Realising the innovative potential of digital research methods: a call from the research community

Open Letter to Michiel Kolman, Senior VP Global Academic Relations, Elsevier

On behalf of research community stakeholders, we are calling on Elsevier to withdraw its current policy on text and data mining (TDM).

TDM is a digital research method which enables the analysis of vast and heterogeneous types of content. It has led to new medical and scientific discoveries and is set to be key to increasing the productivity of research and become an established element of research methodologies. Europe is falling behind in the exploitation of TDM because the lack of clarity in the current European copyright framework is disincentivising the uptake of TDM by researchers. In the UK, an exception for TDM has been introduced into legislation. What this means it that TDM will no longer be an activity that is subject to licence in the UK; any researcher will be free to mine content to which their institution has legal access to. **We see no reason that researchers across Europe and beyond should not have equal rights to mine content to which they have legal access.**

Restrictive licences provided by publishers for access to content for the purpose of TDM have the potential to further disadvantage the research community by enforcing strict parameters around how content can be mined and under what conditions the results may be made available.

We have strong reservations[1] about whether the updated Elsevier TDM policy[2], released on January 31st 2014, will meet the needs of researchers over the long term. The API that Elsevier is offering to access content can be a useful tool for developers, but limiting this service to only text will force researchers to continue to negotiate one to one licences for access to the full content (images, figures, etc.). Also, explicitly preventing direct crawling of content disregards the most common method of performing TDM. What is more, we believe that, because it places conditions on how TDM outputs may be made available, the policy will have a negative impact on the dissemination and transparency of research results. We call on Elsevier to abandon this policy and adapt their licence terms and conditions which will set a positive example that other publishers may follow in order to:

1) Support responsible research and dissemination

In order to ensure transparency and reproducibility of results, it is essential that institutions are permitted to securely store a copy of the individual datasets which researchers have used for their experimentation. Also, we do not agree that any publisher should mandate how researchers licence the output of their TDM research.

2) Protect the academic freedom of the researcher

Presenting researchers with no other option but to register their details and agree to a click-through licence, the terms of which can change at any time, in order to gain access to content for the purpose of TDM is unacceptable. It places undue liability on the researcher and undermines the role of the institution as an intermediary tasked with protecting its researchers' privacy, as well as its ability to protect its intellectual capital.

[1] http://libereurope.eu/liber-responds-to-elseviers-text-and-data-mining-policy

[2] http://www.elsevier.com/connect/elsevier-updates-text-mining-policy-to-improve-access-for-researchers

3) Prevent the creation of a monopoly

The licence for access to content should not limit the choice of text and data mining tools available to researchers, neither those developed by the researcher themselves or provided by a third party.

4) Make technological prevention measures a last resort

Rather than putting technical measures in place to prevent crawling by robots, publishers should encourage the adoption of protocols for responsible crawling of content, which would ensure that the load placed on publishers' servers by TDM activities remains negligible.

5)Deliver value for money

Research institutions are already paying a significant amount to Elsevier and other publishers for electronic access to journals. Given this significant revenue, we believe that publisher infrastructure should be robust enough to cope with the added load of TDM. Nor should licence terms prevent researchers using digital methods to exploit this content. Digital content analysis such as TDM is the most effective way of deriving value out of the large digital collections to which institutions subscribe. It can also help institutions to measure, justify and increase return on investment on its large e journal spend by surfacing valuable elements of the scholarly process such as providing links from new findings to the original publications.

Yours sincerely,

Signed: