



LIBER Response to the STM Statement on Text and Data Mining

On 18 May 2015, STM issued a press release¹ in response to the publication of the Commission's 'Communication on Single Digital Market Strategy for Europe'. In this press release, STM addressed the Commission's suggestion that an exception for text and data mining (TDM) should be introduced as part of the Commission's ongoing work in updating European copyright frameworks.

The STM statement makes two basic points about the Commission's suggestions on a legal solution for TDM: (1) legal certainty already exists for TDM via publishers' licences and (2) creating TDM exceptions to copyright legislation would undermine the investment incentives for ensuring that high-quality content is available.

The purpose of this LIBER statement is to underline that both these premises promulgated by STM are simply wrong.

Legal certainty already exists for STM in the form of licences

Licences could never be described as simple; they are highly complex and can take months or even years to complete. They often refer to laws in other jurisdictions and in most European countries they can override the flexibilities that exceptions are intended to provide². Many licences explicitly forbid TDM associated activities such as crawling of content and the depositing of data in institutional repositories. Also, content mining is not limited only to journal content but the entire open Web. Can this content also be licenced? For big data activities such as TDM, licences simply do not scale.

Recently some publishers have begun imposing requirements on researchers who wish to perform TDM to register and to agree to click-through licences in order to have access to content to which their institution has already subscribed³. Unlike institutional licences, these click-through licences are subject to change at any time and place unfair restriction on how the results of TDM may be made available. Far from providing legal certainty, this increases the lack of clarity for researchers in terms of how and to whom they may disseminate their research. Despite objections from the research and library community, this practice has not changed.⁴

Creating TDM exceptions to copyright legislation would undermine the investment incentives to ensuring that high-quality content is available

An exception for TDM can act as an investment incentive. By implementing the exception for TMD proposed by the Hargreaves review of UK copyright frameworks, the UK government has made a clear statement that legal clarity around activities such as TDM will spur innovation and growth. In the wake of the implementation of this exception tools to support TDM and improve the quality of content have already begun to emerge. Researchers in the UK have developed their own openly available tools for conversion of

¹ http://www.stm-assoc.org/2015_05_19_News_Release_Text_Mining_in_the_EU_STM_Association_Reaction.pdf

² http://www.ifla.org/files/assets/clm/publications/licence_paralysis - protect_copyright.pdf

³ <http://libereurope.eu/blog/2014/03/28/liber-responds-to-elseviers-text-and-data-mining-policy/>

⁴ <http://libereurope.eu/blog/2014/07/01/european-research-organisations-call-on-elsevier-to-withdraw-tdm-policy/>

text files into structured standardised formats⁵. Europe should seek to build on this precedent. What Europe needs is an exception, which is mandatory and which cannot be overridden by contracts, which allow anyone with legal access to content to perform content mining. Other jurisdictions, such as South Korea, already have such legal certainty. The USA has the legal certainty of the ‘fair use’ framework. Europe must have the same certainty as a minimum, and a proposal by the Commission for an exception in the forthcoming copyright reform is the way to achieve it.

Licencing of TDM will undermine investment in TDM by diverting capital away from important TDM research and development activity into licence compliance and monitoring activities. A LIBER member university has modelled the costs of imposing the use of publisher licences on university research. If European universities were to follow the STM model of publisher licences for STM, these are the costs that would accrue to each individual research-intensive University in Europe:

- 1) additional members of staff (in double figures) to monitor compliance with publisher licences
- 2) a costs for academic time, to comply with publisher licence requirements, of up to €0.68 million per university.

A mandatory pan-European Exception for TDM, which cannot be overridden by contracts, is a much more cost-effective solution to the challenges which affect European Universities.

Conclusion

The STM Press Release on a suggested way forward for TDM is mistaken on a number of counts. It is not true that legal certainty for TDM already exists in Europe; and it is not true that the most cost effective way of allowing TDM is via the path of publisher licences. LIBER strongly rejects both propositions. In order to create legal certainty what is needed in the forthcoming European copyright reform is a mandatory exception for TDM/content mining, which cannot be overridden by contract.

European researchers should not have to wait until STM publishers assess the market as being mature enough to provide tools for TDM. Legal certainty will increase demand for, and investment in, TDM tools and will help to foster collaboration across the European research community to develop timely and practical solutions to meet the needs of researchers and accelerate the pace of knowledge discovery.

Computer analysis of content in all formats enables access to undiscovered public knowledge and provides important insights across every aspect of our economic, social and cultural life.⁶ Legal certainty regarding access to facts, data, and ideas to TDM is not only a sensible priority for the Digital Single Market, it is essential in ensuring all European citizens can benefit equally from advances in the availability of digital technology and content.

Release date: 8 June 2016

About LIBER

LIBER (Ligue des Bibliothèques Européennes de Recherche – Association of European Research Libraries) is the main network for research libraries in Europe. It represents over 400 research institutions from over 40 countries.

For press enquiries please contact the LIBER Office. Email: liber@kb.nl Phone: +31 (0)70 314 01 60

⁵ <http://contentmine.org/software>

⁶ <http://thehaquedeclaration.com/the-hague-declaration-on-knowledge-discovery-in-the-digital-age/>