Researchers to take on publishers over new EU copyright laws

Éanna Kelly, Science|Business

The scene is set for intense lobbying by scientists and journal publishing companies between now and the autumn, when long-awaited reforms of EU copyright law must address the contentious issue of text and data mining

Researchers want to see provisions in a new European copyright law allowing them to use computer programmes to harvest facts and data from research papers, a practice which to date has been tightly controlled by journal publishers in Europe.

For advocates on the issue, the choice for lawmakers when they set out a legislative proposal this autumn is clear cut. "You either leave it to a few monopolies in a dysfunctional market or open it up to allow free flow of information, benefitting all and not hindering researchers," said Stephan Kuster, head of policy affairs with Science Europe, which represents the interests of national public research funding bodies.

Publishers have already lowered barriers to their academic stock, but not fast or far enough for researchers' liking. Because big academic publishers like Macmillan, Wiley and Elsevier are concerned their content might be redistributed for free, they block data-mining software programmes by default, and distribute special licence permissions to academics and university libraries.

Text mining, essentially data extraction on a large scale, goes several steps beyond keyword search tools like Google. While researchers can type 'breast cancer' into Google, and receive a list of all the documents that contain these words, the hard work of reading and rating their relevance is still ahead of them.

Text miners use computer programmes to speed-read and synthesise thousands of pages of academic literature. The result is a visual map linking unseen patterns that can lead users down new pathways of scientific discovery.

The potential of the technique for science is evident, researchers argue. One report **published by McKinsey Global Institute**says text mining could create billions of annual value to Europe's economy, if researchers were allowed to make full use of it.

Without changes to statute books, publishers across Europe can claim a right to grant or refuse the mining of their works on the basis of copyright law, EU database protection law, and provisions in licensing law. The UK is the only country in Europe that has said automated computer crawling is exempt from copyright law. Something called the 'fair use doctrine' in the US and Canada has been interpreted as covering text mining for non-commercial research. Japan and South Korea have also made exceptions.

Licence system: the start rather than the end

The problem faced by researchers is that often the content they want to mine is owned by a number of rights holders, which would involve negotiating individual licensing terms with publishers.

For scientists with a small budget and limited time, negotiations with publishers are too costly; for big research projects, which might require text mining across numerous rights holders, the process is too complicated.

Getting a licence from a publisher is a "tedious exercise" says Sergey Filippov, an associate director of the Lisbon Council, a think tank in Brussels, in a report on the topic. http://www.lisboncouncil.net/publication/publication/109-mapping-text-and-data-mining-in-academic-and-research-communities-in-europe.html "In many instances, requests are handled on a case by case basis, involving several account managers of the publishing company."

Elsevier is one publisher which says it has **updated its services** to make the process quicker and easier for scientists. Scientists can mine with Elsevier under the following conditions: they must publish the products of their text-mining work only under a licence that restricts use to non-commercial purposes, and must include links to original content. The Human Brain Project, a vast European project which is attempting to create a computer model of the brain, has a text mining licence with the publisher.

Even so, many researchers feel that computer reading should require no more permission than human reading. "The right to read is the right to mine," is a mantra they repeat.

An expert panel advising the Commission on reforming text mining law commended the publishers' licence approach **in a report last year** but said it should not be the end of the road. "Licences should be seen as a prologue to legal reform, not an end in itself," the panel said.

Susan Reilly, executive director of Liber, an association of European research libraries, agrees with the panel. "Licences will never provide legal clarity as terms vary, can change, may not be cross border and are not scalable to all of the content we could mine," she said. "The only way is via a mandatory, not to be overridden exception for text mining, one without non-commercial limitations."

The controversy surrounding the issue was brought to the fore in May 2013 when Reilly and a group of researchers and librarians walked out of EU talks on how to develop to a better regime on text mining, because, they said, only the licensing approach was being discussed.

Text mining, from the publisher's side

Text mining has become another front on the research "open access" movement and one that threatens to significantly disrupt the core business of publishers, which only have to look to the music industry and its new free-for-all file-sharing mentality to see omens.

For publishers to fully mandate text and data mining would involve significant investment in site infrastructure, plus investment in systems to hold content in a secure manner.

Some publishers have reported little interest in text mining from researchers. For John McNaught, who is deputy director of Manchester University's National Centre for Text Mining, this does not mean you can say there are pockets in Europe where text mining is not in demand.

"It's essentially a chicken and egg problem: not much text mining going on in relation to massive amounts of full texts, because of lack of access to these due to copyright restrictions, allows publishers to claim there is little interest," he said.

Sweeping legislation can be avoided, some in the industry argue. The International Association of Scientific, Technical & Medical Publishers has sought to standardise the existing practices and to develop common rules for text and data mining, as the self-regulatory way forward.

Others have made steps to accommodate researchers better. PubMed allows unrestricted text mining without permission, although this applies to academic abstracts only. Global publishers Taylor & Francis and Routledge, ring-fence a number of pure open access journals, with no subscription content. Authors also have the option of publishing their open access article under a Creative Commons Attribution license.

A number of publishers were invited to comment on this article. A MacMillan spokesperson replied to say it wasn't able to produce anyone to speak "just at this moment".

What will the Commission do?

There will be a text and mining provision in new EU copyright reform to be unveiled in the autumn, but the question of what it says remains.

For reformers, the text the European Commission used on Wednesday when it launched its digital market strategy, which is essentially a to-do list, is encouraging. A new copyright law would look at, "Harmonising exceptions for important activities such as research, education, text and data mining," it reads.

For Reilly, the chances of a good deal are "extremely realistic". However, McNaught cautioned there is still a long road to travel.