

# A Copyright Exception for Text and Data Mining

A mandatory and non-overrideable copyright exception for text and data mining (TDM) for both commercial and non-commercial activity is needed. By modernising European copyright laws to support TDM, researchers will be enabled to make new discoveries and, in turn, to help drive science, competitiveness and innovation.

## 1 Why is a Copyright Exception Important?

Europe's information infrastructure cannot support world-class research unless it is underpinned by an efficient and fit-for-purpose copyright framework.

The Commission is already concerned that European TDM-based research is falling behind other areas of the world, partly due to a lack of clarity around the legality of TDM<sup>1</sup>. Research done in 2014 by the Lisbon Council indicates that this is, indeed, the case.

"U.S. nationals are responsible for almost half of all publications and patents in the text and data mining field. By contrast, European universities and scholars lag behind," the Council reported<sup>2</sup>.

*"Prolific use of TDM would add tens of billions of Euros in value to the EU's aggregate GDP."*

*-EC Expert Group TDM report*

## 2 What Type of Exception is Needed?

Any exception must be mandatory and should apply to commercial activity.

**MANDATORY** - The best way to provide legal clarity is to implement an exception to copyright law which makes clear that all Europeans have the right to mine, as long as they have legal access to the content. The only way to achieve this is via a mandatory exception at a European level, which cannot be overridden by individual contracts.

**COMMERCIAL AND NON-COMMERCIAL** - Since the objects of TDM are non-copyrightable (facts and/or data), there is no reason to limit the solution to non-commercial uses, when the very purpose of the reform is to modernise the current copyright framework.

Research does not recognise borders. Scientific research serving commercial purposes may be societally as important as strictly non-commercial research. Many research institutions receive funding from both commercial and non-commercial sources, and are mandated to engage in public-private partnership and knowledge transfer.

*"There is a continuum from public research with no commercial purposes at all to public research with clear commercial purpose; setting a boundary would be artificial and make the positioning of a given research activity very difficult."*

*-Science Europe<sup>3</sup>*

1. [http://ec.europa.eu/research/innovation-union/pdf/TDM-report\\_from\\_the\\_expert\\_group-042014.pdf](http://ec.europa.eu/research/innovation-union/pdf/TDM-report_from_the_expert_group-042014.pdf)

2. <http://www.lisboncouncil.net/publication/publication/109-mapping-text-and-data-mining-in-academic-and-research-communities-in-europe.html>

3. [http://www.scienceeurope.org/uploads/PublicDocumentsAndSpeeches/WGs\\_docs/SE\\_Briefing\\_Paper\\_textand\\_Data\\_web.pdf](http://www.scienceeurope.org/uploads/PublicDocumentsAndSpeeches/WGs_docs/SE_Briefing_Paper_textand_Data_web.pdf)

## PUBLIC-PRIVATE PARTNERSHIP CASE STUDIES

**The European Bio-Informatics Institute (EMBL-EBI)** — This institute manages some of the world's largest molecular databases. It works with industry and about a fifth of its users are involved in industrial R&D. **The institute used TDM to examine 800,000 Open Access (OA) papers for a link between genes and inflammatory bowel disease. They found over 6,000 relevant sentences in less than one day.** Many more links could have been identified, and the chances of a significant discovery increased, if the institute were able to access all relevant literature — not just OA articles.

**Agro-Know** — Foodborne disease accounts for 351,000 deaths each year<sup>4</sup>. Agro-Know aims to reduce this risk by aggregating heterogeneous information from a variety of sources (e.g. publications, government reports, datasets, social media) and making it available in user-friendly formats to decision-makers and scientists in the public and private sectors, as well as consumers who need to be informed about food-safety issues. **Their work is limited by the fact that data about food supply chains is often not available as open data, and by the lack of a clear legal framework for the sharing of information related to foodborne diseases.**

**ChemicalTagger** — This open-source tool for semantic text-mining in chemistry has already examined over 500,000 patents and identified solvents from their linguistic context with >99.5% precision. **Peter Murray-Rust, a chemist at the University of Cambridge and one of ChemicalTagger's developers, believes that chemical content mining such as this 'could create new information worth over a billion euros, and help to develop new drugs and better reproducible chemistry and educational products'.** Tools such as ChemicalTagger are, however, currently limited because they can only access open data and most publishers of chemistry papers are not universally OA.

### 3 Why Won't Licenses Work?

Licences do not provide the legal clarity needed to enable TDM at scale and across borders. Instead of enabling research, they create additional obstacles for access to research online and run contrary to the principle of a digital single market.

We must remember that the content of academic journals is only a small fraction of the content openly available and which researchers might want to mine (eg. databases, blogs, digitised cultural heritage, video clips and voice recordings). In order to make breakthroughs, discoveries and world-firsts, researchers should be able to optimise data analysis at scale, across disciplines and formats.

No licence or combinations of licences can ever facilitate the true potential scale of TDM.

### 4 What Role Does LIBER Play?

As the Association of European Research Libraries, we represent over 400 universities, national libraries and research institutions across 41 European countries. Our mission is to create an information infrastructure which enables world-class research. This can only happen when data is freely accessible, and when researchers have the tools and skills to analyse that data.

To this end, LIBER led the creation of The Hague Declaration<sup>5</sup>. It calls for change to intellectual property law and access to technology, so that facts, data and ideas are equally accessible to everyone. Over 200 organisations and nearly 500 individuals have so far signed the Declaration.

LIBER also strongly supports Open Data and initiatives such as the European Commission's current work towards the establishment of a European Open Science Cloud. We work with the Commission through our participation in several FP7 and H2020 projects, and by contributing to EU research forums, meetings and events.

4. <http://www.who.int/mediacentre/news/releases/2015/food-safety/en>

5. <http://www.thehaguedeclaration.com>