

Background to the Letter from Publicly funded Research Sector, Researchers, Open Access Publishers and EU Technology Companies

Licences for Europe

Licences for Europe was announced in the Communication on Content in the Digital Single Market (18 December 2012) and is a joint initiative led by Commissioners Michel Barnier (Internal Market and Services), Neelie Kroes (Digital Agenda) and Androulla Vassiliou (Education, Culture, Multilingualism and Youth) to “deliver rapid progress in bringing content online through practical industry-led solutions”.

Licences for Europe aims to engage stakeholders in four areas:

1. Cross-border access and the portability of services;
2. User-generated content and licensing;
3. Audiovisual sector and cultural heritage;
4. Text and Data Mining (TDM).

Licences for Europe and the Modernisation of the EU Copyright Legislative Framework

Referring to Licences for Europe as well as the need to review European copyright law (the Information Society Directive dates from 2001), the 2012 Communication from the Commission on Content in the Digital Single Market states:

Whilst good progress has been made in delivering the copyright-related actions identified in the Digital Agenda and the Intellectual Property Strategy, there remains work to be done to ensure an effective single market in the area of copyright.

The Commission will therefore work on two parallel tracks of action. On the one hand, it will complete its on-going effort to review and to modernise the EU copyright legislative framework.

In parallel, the Commission will set out to address a number of issues on which rapid progress is necessary and possible. Thus, a structured stakeholder dialogue will be launched by the Commission, with the assigned objective of delivering by the end of 2013 practical industry-led solutions to these issues, without prejudice to further public policy action, including legislative reform, as appropriate. Under the name of "Licensing Europe", this process will seek to tap the potential and explore the possible limits of innovative licensing and technological solutions in making EU copyright law and practice fit for the digital age.

Making sense of Big Data using Text and Data Mining Tools

McKinsey Global Institute reported in 2011^[1] that effective use of ‘big data’ in the US healthcare sector could be worth more than \$300 billion a year, two-thirds of which would be in the form of a reduction in national health care expenditure of about 8%. In Europe, the same report estimated that government expenditure could be reduced by €100 billion a year. TDM has already enabled new

^[1] Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. 2011.
http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation

medical discoveries through linking existing drugs with new medical applications, and uncovering previously unsuspected linkages between proteins, genes, biological systems and diseases^[2]. A JISC study on TDM found it could reduce “human reading time” by 80% and increase efficiencies in managing both small and big data by 50%^[2].

Text and Data mining of the web is being used by many technology companies to develop new products and services. They include large multinational companies and many European-based Small and Medium Enterprises (SMEs). Many business currently use TDM to help the corporate sector profile current and future trends on the web, to provide customers (including governments) with TDM tools that save time managing their online assets and to enable scientists to mine data and text they have bought or have lawful access to on the web.

Copyright, Database Rights, Contracts and TDM

To what extent copyright law and database rights regulate the TDM process is debatable. While some techniques do not create copies of the content mined (and therefore copyright and database law would appear to not apply), others techniques may make copies and therefore may be regulated by copyright and database laws in the EU. Even if copyright law is not relevant, most purchased digital content is subject to contract law as the content is licensed. To this end, in order to compete globally Europe needs both a TDM exception and a way to ensure that contracts cannot override this access once lawful access to content has been gained.

TDM does not trade on the creative and artistic expression that copyright exists to protect, and this has led countries like Japan and the UK to introduce specific technology exceptions to allow text and data mining. In “fair use countries” like the United States, Israel and South Korea the law offers an exemption or defence for people and companies that text and data mine.

Commercial Uses for TDM – the link between academia and industry

The academic sector strongly believes that a limitation and exception for TDM is needed for both commercial and non-commercial uses. This is because of the huge interplay between the research sector and industry. The so-called “knowledge transfer” agenda is just one example of this interaction.

For example, new medical discoveries may be made in a not-for-profit research institute but it will be a medical company that commercialises these discoveries by manufacturing new drugs. Within the pharmaceutical industry, given the huge costs of developing new drugs, we see that partnerships and information sharing between different companies and organisations is increasingly the norm.

Significant numbers of technology companies ranging from multinationals through to EU start-ups are currently using TDM. Those based in the EU are doing so in an environment of legal uncertainty.

Library Spend on Content and Licensing

In 2011 \$23.8 billion was spent by libraries globally on content, and European libraries purchased an estimated \$5.4 billion worth of content. (2012 Library Market Size, Share & Forecast. Outsell.)

In the higher education sector, universities (and national consortia on their behalf) negotiate directly with publishers and aggregators for licensed rights to use publishers’ e-content and databases. This has happened since circa 1995 when electronic publishing of scholarly content became the norm. Despite universities having already paid such large sums to license this content, the premise of WG4

^[2]Text Mining and Data Analytics in Call for Evidence Responses. UK Government <http://www.ipso.gov.uk/ipreview-doc-t.pdf>

^[3]Value and Benefits of Text Mining. Dr Diane McDonald. Joint Information Systems Committee. 2012. <http://www.jisc.ac.uk/publications/reports/2012/value-and-benefits-of-text-mining.aspx>

appears to be that additional licences are required for computers to “read” and analyse this content in the context of TDM. Many in the research and higher education sector fundamentally disagree with this, claiming instead that the “right to read is the right to mine”.

EU Public Investment in R&D

According to the OECD European public investment in research totalled €91 billion¹ in 2008.

Open Access Publishing

The European Commission has made Open Access a general principle of Horizon2020 in order to boost innovation capacity². ‘Open Access’ publications make scholarly literature freely available on the internet, so that it can be read, downloaded, copied, distributed, printed, searched, text mined, or used for any other lawful purpose, without financial, legal or technical barriers, subject to proper attribution of authorship. Open Access improves the pace, efficiency and efficacy of research. It heightens the visibility of authors and the potential impact of their work. It removes geographical and structural barriers that hinder the free circulation of knowledge. It therefore contributes to increased collaboration, ultimately strengthening scientific excellence and societal progress. There are two routes to Open Access; self- archiving in repositories (green) and Open Access publications (gold), whereby a publisher is paid up-front in order to publish a book or an article. In return for this, the terms for reusing the content are intended to be unrestricted. Increasingly, research funders³ are mandating that the outputs of the research they fund be made available in Open Access.

Who are the organisations leaving WG4 on Text and Data Mining?

The organisations who have decided no longer to continue in the TDM working group 4 represent a wide spectrum of interests. They include universities and research institutions who purchase and license content, open access publishers, university researchers and research funders. Notably, the withdrawal includes that of COADEC – the only organisation that represented EU technology companies that was at the table.

A letter⁴ that was sent to the Commissioners from the same participants on 26th of February 2013 expressing their concerns about WG 4 was also signed and supported by a wide range of notable stakeholders, including Nobel prize winners for science, research institutes, open access publishers, technology companies, research funders, research libraries and learned academies.

Press Contacts:

European Technology SMEs: Sara Kelly | sara.kelly@coadec.com | +44 (0) 78 9787 1789 | <http://www.coadec.com/>

Researchers: John McNaught | John.McNaught@manchester.ac.uk | <http://www.nactem.ac.uk/>

Research Sector: Susan Reilly | susan.reilly@kb.nl | +31 615189506 | <http://www.libereurope.eu/>

Open Access Publishers: Brian Hole | brian.hole@ubiquitypress.com | +44 785 769510 | <http://www.ubiquitypress.com/>

¹OECD, Government budget appropriations or outlays for RD, data from 2008.

²http://europa.eu/rapid/press-release_IP-12-790_en.htm

³<http://www.rcuk.ac.uk/research/Pages/outputs.aspx>

⁴ <http://www.libereurope.eu/news/licences-for-europe-a-stakeholder-dialogue-text-and-data-mining-for-scientific-research-purpose>

FAQs

Would an exception in copyright law require a content provider to give free access to content?

No. Universities and other research organisations spend billions of euros annually on subscribing to scholarly publications, and an exception would not affect this at all. An exception would simply allow someone who already has access to electronic material to use their computers to read and analyse what they have bought (or have access to already on the open internet) without having to seek separate permissions, or relicense what they already have access to.

Will this put undue pressure on content providers' servers?

Open Access publishers have not reported such problems, and have not reported unmanageable server traffic. In the unlikely event proprietary publishers do experience server capacity issues, these extra costs can be passed on to the research organisations that subscribe and pay annually for the content. Another solution would be for libraries or data centres to host this material, and only give access to those who have subscribed to the content. For researchers or technology companies that wish to mine proprietorially published material that they do not have legal access to already, commercial services that aggregate and offer access to this content for text and data mining purposes would be very welcome and should be encouraged.

Aren't the TDM issues more than just legal ones?

Yes. There are many issues from needing the right skill sets and technological tools, through to needing access to the content. Removing one extra barrier, namely the need to re-license content that you already have legal access to, we believe would be immensely valuable and encourage scientific and business innovation. Other than through copyright law and legislators it is also difficult to imagine who can give permission for mining the open internet. Researchers also do not just wish to mine scholarly content in isolation e.g. social science researchers are also interesting in mining social media content such as blogs and Twitter.

Will this undermine the interests of content providers?

A legal exception for text and data mining would not allow users to share anything that was substitutable or replace the purchase of the original text or database. All it would allow would be the sharing of information, such as facts, data or a correlation derived from the original text or database. A legal exception for text mining exists already in Japan, and is being introduced in the UK. Technology companies and researchers in the US, Israel and other fair use countries also assert that fair use allows text and data mining.

Is the extraction of facts subject to copyright or database law?

No – facts are not subject to intellectual property law. Copyright law exists to protect the creative and artistic expression of an author, artist or creator, and not facts or data.

How is text and data mining related to copyright or contract law?

While the extraction of facts is not subject to copyright or database law, the copying of in-copyright materials prior to text and data mining could be if not subject to an exception. Moreover contract law, which overrides copyright law in most of the EU, controls all use of paid for content now and not copyright law. This means that without express permission to text and data mine, even if facts themselves are not subject to copyright law, their reuse may be prevented by contract.

Would an exception prevent text and data mining commercial services being developed?

No – it is not a zero sum game. Text and data mining is complex and the existence of “off the shelf” text and data mining services with aggregated content will continue to be attractive to researchers and technology companies. Scientific and industry innovation requires as many paths to content mine to be as open as possible, with as many barriers as possible removed.

Is text and data mining like a search engine “search”?

It is not to be confused with the process employed by individuals when using a search engine. In text and data mining, natural language processing technology as well as other technologies are used to find and analyse relationships in specified datasets, rather than simply rank search results by relevancy.

Why are scientists text and data mining?

Scientists are text and data mining because the volume of data available is too much for them to read, and new health or scientific discoveries can be made by using computing technology. For example it is estimated that globally more than 1.8ZB (equivalent to 1.8 quadrillion digitised books) of data is stored on servers and computers.

Why are technology companies text and data mining?

Technology companies are mining to provide a wide range of products and services for their customers. This ranges from price comparison web sites, through to companies that predict future market trends, or mine content to make economic efficiencies for companies.

Can't an exception just be for non-commercial uses?

No. Whereas health research is often undertaken by universities or charities it is commercial companies that develop the cures for diseases. Also European based technology companies who create jobs, can be encouraged to grow by providing them with a sound legal underpinning.

Won't this just help non-EU companies?

An exception in EU law will not discriminate regarding the nationality of a company, it will simply provide a “safe harbour” for companies that choose to base themselves here to take advantage of text and data mining friendly copyright laws.