

The Text and Data Mining Exception and the Enhancement of Access to Scientific Information in Europe

Statement prepared by the ALLEA Permanent Working Group Intellectual Property Rights*

November 2017

This short statement of ALLEA invites the European Union to adopt a broad exception for text and data mining (TDM). In September 2016, the European Commission issued a proposal for a Directive on copyright in the Digital Single Market¹ (“the DSM directive”). Its Article 3 provides for a mandatory exception to copyright and database right allowing for TDM. However, the proposed exception is presently too narrowly defined, as it would not apply (i) when the results are used for publication, (ii) to independent researchers unaffiliated with research organisations, (iii) to information providers or journalists, and (iv) to private foundations or for profit entities involved in joint research projects. Such a narrow exception will not facilitate the research of various corpora that ALLEA members are undertaking in several scientific disciplines and in various circumstances.

ALLEA’s demand comes timely as the trilogue on the adoption of the DSM directive is ongoing.² ALLEA would be pleased to work with the European Commission, the members of the European Parliament and the Council to further elaborate a balanced exception for TDM in favour of the development of science and research in Europe.

¹ COM(2016) 593 final of 14.9.2016.

² The legislative process involving the Council, the European Parliament and the European Commission should be finalized at the end of 2017.

* Members of the ALLEA Permanent Working Group Intellectual Property Rights: Joseph Straus (Chair), Delegate of the Union of the German Academies of Sciences and Humanities; Alain Strowel (Lead Author), Université Catholique de Louvain, Saint-Louis University; Bernt Hugenholtz (Co-author), Royal Netherlands Academy of Sciences and Humanities; William Cornish, British Academy; Carlo D’Adda (ex officio, ALLEA Board Member), Accademia Nazionale dei Lincei, Italy; Yuriy Kapitsa, National Academy of Sciences of Ukraine; Paul O’Brien, Royal Society, United Kingdom; Are Stenvik, Norwegian Academy of Arts and Sciences; Belgium, Ünal Tekinalp - Bilim Akademisi (Science Academy), Turkey; Tomasz Twardowski, Polish Academy of Sciences; Sylvester Vizi, Hungarian Academy of Sciences.

The Text and Data Mining Exception and the Enhancement of Access to Scientific Information in Europe

Text and data mining arguably falls under EU copyright

TDM is a research method that extracts data and patterns from large datasets.³ The proposed DSM directive defines TDM as “any automated analytical technique aiming to analyse text and data in digital form in order to generate information such as patterns, trends and correlations” (Art. 2(2)).

Many copyright experts believe that text and data mining should remain outside the scope of copyright protection, since copyright protects original expression only, and ought not to extend to underlying facts and ideas (the idea/expression dichotomy). Also TDM operations usually do not affect the normal exploitation of works that copyright aims to guarantee. Moreover, most if not all scientific authors agree that their works be freely used for searching and indexing, and for developing new scientific insights in the humanities (for ex. linguistic occurrences), the social sciences (for ex. economic data), or the natural sciences.

However, the notion of reproduction, as broadly defined in Article 2 of the 2001/29 EU Directive on copyright in the

³ R. Kitchin, *The Data Revolution*, Sage, 2014, p. 104.

information society (“Infosoc directive”) to cover all “direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part”, might be interpreted as presently prohibiting acts of digitally processing data and works that are part of TDM operations. The current EU framework for copyright might thus pose an obstacle to TDM.⁴

Need for a broad exception for TDM

In the U.S., the fair use provision can function as a way to delimit the author’s exclusive rights in appropriate cases. In the *Google Books* case (*Authors Guild v. Google, Inc.*), the U.S. Court of Appeals for the 2nd Circuit, decided that the various acts committed by Google in reproducing books, including for searching and TDM, are “fair use”.⁵

⁴ In 2011, the Hargreaves Report mentioned TDM as « one current example of a new technology which copyright should not inhibit, but does » (*Digital Opportunity. A Review of Intellectual Property and Growth*, p. 48).

⁵ *Authors Guild v. Google, Inc.* 13-4829-cv available on <https://www.unitedstatescourts.org/federal/ca2/13-4829/230-o.html> (On April 18, 2016, the Supreme Court denied the petition for writ of certiorari, leaving the Second Circuit ruling in Google’s favor intact.) To make available parts of the corpus of books, Google has scanned the digital copies and established a publicly available-search function, the ngrams tool.

While no general “fair use” exemption exists under EU law, it is possible for the European legislature to introduce a specific exception for TDM. Following the example of several Member States including Germany, France and the UK, the European Commission now proposed a TDM exception in its proposal for a DSM directive.⁶

The envisaged exception, however, would arguably not allow research when it results in publications or when it is conducted in a commercial context as the exception is limited to (public interest) “research organisations” and “for the purposes of scientific research” (Art. 3). According to Article 2(1), “research organisations” means “a university, a research institute or any other organisation the primary goal of which is to conduct scientific research or to conduct scientific research and provide educational services”. This could be too narrow as research conducted by universities and research entities often involves private parties (such as private foundations) whose primary goal is not to conduct research or partners acting for profit. Recital 10 provides

that research organisations can “benefit from the exception when they engage into public-private partnerships”, but this could be too narrow as the private partners in those PPPs would not be able to access and use the data retrieved from TDM, making the functioning of those consortia quite complex. Research in corpuses conducted for the purpose of journalism by information providers and members of the press (which do not fall among the beneficiaries of the TDM exception) should also be allowed when the researcher has lawful access to the corpus. Also, independent researchers who are not affiliated with a research organisation would not enjoy the freedom of conducting TDM. For these reasons, the proposed wording on the TDM exception is too narrow.

The Commission rightly states that “there is a risk that the Union’s competitive position as a research area will suffer unless steps are taken to address the legal uncertainty for text and data mining” (recital 9), but it does not draw the right conclusion by supporting a broader TDM exception.

⁶ See Article 3 of the draft DSM directive (COM(2016) 593 final). The December 2015 Communication of the Commission recognizes the need to reduce the uncertainties by adding a TDM exception to boost research and innovation activities: Communication from the Commission, *Towards a modern, more European copyright framework*, COM(2015) 626 final, 9/12/2015, p. 7 («The need to better reflect technological advances and avoid uneven situations in the single market is also clear with text-and-data mining (TDM), through which vast amounts of digital content are read and analysed by machines in the context of science and research. The lack of a clear EU provision on TDM for scientific research purposes creates uncertainties in the research community.»).

A broad exception on TDM fits with the other ALLEA policies that aim to accelerate the move towards access to scientific works and data

In its April 2012 declaration 'Open Science for the 21st century', ALLEA stressed the need to promote (i) access to the scientific publications as soon and as freely as possible ("Open Access"), (ii) the development of open platforms allowing access to research data that are discoverable and re-usable ("Open Data"), (iii) support for interoperable e-infrastructures to manage the scale of future data flows ("Open e-Infrastructure"), (iv) the culture of open science based on online collaborations and high standards of quality and

integrity ("Open Scientific Culture").

ALLEA's statement on TDM is aligned with its position that supports the development of Open Science.

Obligation to make results of TDM research available in Open Access or as Open Data

ALLEA's general support of Open Science implies that ALLEA would like to see the results of TDM research be made available in the form of Open Access publication or as Open Data, regardless of whether the research was conducted by a publicly funded research organisation or by a private entity.

About ALLEA (All European Academies)

[ALLEA](#), the European Federation of Academies of Sciences and Humanities, was founded in 1994 and currently brings together 59 Academies in more than 40 countries from the Council of Europe region. Member Academies operate as learned societies, think tanks and research performing organisations. They are self-governing communities of leaders of scholarly enquiry across all fields of the natural sciences, the social sciences and the humanities. ALLEA therefore provides access to an unparalleled human resource of intellectual excellence, experience and expertise.

ALLEA Permanent Working Group Intellectual Property Rights

Intellectual Property Rights (IPRs), be it patents or copyrights, play an important role in all academic activities. The [ALLEA Permanent Working Group Intellectual Property Rights](#), which has been in existence since the 1990s, has prepared and issued reflections, declarations and recommendations on the most challenging topics of IPRs. At present it is developing a horizon-scanning mechanism to identify emerging issues and suggest appropriate solutions.