

## **Frequently asked questions about Text and Data Mining for scientific purposes**

### ***Why do we need licences?***

Licences permit tailor-made uses for researchers and all potential miners. Licences will make sure that the mining is reliable, effective and efficient while not competing in any way with the original content.

### ***What can be mined?***

Scientific publications (STM)? All publications (books, newspapers, magazines)? The entire internet?

Mining the entire open internet for science may be interesting for some social science and sentiment analysis, even though there are important data protection issues also that need to be addressed. Mining the open internet in an uncontrolled manner will not deliver real and true science but risks leading to low-grade TDM results. There is a high risk of poorly executed research of low scientific merit and with unverifiable outcomes, when what is needed is high-quality TDM with guaranteed and verifiable quality content input and output.

STM publishers are willing to collaborate on very reasonable terms. Other publishers are equally open to accommodating TDM via licensing arrangements when they have requests. Priority must be given to the highest quality TDM environment - conducting more TDM and in a more meaningful way.

### ***Where can you mine? Is mining limited to the country where the researcher is based?***

Licences permit users to mine the content and then use the results of the mining anywhere in the world. Exceptions would eventually entail mutual recognition across Europe, far from the scope of what licences will offer.

### ***Who can mine?***

Researchers need legal access to the publications they wish to mine. Publishers are also working with researchers who just want to mine content for which they don't have legal access and publishers are finding the appropriate solutions. They equally are working with commercial companies to provide them with solutions.

### ***How can we remedy uncertainty for researchers wanting to do TDM?***

Researchers know that exceptions will not obviate for demand for a more successful TDM environment to surface. Scientific publishers have committed to providing TDM access to researchers who have licences for their publications (see the declaration made by STM at the occasion of Licences4Europe). For the record, we also have evidence to show that TDM research is more developed in the EU than in the US.

***Is there a market failure which would require a public intervention?***

STM publishers are including in their licences the possibility to mine and press publishers are already licencing TDM when requested by bone fide researchers. As far as the open web is concerned, this calls for the examination of data privacy as well as competition issues. With licensing mechanisms already in place, which continue to be made more effective, there is no market failure requiring a public intervention.

***Why there is a distinction between commercial and non-commercial purposes?***

In Licences4Europe, STM publishers committed to providing mining permission, at no extra cost, for not for profit researchers who have a legal access to their content. STM publishers and other publishers, when requested by commercial companies, are providing licences to exploit their content by TDM. Newspaper publishers, for example, actively license their content for re-use by media monitoring companies and other players in the commercial sector. Therefore any exception could seriously undermine legitimate revenues streams for publishers.

Recital 42 of the 2001/29 Directive addresses the issue of the funding of the research institutions: *the non-commercial nature of the activity in question should be determined by that activity as such. The organisational structure and the means of funding of the establishment concerned are not the decisive factors in this respect.*

Therefore, it may be mainly commercial intermediaries that claim the line between commercial and non-commercial uses are blurred - which they are not.

These aim for a free lunch by wanting to stand in the shoes of their non-commercial customers. But TDM is not the place to shield them like conduits from liability (as in the e-commerce directive) for they are not neutral to the content: they are actually monetising that content and relying on its quality without wanting to pay their share for it. The intermediaries are engaged in commercial uses for which they can get a licence, like any other commercial actor can.

Brussels, 15 July 2014