by the public consultation on the review of the EU copyright rules of December 2013. The replies to the consultation included proposals to depart from licensing in TDM practices and to introduce a new mandatory exception for TDM for the purpose of commercial and non-commercial scientific research.

Finally, on 9 December 2015, the Communication from the Commission set out a long term agenda for modernisation of EU copyright rules. It asserted that the Commission would consider legislative proposals in order to "allow public interest research organisations to carry out text and data mining of content they have lawful access to, with full legal certainty, for scientific research purposes".

The Proposed Directive and proposed legislative amendments

It is possible to identify three main problems with the proposed provision on TDM. First, the beneficiaries of the exception are limited to non-commercial research organisations. Second, TDM is allowed only for scientific research purposes. Third, TDM techniques can be used only in relation to content to which there is lawful access.

Over the course of the legislative process, different Committees of the European Parliament have taken varying approaches to these problems.

The draft opinion of The Culture and Education Committee (CULT) stipulates for a fair compensation for the harm incurred by rights holders due to the use of their works, limits the scope of the exception to certain areas of scientific research and requires research organisations to delete the reproduced subject matter after the mining.

In contrast, the Committee on the Internal Market and Consumer Protection (IMCO), amends the Proposal such that the distinction between commercial and non-commercial research organisations is abolished and the scientific purpose specification is deleted.

The Industry, Research and Energy Committee (ITRE) takes an extensive approach, similar to that of the IMCO. Most notably, in its draft opinion, the beneficiaries of the limitation are identified as "public entities, private entities and individuals".

Finally, the draft report of the Committee on Legal Affairs (JURI) provides for a comprehensive set of amendments. Beneficiaries are amended to "anyone", and it abandons the purpose limited approach of the Proposal. The JURI Committee addresses the "lawful access" restriction by introducing an obligation on the part of the right holders to allow research organisations to access datasets containing works marketed by them for TDM purposes. Member States may adopt a right to request compensation for the right holders in return for the access permission, on the condition that the compensation relates to "the cost of formatting these datasets". Furthermore, it provides for establishment of storage facilities for datasets used for TDM, to be accessed only for verification of the research.

Concluding Remarks

On balance, it seems that the Proposal does not come close to offering a sound solution to the problems faced by the users of TDM techniques in the EU. The amendments proposed by JURI and IMCO on the removal of the distinction between commercial and non-commercial research organisations is promising. This is because the limitation on beneficiaries does not support commercial application of research findings and collaborative approach to research. Moreover, TDM should be allowed for any purpose. In fact, if the aim is to make the EU's single market fit for the digital age, it can be argued that the legislative framework should promote these activities by making them available not only for universities but also for research and innovation across the modern economy and in society at large. Nevertheless, and somewhat alarmingly, JURI is the only committee to address the "lawful access" restriction on the subject matter that can be mined. This aspect is strongly criticised by the European Copyright Society for good reason: the refusal of access by the right holder or the imposition of a conditional access may have detrimental effects, such as an increase in subscription fees, thereby frustrating the realisation of the full potential of TDM for research and innovation, and more broadly for the European economy.

In sum, the proposed exception for TDM should be clarified and broadened. As argued in a recent in depth expert analysis for the European Parliament, there are good arguments supporting an exception that is mandatory in nature, applies to commercial and non-commercial use, and cannot be overridden by contract or technological measures.