

# 8 Reasons Why...

## Europe needs a simple exception for Text and Data Mining

Text and data mining (TDM) has huge potential to advance science, growth and jobs in Europe. At a time that the EU is looking to boost its competitiveness internationally, it cannot afford to delay progress. TDM needs to take place under a clear legal framework that means that **all researchers, journalists, individuals and businesses** in Europe can analyse works to which they have **legal access**, using **the tool they choose**, and with the right to keep **secure copies to avoid fraud or malpractice**.

The following amendments would achieve this: JURI: 5, 7, 8, 10, 32, 33, 34. Also: 117, 119, 123, 125, 127, 134, 136, 150, 152, 159, 160, 162, 164, 532, 538, 548, 551, 552, 553, 554, 555, 559, 560, 565. IMCO Opinion: 5, 6, 8, 9, 40, 41; ITRE Opinion: 1, 3, 4

Here's why:

**1** **Because it doesn't make sense to pay twice:** access to academic publications is negotiated, with publishers offering a price calculated according to the readiness of the buyer to pay. Large companies pay more, small research institutes less. The potential value of the results of TDM is, logically, factored into this price. There is no need for a second payment.

Negotiations around the journal subscriptions are notoriously secretive, although in recent years, freedom of information requests have made it possible to show that publishers are increasing prices far above inflation. Given that prices are negotiated, publishers can already include in their offer the potential 'value' of TDM to the subscribing institution.

**2** **Because the boundary between commercial and non-commercial isn't clear:** the EU and its governments have long promoted partnerships between public research institutes and business, with a view to bringing new ideas and innovations to market. The boundary between those public-private partnerships that are and are not allowed to enjoy the exception as currently proposed will be impossible to manage in daily life.

The EU has actively sought to support university-business cooperation in recent years, seeing it as a means of accelerating knowledge transfer, creating long-term partnerships and opportunities, and driving innovation, entrepreneurship and creativity. The proposed rules around public-private partnerships will set this agenda back by creating confusion.

**3** **Because there's no limit on publishers offering their own tools for TDM:** effective tools for TDM will be essential, especially for those who cannot invest in creating their own. Competition will provide the greatest incentive to develop the best solutions. Publishers are well placed in this new market, but should not be allowed to prevent others from entering.

Publishers have worked together to produce CrossRef, which identifies works which could be mined for data. However, this does not help those whose subscriptions do not allow for TDM, and there has been limited demand. Publishers own tools can be useful, but can serve to monitor what researchers are doing, which can inhibit their work.

**4** **Because researchers gain from access, not from restrictions:** in the case of most journals, authors are not remunerated for their work. Indeed, under some open access models, they in fact pay publishers. Allowing publishers to monopolise TDM will not mean any additional money for researchers. A simple TDM exception will, however favour access and use.

Researchers benefit when their work is read and used, as this increases the impact score on their work, potentially leading to tenure or promotion. Especially for those not publishing in the highest-profile journals, TDM offers new possibilities for discovery.

**5** **Because bigger firms can pay whatever happens:** charging users to mine works they can already access legally will play into the hands of the richest actors. Making a distinction between SMEs and larger firms, while favouring smaller businesses, will create a disincentive to growth. As highlighted above, it is in negotiations over the cost of initial access that publishers should be able to earn their revenues.

When undertaking a review into the value of a TDM exception in the UK in 2013, the UK Parliament surveyed SMEs and larger pharmaceutical firms. It found that SMEs were heavily limited by the cost and uncertainty of the current situation, while the bigger firms had the budgets to buy whatever rights they needed. This served to benefit larger companies, and hurt small ones.

**6** **Because keeping copies is essential for science:** science relies on the possibility of checking out a researchers' methods and data. Without the opportunity to look at the datasets used for analysis, other researchers cannot confirm – or disprove – findings, undermining overall scientific progress. At a time that the effects of questionable science – climate change denial, the anti-vaccine movement – are imposing serious costs, we need to defend proper scientific methods.

The Wakefield scandal in the UK saw a medical researcher claim that a new vaccine (Measles, Mumps and Rubella) caused autism in children. His article is associated with a serious drop in vaccinations, and rise in disease. It was access to his data – that showed not only the use of unreasonably small samples, but also very poor practice – that made it possible to debunk his work and remove him from the medical register.

**7** **Because illicit access and sharing will still be illegal:** Original access to works must be legal. Copyright, justifiably, places the right to distribute works in the hands of the rightholder, and so also the right to charge for access. Library and users are not allowed to distribute copies further – for example by uploading them to the internet or selling them. This will not change under an exception for text and data mining.

Libraries understand the importance of high quality publishing, and indeed pay around \$30bn a year in order to access works, as underlined in Outsell's work on library market size Under the proposed exception, libraries and others will continue to have to pay for access.

**8** **Because a patchwork of different access policies just creates confusion:** the more scope there is for countries – and companies – to vary access and use rules, the further Europe will move from keeping up with the rest of the world on text and data mining. Optional exceptions, in particular when they can be overridden by contracts, merely give a formal legal status to the confusion that exists today.

As the Lisbon Council underlined in its 2016 report on Text and Data Mining, the EU has lost ten points in its share of global publications using the technique. Different rules between countries, and between publishers – the report suggests – are behind this, as jobs and revenues transition to countries where there is a simple and legally secure rule.

Europe can realise its innovative potential, without jeopardising the right of publishers to charge for original access, or their right to offer TDM tools to users. **A clear exception allowing for TDM of legally accessed content is essential.**