

**Response to Draft Legislation to modernise copyright exceptions
published for technical review
New Exception for Data Analysis for Non-commercial Research**

(See <http://www.ipo.gov.uk/types/hargreaves/hargreaves-copyright/hargreaves-copyright-techreview.htm>)

Background

This is a collaborative submission from a group of academics based in the UK with expertise in intellectual property and information technology law and related areas.

The preparation of this response has been funded by (1) British and Irish Law, Education and Technology Association (“BILETA”) <http://www.bileta.ac.uk/default.aspx> and (2) CREATe (Creativity, Regulation Enterprise and Technology) the RCUK Centre for Copyright and New Business Models in the Creative Economy <http://www.create.ac.uk/>

This response has been prepared by Christian Geib and Lilian Edwards. Christian Geib is a PhD Student at Strathclyde University. His main research interests are within data mining, (consumer) profiling, search engine optimization (SEM), internet policy issues concerning net neutrality, internet architecture and quantitative legal predictions. Lilian Edwards is Professor of E-Governance at Strathclyde University.

This response has been approved by the Executive of BILETA (the British and Irish Law, Education and Technology Association <http://www.bileta.ac.uk/default.aspx>) and CREATe and is therefore submitted on behalf of BILETA.

In addition, this response is submitted by the following individuals:

Christian Geib, University of Strathclyde
Professor Lilian Edwards, University of Strathclyde

New exception for copy made for the purposes of carrying out an electronic analysis of anything recorded in that work:

As described in 'Modernising Copyright', the Government intends to amend the Copyright, Designs and Patents Act 1988 so that it is not an infringement of copyright for a person who already has a right to access a copyright work (whether under a licence, or otherwise) to copy the work as part of a technological process of analysis and synthesis of the content of the work for the sole purpose of non-commercial research.

A licence governing access to a work will not be able to prevent or restrict use of the work in accordance with this exception, but it may impose conditions of access to the licensor's computer system or to third party systems on which the work is accessed.

Therefore the exception will not prevent a publisher from applying technological measures on networks required in order to maintain security or stability, or from licensing higher volumes of access to research outputs at an additional cost.

Exceptions to copyright for research are permitted by Article 5(3)(a) of the Copyright Directive.

Commentary on the Draft Legislation

The exception is drafted as Section 29A of the Copyright, Designs and Patents Act.

Q1: Are these provisions an effective implementation of the Government's policy?

The provisions outlined in the draft section give effect to the governmental policy of creating a more permissive copyright environment for carrying out data analysis, ie. when applying text and data mining in order to, eg, create additional value out of already published scientific articles by discovering hidden common patterns in seemingly unrelated fields.

We agree with the evidence presented at the IPO Copyright Consultation event: Proposed Exception for Text and Data Mining in London on 13 March 2012 that there is a proven need for an exception since licensing alone is not currently producing an adequate market solution – especially given the lack of legal expertise among scientific researchers, the lack of clarity as to the scope of current exceptions eg private research and study, the tendency of universities especially to wish to avoid any risk of copyright infringement suit, and the lack of resources for negotiating licensing solutions both in the research community, and for smaller publishers. There is a strong argument furthermore for facilitating non-commercial exploitation of scientific datasets and texts created with in UK HEIs given their likely at least partial original funding by public money. There is a strong connected argument we support that rightsholders should not be entitled to charge twice for effectively the one use of the copyright work (license for lawful use/access and license to datamine) given that it is the researchers who add the value to the text or dataset mined not the publishers. A licensing solution alone gives publishers total freedom to exploit notwithstanding these arguments and thus runs the risk of unduely burdening researchers attempting to produce innovative products for public good in this field.

Draft Subsection 1 would also be clearly legal under Art. 5 (3a) of Directive 2001/29/EC¹ and put into practice the conclusion by the Hargreaves Report that the UK could achieve many benefits by taking up more fully the copyright exceptions available under EU law.

Q2: Do these provisions have any effect that is not consistent with the policy aim?

The aim behind subsection (1) is to create an exception to copyright that will enable researchers to use data analysis without infringing copyright. This policy aim is clearly met with subsection (1), provided that the conditions set forth under subsection (1) are met.

Subsection (2) is consistent with the aim of “robustness” in the *Government Response*², i.e. by providing limitations to the distribution for any purpose other than non-commercial research and thereby protecting the rights of the copyright owners.

Subsection (3) ensures effectively that this exception can not be undermined by contract and thereby ensures the effectiveness of creating a copyright exception.

However, the requirement of acknowledgement unless “this would be impossible for reasons of practicality or otherwise” in subsection (1) might still be too difficult for researchers in the context of text and data mining. With the large amount of data processed it is questionable if “impossible” imposes an unreasonable burden on researchers and may simply not be scalable.

We might suggest instead “this would be impractical to achieve given reasonable effort” as more consistent with the policy aim.

Subsection (1) establishes that where a person has lawful access to a copy of a copyright work, copyright is not infringed where a copy is made for the purposes of carrying out an electronic analysis of anything recorded in that work

Q3: Is the term “lawful access” effective?

Particularly in the context of text and data mining access to copyright work is the crucial factor. The term “lawful access” in Subsection (1) therefore seems effective and less unduly restrictive than the term “lawful use” as used in Directive 2001/29/EC³.

It seems essential that those who are internally authorised to access the dataset or text but are not themselves necessarily named licensees should have the right to benefit from the new exception. For example, PhDs working in a team under a senior researcher at a university which has acquired a license for said named senior researcher and a certain number of other accesses to a certain dataset.

¹ Hargreaves, Prof Ian (May 2011). *Digital Opportunity – A Review of Intellectual Property and Growth*, UK Intellectual Property Office, p. 6.

² The Intellectual Property Office, *Modernising Copyright: A modern, robust and flexible framework-Government response to consultation on copyright exceptions and clarifying copyright law*, p. 2, Newport: The Intellectual Property Office.

³ Art. 5 (1b) of Directive 2001/29/EC.

Q4: Does the term “electronic analysis” capture the range of analytical techniques used in scientific research?

Although “electronic analysis” of Subsection (1) might vaguely encompass the range of conceivable analytical technologies, it would be preferable to have more specific terms or some examples used in order to describe the technologies employed for data analysis. An *eiusdem generis* rule can then be applied.

Clearly the “future readiness”, as expressed in the *Government Response*⁴, was the legitimate rationale for the use of a broad term such as “electronic analysis”. However, with the widespread uncertainty among the research community at present (and to be expected after the introduction of new copyright exceptions) it would be of great help to the researchers, who may not have the benefit of legal advice, to use some common examples of “electronic analysis” and possibly provide some definitions (such as text or data mining or at least structured and unstructured data).

This is especially relevant where the researcher might be in doubt whether “electronic analysis” of structured data (data mining)⁵ in fact is treated the same as “unstructured data”/“free text” (text mining)⁶. Also, for the sake of clarification it should be made clear here that the term “electronic analysis” indeed does include all the major activities such as information retrieval, natural language processing, information extraction etc⁷.

Subsection (2) establishes that a copy made for these purposes may not be distributed or used for any purpose other than non-commercial research. Any such action infringes copyright. This protects the owners of rights in the copied works.

Q5. Is this wording effective?

It is, but we would ideally like to see future discussion about the exception being extended to cover *commercial* data or text mining uses, albeit with reasonable pre conditions, for the reasons cogently explained by the OKF at <http://science.okfn.org/2012/03/21/response-to-ipo-consultation-on-text-mining-copyright-exception/>. “Researchers in both academia and industry are often reliant on the same information e.g. libraries of chemical structures”. Therefore, to impede commercial, but not non commercial, access to mine text and data would result in duplicated time, effort and expense to obtain the same information, reducing public utility. Furthermore in modern universities, the lines between commercial and non commercial research are often extremely blurred and the distinction increasingly artificial. We recognise there may be difficulties in bringing commercial research under an EU approved exception however given the limits of art 5(3) 2001/ 29/EC. This may be a

⁴ The Intellectual Property Office, *Modernising Copyright: A modern, robust and flexible framework-Government response to consultation on copyright exceptions and clarifying copyright law*, p. 14, Newport: The Intellectual Property Office.

⁵ Sophia Ananiadou, Paul Thompson, James Thomas, Tingting Mu, Sandy Oliver, Mark Rickinson, Yutaka Sasaki, Davy Weissenbacher, John McNaught: *Supporting the education evidence portal via text mining*, in: *Phil. Trans. R. Soc. A* (2010) 368, p. 3831.

⁶ Kristin Sahini, “Mining Biomedical Literature: Using computers to extract knowledge nuggets”, in: *Biomedical Computation Review*, Summer 2008, p. 18.

⁷ *Ibid.*

matter that has to be reconsidered at supranational level. We would suggest such an extension should only occur in circumstances where there is no prejudice to the rightsholder in the reasonably foreseeable exploitation of their own market. For example a publisher of scientific biochemistry journals is rarely also involved in the market for data mining to produce new drugs and as we note above bring nothing new to the table to justify charging twice for access to read and then permission to data mine. Again as the OKF say, it seems reasonable that : “At a minimum, any subscriber to closed access journals regardless of their commercial status should be able to mine the information for which they have paid subscription fees”.

Subsection (3) protects the exception to copyright from override by contract, while not restricting other contractual terms, as outlined above.

Q6. Is this wording effective?

The wording of Subsection (3) appears effective.

End of submission