Title: Exception for copying of works for use by text and data analytics **IA No:** BIS0312 **Lead department or agency:** IPO **Other departments or agencies:**	**Impact Assessment (IA)** **Date:** 13/12/2012* **Stage:** Final **Source of intervention:** Domestic **Type of measure:** Secondary legislation **Contact for enquiries:** joanna.huddleston@ipo.gov.uk

Summary: Intervention and Options

RPC Opinion: GREEN

Cost of Preferred (or more likely) Option				
Total Net Present Value	**Business Net Present Value**	**Net cost to business per year** (EANCB on 2009 prices)	**In scope of One-In, One-Out?**	**Measure qualifies as**
£0m	£0m	£0m	Yes	Zero Net Cost

What is the problem under consideration? Why is government intervention necessary?

Text and data and data analytics methods extract data from existing electronic information, to establish new facts and relationships, building new scientific findings from prior research. These new methods involve copying of prior works as part of the process to extract data. These methods have emerged since earlier revisions to the copyright framework. Copyright law currently requires a licence for copying substantial parts of works of part, whole or collections of work, which could inhibits use of text and data analytics. Copyright is not intended to prevent use of facts for research, and this exception is intended to remove the block on reuse of materials for research using these tools.

What are the policy objectives and the intended effects?

The aim is to permit copying where it is for the purpose of applying analytic technologies, in cases where access to articles and / or data has already been gained (eg by subscription), and the works have been provided to the user, separably from the publisher or provider's system. Therefore it would apply to those circumstances where the added value of the technology is supplied by the user, rather than by an ongoing service by the publisher. This limited exception would be protected from override by contract. The exception would not permit users uncontrolled copying of the content on publishers' systems .

What policy options have been considered, including any alternatives to regulation? Please justify preferred option (further details in Evidence Base)

- Option 0 - do nothing
- Option 1 - create a limited and targeted copyright exception in UK law to allow text and data mining for non-commercial research / amend the current research exception, in order that copying and/or any related use of works for the use of text and data analytics will not require specific authorisation, where the user has permission to use the works on their own systems.

Option 1 is our chosen option.

The Government will also explore what further support or measures might be provided to enable publishers and other stakeholders to improve support for and deployment of these technologies, including by clarifying what can already be done under the existing law, and by supporting development of licensing process where licensing is the best solution. However, these measures would be in addition to this limited exception.

Will the policy be reviewed? It will/will not be reviewed. **If applicable, set review date:** Month/Year

Does implementation go beyond minimum EU requirements?			No		
Are any of these organisations in scope? If Micros not exempted set out reason in Evidence Base.	**Micro** Yes	**< 20** Yes	**Small** Yes	**Medium** Yes	**Large** Yes
What is the CO_2 equivalent change in greenhouse gas emissions? (Million tonnes CO_2 equivalent)			**Traded:** n/a	**Non-traded:** n/a	

I have read the Impact Assessment and I am satisfied that (a) it represents a fair and reasonable view of the expected costs, benefits and impact of the policy, and (b) that the benefits justify the costs.

Signed by the responsible Minister: Younger of Leckie Date: 23 March 14

Summary: Analysis & Evidence

Policy Option 1

Description:

FULL ECONOMIC ASSESSMENT

Price Base Year 2012	PV Base Year 2012	Time Period Years 10	Net Benefit (Present Value (PV)) (£m)		
			Low: 0	**High:** 0	**Best Estimate:** 0

COSTS (£m)	Total Transition (Constant Price)	Years	Average Annual (excl. Transition) (Constant Price)	Total Cost (Present Value)
Low	0	1	0	**0**
High	0		0	**0**
Best Estimate	0		0	**0**

Description and scale of key monetised costs by 'main affected groups'

It has not been possible to monetise the costs due to lack of available data and limited consultation responses containing evidence.

Other key non-monetised costs by 'main affected groups'

To electronic publishers there may be: loss of revenue from existing and future licensing for non-commercial research; increased costs through the provision of server capacity to enable access to material for text and data mining; legal costs; reduced incentive to develop 'in house' analytics products and services or; loss of control over the market for providing analytic technology services.However costs for any individual publisher or rights owner will be limited because this exception covers only non-commercial research.

BENEFITS (£m)	Total Transition (Constant Price)	Years	Average Annual (excl. Transition) (Constant Price)	Total Benefit (Present Value)
Low	0	1	0	**0**
High	0		0	**0**
Best Estimate	0		0	**0**

Description and scale of key monetised benefits by 'main affected groups'

It has not been possible to monetise the benefits due to a lack of available data. However we have used the JISC report "Value and Benefits of Text Mining to UK Further and Higher Education" to demonstrate the benefits using illustrative case studies.

Other key non-monetised benefits by 'main affected groups'

To non-commercial researchers: use of analytic technologies on large and multiple data collections will make research easier, so more and higher quality research takes place; reduction in administration costs through the simplification of licensing agreements across multiple publishers; licence fees specifically for mining reduced or removed. Overall public benefits from more and higher quality research, including development of new technologies and gains to public authorities (health, education, defence) from additional research results at lower cost.

Key assumptions/sensitivities/risks **Discount rate (%)** 3.5

Assumption: achieving the optimum benefits will depend on framing and administering an exception so that it does not lead to a substantially higher rate of copyright infringement of electronic publications/archives.

Cost and benefits of the exception could well grow as further technological applications emerge.

BUSINESS ASSESSMENT (Option 1)

Direct impact on business (Equivalent Annual) £m:			In scope of OIOO?	Measure qualifies as
Costs: 0	**Benefits:** 0	**Net:** 0	Yes	Zero net cost

Evidence Base

Background

Copying is an act restricted by copyright. However, current UK and EU copyright laws include a number of exceptions allowing copying (without the need for permission or a licence) in specific circumstances, including for non-commercial research under certain conditions. Copyright is intended to offer incentives to creators through the offer of a temporary monopoly; it is not intended to act as a barrier to the development and use of new technologies which do not unduly undermine the offer of those incentives.

Copyright law was not established to regulate the deployment of analytic technologies, but because these technologies involve acts of copying, copyright does take effect in this area, even where the products of the technologies do not contain protectable expression or affect the markets for the primary works. Text and data analytics extract data from existing works and then analyse and synthesise it to develop new research and findings. The process involves copying and otherwise synthesising part or all of the primary works. Copying a substantial part of a work is an act restricted by copyright, and so requires permission from the rights holder. Therefore there is potential for copyright to establish a barrier to deployment of these technologies even where legitimate access has been gained to the primary works.

Text and data analytics were not developed when the current copyright exceptions regime was developed. Copyright has not been designed to create, in effect, an additional right for rights-holders to restrict future deployment of analytics. Therefore it is necessary to assess whether it imposes undesirable barriers to the realisation of the overall public benefits of these new technologies.

Problem under consideration

As a result, these restrictions appear to be inhibiting wider usage of text and data mining technologies and are hampering new research in the UK Further and Higher Education. As the Joint Information Systems Committee (JISC) highlight in their report looking at the value and benefits of text mining[1] "without wider usage, the potential for text mining to generate gains for the economy and society cannot be exploited and the UK economy will be less able to take advantage of its strong public research base. This carries dangers of being left behind as other competitor countries adopt a more liberal approach that encourages text mining usage."

The Government believes that it is important to protect publishers / suppliers from the risk of copyright infringement, but that where risks can be minimised and publishers are primarily providing digital copies of works, it should not be necessary that users should require an additional permission for analytic uses of those works for non-commercial research, provided that such use does not significantly undermine the security and stability of publishers / providers' networks.

There is evidence from some stakeholders that the transaction process of negotiating contracts/licences with copyright holders (publishers) specifically for

[1] JISC (2012) The Value and Benefit of Text Mining to UK Further and Higher Education. Digital Infrastructure. Available at http://bit.ly/jisc-textm

using analytic technologies and in addition to normal access subscriptions, on different journals across multiple publishers, can have a non-negligible cost implication, and also result in many failures to gain permission, in effect preventing some research from taking place.

JISC explain "The McKinsey Global Institute's report on "Big Data" [2]found that the generation of information and data has become a "torrent", pouring into all the sectors of the global economy and is predicted to increase at a rate of 40% annually. Exploitation of this cast data and information resource can generate significant economic benefits including enhancements in productivity and competitiveness, as well as generating additional value for consumers. However the full economic and societal potential afforded by this cast sea of information and data is not yet being realised within the UK. Realising the potential requires text and data analytical capability, access to the information and data sources, and involves a range of computerised analytical processes, not all of which are readily permitted within the current UK legislative environment."[3] In academic research 1.5million new peer reviewed articles are published globally each year. No research can keep pace with this without digital technologies.

JISC believe that "the barriers limiting uptake of text mining appeared sufficiently significant to restrict seriously current and future text mining in UKFHE, irrespective of the degree of potential economic and innovation gains for society,"

Rationale for intervention

As above, copyright has an essential role in offering incentives to publishing. However, copyright should not present unnecessary obstacles to the deployment of analytic technology for non-commercial research where that is not necessary to its core function in offering those incentives.

The aim of this measure is to ensure that copyright does not act as an undesirable barrier to deployment of these technologies for non-commercial research, where legitimate access is gained to works and as long as the activity does not undermine the primary market for access to works, or the security and stability of providers' networks. The Government believes that it is important to protect publishers / suppliers from risk of copyright infringement, but that where risks are minimised and publishers are primarily providing digital copies of works, it is not necessary that users should require an additional permission for analytic uses they make of those works.

Within these limits, the policy aims to support deployment where the added value of the technology is supplied primarily by the user or an authorised agent, rather than by an ongoing service by the publisher. It will not prevent the publisher / provider from applying technical protection measures on networks used in order to maintain security or stability. It will not prevent the publisher / provider from offering licensed services, or from licensing all commercial activities in this area.

---

[2] http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation

[3] JISC (2012) The Value and Benefit of Text Mining to UK Further and Higher Education. Digital Infrastructure. Available at http://bit.ly/jisc-textm

The Government notes that publishers have been working to streamline licensing for text and data mining. This activity will continue to be valuable, in particular for the licensing of commercial uses, and of uses beyond the scope of this exception.

Policy objective

The opportunities which text and data analytics offer to non-commercial research are fully realised, while preserving the core aim of copyright in offering incentives to publishers' and protecting them from risk of copyright infringement.

Options Considered

The options we have considered are:

Option 0 - no change to copyright law.

Option 1 - Create a copyright exception to cover text and data analytics for non-commercial research within certain restricted limits

To amend the Copyright Designs and Patents Act to provide that it is not an infringement of copyright for a person with a right of access to a work (whether under a licence or otherwise and regardless of the conditions of access) to copy the work as part of a technological process of analysis and synthesis of the content of the work (whether with or without other works) for the purpose of non-commercial research.

This will not prevent the publisher / provider from applying technical protection measures to control or prevent activity on networks used in order to maintain security or stability.

The Government will also explore what further measures may be useful to assist publishers and users in supporting deployment of these technologies, including by clarifying what can already be done under the existing law, and by supporting development of licensing procedures for added value services. However, these measures would be in addition to this limited exception.

Option 0: Do nothing

Copying whole copyright works for use by text and data mining would continue to require specific permission from copyright holders for each work.

- Costs and Benefits of Option 0 (including administrative burden);

No additional costs or benefits would accrue from this option. Doing nothing would result in no reductions in transaction costs for research using text and data analytics and research opportunities would continue to be hampered by the hold-up effects of requiring permissions to mine each separate work. The Government would not be able to meet its policy objectives, and the problems may worsen over time.

- Risks and assumptions;

We have assumed that no comprehensive non-statutory mechanism develops or develops quickly enough to establish permissions easily across different copyright owners to achieve the desired benefits to non-commercial research, and the need for multiple additional licences remains.

- Wider impacts

There would be none of the improvements sought to the non-commercial research environment, and so none of the subsequent public benefits of more and higher quality research.

Option 1 - Create a copyright exception to cover text and data analytics for non-commercial research within certain restricted limits, in order that copying and/or any related use of works for the use of text and data analytics will not require specific additional authorisation. Any contractual provision contrary to this shall be null and void.

A licence governing access to the 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 system or to third party systems on which the work is accessed. Therefore the measure will not prevent the publisher / provider from applying technical protection measures on networks used in order to maintain security or stability.

The exception will not prevent rights-holders /suppliers from offering licensed services such as the provision of analytic tools and services.

This measure does not provide a "right to data mine" works to which the researcher does not already have a right of access.

Any material resulting from the analysis and synthesis which does not contain substantial protectable expression taken from the copyright work, and which is made available to the public will not infringe. This measure would not affect rights-holders' rights to restrict resulting works which do contain such protectable expression.

- **Costs and Benefits of Option 1 (including administrative burden);**

Costs

Electronic publishers would forego opportunities specifically to licence text and data mining for non-commercial research. However, they would retain the right to offer "in-house" text and data mining services, and additional value-added services, and would continue to control commercial uses of text and data mining. No evidence was provided during the consultation on the current level of text and data mining licence fees or the exact number of licences granted per year. There is some variation in reports of the proportion of text mining requests that are currently granted.

This exception is framed to operate in those circumstances where publishers are not obliged to undertake very significant specific additional investment, or to accept significant additional risk so we expect transitional costs to be minimal.

Publishers may find it necessary to develop technical support and administration to allow users who have legitimate access to works to use these technologies in relation to those works. If there is a significant increase in demand, publishers may need to act to manage this traffic, including demand placed on servers by users with lawful access downloading material for text and data mining. However no evidence was provided to allow us to calculate the size of these costs. Some publishers e.g. the open access publishers PLoS already actively permit text and data mining of their entire corpuses. PLoS have reported that they have not experienced any issues with server load caused by text mining[4] [5].

Some respondents raised concerns about additional legal costs. No evidence was provided to allow us to calculate the size of these potential costs.

The exception will not prevent the publisher / provider from applying technical protection measures on networks used in order to maintain security or stability. If there is a significant increase in demand for text and data mining for non-commercial research, there are likely to be some costs to ensuring security and stability.

A number of consultation responses raised concerns that the exception would result in a loss of incentive to develop new technologies in the text mining area, however as the exception only applies to non-commercial research we do not expect these costs to be realised as there will still be a strong incentive to innovate for commercial research.

Benefits:

The text and data mining exception for non-commercial research will benefit all users of information and data and will result in cost savings and productivity gains. A research report by JISC into the Value and Benefits of Text Mining to UK Further and Higher Education[6] found that “the existing legal restrictions on text and data mining meant that it proved very difficult within the course of this study to source sufficiently robust data to enable quantification of the benefits arising”, However they have provided case studies and given an indication of the likely scale and significance of the value that can be generated through text and data mining. IPO appreciate that text mining is current possible for some research e.g. using open access material but that the transaction costs and ad hoc approach means that the barriers to text mining can be high and, significantly, not scalable to larger volume use. These barriers will be significantly reduced by the exception and this will allow both development of text and data mining techniques and greater utilisation of these technologies. .

[4] JISC (2012) The Value and Benefit of Text Mining to UK Further and Higher Education. Digital Infrastructure. Available at http://bit.ly/jisc-textm

[5] http://libereurope.eu/blog/the-perfect-swell-at-the-british-library

[6] JISC (2012) The Value and Benefit of Text Mining to UK Further and Higher Education. Digital Infrastructure. Available at http://bit.ly/jisc-textm

Reduction in costs for licences: There would be a reduction in costs for researchers and institutions from licences no longer required for non-commercial research and from the transaction costs currently required to establish specific contractual permissions for text and data mining across quantities of materials with different owners, for such research. The improved clarity and the removal of these costs should result in increases in a) lower costs for ongoing research, and b) increased volumes of research.

The Publishing Research Consortium commissioned a survey about content mining, for which it polled 190 journal publishers with more than half of publishers responding that they decide on a case by case basis whether to allow access. As the current requests for text and data mining licences are ad hoc and are dealt with on a case by case basis it is not possible to calculate an average licence fee that would be saved as a result of the exception. However as this saving would also be a loss of income for the publishers, this benefit would have a net zero effect.

Administrative costs

The JISC report identified transaction costs as, with entry costs, the most significant costs in text mining. Publishers and researchers have noted that individual negotiations are often positive and successful, but researchers have objected that specific additional negotiations for different uses, in addition to those for access, do not represent a scalable approach, and these technologies work most effectively at scale.

There is some variation in reports of the proportion of text mining requests that are currently granted. However, the process clearly occupies considerable resources on all sides both in terms of time and money.

The Wellcome Trust in their response to the JISC report highlighted the high transaction costs using a case study looking at text mining for research on malaria. They found that in the free to access, UK PubMedCentral repository there are 2,930 full-text articles published since 2000, which have the word ‘malaria’ in the title. Of those 62% are open access and this suitable for text mining without having to seek permission. However the remaining 1,112 articles are not Open Access, and thus permission from the rights-holder to text mine this content must be sought. The 1,112 articles were published in 187 different journals, published by 75 publishers. As publisher details are not held in the UKPMC database, the permission-seeking researcher will need to make contact with every journal. Wellcome used a conservative estimate of one hour research per journal title to find contact details, make contact and seek permission etc. (The British Library estimated that it took 302 hours just to identify 299 rights holders) Therefore this task would take 187 hours. Assuming that the researcher was newly qualified, earning around £30,000pa, this single exercise would incur a cost of £3,999. They then went on to point out that in reality a researcher would not limit their text mining analysis to articles which contained a relevant keyword in the title. They expand the case study to find any full text research article in UKPMC which mentions malaria post 2000 and found the cohort increases to 13,757 of which 49% were not Open Access. In this example the transaction case based on the assumptions above would be £18,630. Whilst this is

only a case study and cannot be used to extrapolate to provide a robust estimate of the potential transaction cost savings, it does provide a good example of just how high the transaction costs are and the potential scale of the benefits. 89% of the life sciences literature indexed by Europe PubMedCentral published in 2011 was not eligible for text mining without getting special permission from the publisher.

The National Centre for Text Mining identified an additional cost in the need to explain the purposes of requests.

Greater outputs from investment: Research funding in UK Higher Education Institutions (2009/10) was £6.3 billion. In 2008, UK HE alone (ie not including FE) directly employed more than 372,000 people. The Universities UK report, "The impact of universities on the UK economy",[7] states that they contributed £59 billion to the UK economy in 2009 and generated over £33bn of GDP. A measure which can derive more value from the outputs of this investment has considerable potential.

The Wellcome Trust points out that a study by Cambridge Economic Policy Associates (CEPA)[8] and commissioned by the UK Research Information Network identified costs in the global 'scholarly communications system' at £175bn: £115.8bn for research production, £6.4bn for publication and distribution, £2.1bn in providing access to the articles (library costs) and £50.3bn in user search and print cost and user reading of the articles. The report concluded that the most resource-intensive elements of the system were concentrated, not in the publication stages, but in the original research production, in peer review and in the costs incurred in access and in the consumption and use of the research articles.

The JISC report states: "As the CEPA analysis has shown, the publisher investment in the scholarly communications process is an important but relatively small overall part of the investment required. The majority of the investment and risk burden is borne by the public purse, philanthropic funders and researchers and users themselves. Hence, the broader interests of equity may support the case for an exception to enable text mining so that society can maximise the potential returns from an asset in which society has made the lion's share of investment and taken the vast majority of the risk."

To provide an illustration of the potential efficiency savings of text mining the JISC report uses Tenopir and King's[9] longitudinal study of US academics' reading practices to show that on average it takes 31 minutes to read a paper in order to summarise its content. They estimate that it takes only 5 minutes to read an automated summary resulting in a 26 minute time saving. They assume that the average academic salary is £48,000 and they work 1,650 hours per annum. This therefore provides a £12.61 cost saving per summary read. Tenopir et al[10] estimate that the average academic in the sciences reads on average 204 unique articles per year. Assuming the same reading behaviour across all disciplines, automated summarising through text mining could therefore lead to costs saving per academic

[7] http://www.universitiesuk.ac.uk/Publications/Documents/EconomicImpact4Full.pdf

[8] http://www.cepa.co.uk/documents/Dynamicsoftransitionforscreen_004.pdf

[9] C. Tenopir, D.W King, S. Edwards and L.Wu, "Electronic journals and changes in scholarly article seeking and reading patterns" Aslib Proceedings, vol 6`. No.1 pp5-32, 2009.

[10] ibid

per year of £2,572. With over 144,000 academic staff in UK Higher Education this would imply possible research efficiency savings of over £370m per annum.

Whilst these are benefits of text and data mining rather than of the text and data mining exception, currently a large proportion of journals are restricted and text mining is not available. Therefore the exception would mean that information that is not currently easily text mined will now become available and the efficiency savings above can be realised.

Haeussler and Bergman, academics at the University of Manchester, have started a project to log the problems with gaining permission to mine their content[11].They found that although freely available patents and article abstracts are open for text-mining, material behind pay walls is not, even when institutions have paid for a site licence. Of the 2.4million abstracts listed by PubMedCentral only 17% are licenced for text mining. Therefore if we assume that the remaining 83% journals become available as a result of the exception then we could expect the efficiency savings to be around (£370m x 0.83) £300 million.

More efficient use of existing services: Several respondents pointed to the advantages of more researchers being able to use existing and central services. Cambridge University Library: "The University of Cambridge, for example, has a very active community of researchers working on the development and application of text and data mining tools, such as the Computer Laboratory's Natural Language and Information Processing Research Group and the Unilever Centre for Molecular Science Informatics. However, contractual restrictions imposed by those who own the rights to these materials mean that our users are barred from taking advantage of the technology's potential, either because the licences explicitly forbid this, or because their lack of clarity on the subject makes researchers and librarians conclude that the risk of infringement is too great."

There are potentially significant broader public benefits (UK and worldwide) from scientific/medical advances, deriving additional value from existing research, some of which will have been publicly funded. To non-commercial scientific/medical researchers and organisations: greatly expanded capacity to undertake non-commercial scientific and medical research. To give a few examples, data mining techniques have been used to identify previously unknown connections between illnesses[12], to identify genes linked to osteoporosis[13], and to understand sociological phenomena such as when new words are most likely to arise[14].

Development of services: There may also be subsequent benefits to innovative companies and to the economy from opening up the market for development and deployment of analytic technologies. This could have overall benefits to the UK economy. A number of respondents quoted Victor Henning of Mendeley: "Due to the uncertainties with UK legislation, we are currently exploring opportunities for setting up text mining projects through our US subsidiary, on US-based cloud computing

[11] http://www,nature.com/news/trouble-at-the-textmine-1.10184
[12] http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1002141
[13] http://www.sciencedirect.com/science/article/pii/S8756328206009471
[14] http://www.nature.com/srep/2012/120315/srep00313/full/srep00313.html#f2

infrastructure. However, since most of our team and infrastructure is based in the UK, this introduces delays and overhead cost, and will potentially lead to Mendeley creating future jobs in the US rather than the UK."

An objective of policy is to ensure that the exception will apply in cases where access to articles and/or data has already been gained (eg by subscription). If users are able to create additional value through the exception this may increase demand for articles and/or data and so there will be additional benefits.

Total Productivity Gains as a result of the exception

As it is not possible to robustly quantify the benefits above, we have provided illustrations that show the impact the exception could have on non-commercial research. If all the benefits above occur then we believe there will be significant productivity gains. JISC estimate that "there are currently over 144,000 full time equivalent academic professionals (teaching and research) working in UK higher education. Using data from the Higher Education Statistics Agency for UK academic salaries we can calculate that the median salary for a UK academic falls into a band of between £42,000 and £55,000, which translates to between £26 and £33 per hour. If text mining exception enabled just a 2% increase in productivity – corresponding to only 45 minutes per academic per week this would imply over 4.7million working hours and additional productivity worth between £123.5m and £156.8m in working time per year.

We believe that this 2% productivity gain is a very conservative estimate because whilst there is limited robust empirical evidence on the contribution of text mining technologies to productivity growth, extensive research reveals a link between firm-level use of ICT and productivity. For example, a recent study finds that a doubling of the IT stock is associated with an increase in productivity of 5.35% for US multinationals and 4.5% for a non-US multinationals.15 Similarly, using a sample of EU countries it has been estimated that an increase in broadband use from 40% to 100% of a firm's workforce nearly doubles its labour productivity.[16] These effects are larges in knowledge based services (which include research).

ICT investments also drive productivity in the public sector. There is significant interest in improving usage of public sector data to inform policy making (e.g. Stephan Shakespeare's review of Public Sector Information[17] and the Information Economy Strategy[18]). Improvements in text and data mining technologies facilitated by the proposed reforms are likely to support efforts to better use public data. ICT investments in schools undertaken by the UK government in recent years have contributed to the increase in educational performance in primary schools.[19] This

[15] Bloom, N., Sadun, R., Van Reenen, J., 2007, Americans do I.T. better: US multinationals and the productivity miracle, NBER Working Paper Series.

[16] Franklin, M. Stam, P. & Clayton T., 2008, "ICT impact assessment by linking data across sources and countries".

[17] https://www.gov.uk/government/publications/shakespeare-review-of-public-sector-information

[18] https://www.gov.uk/government/publications/information-economy-strategy

[19] Machin S. & McNally S. & Silva O., 2007. "New Technology in Schools: Is There a Payoff?," Economic Journal, Royal Economic Society, vol. 117(522), pages 1145-1167, 07.

implies productivity gains related to ICT also for non-commercial use. Taking these findings together with anecdotal evidence on the productivity gains related to text mining, there seems to be a strong case to remove potential barriers to its use. Evidence from the consultation responses suggesting that copyright licenses are not granted for text mining purposes in many cases or the barriers to obtaining such licences are unreasonably high thus calls for government intervention.

Risks and assumptions

We have assummed that exercise of the exception will not act as a channel for copyright infringement, through abuse of the exception to develop substitutable products or an enduring archive of the originals.

That text and data mining in general have the potential to deliver benefits for UK research, and it is therefore appropriate to seek measures to support their deployment, consistent with not undermining research or publishing in other respects.

That publishers will continue to be able to offer enhanced services outside the scope of this exception.

Cost and benefits of the exception could well grow as further technological applications emerge.

The exception will only be available for non-commercial research, only where researchers have obtained legitimate access to works, and will not prevent the publisher / provider from applying technical protection measures on networks used in order to maintain security or stability.

The market for analytic technologies is developing, so it is not clear what the full potential is. Both costs and benefits could grow as further technological applications emerge. However, the exception would remain limited to non-commercial research.

Summary and preferred option with description of implementation plan

Option 1 - Create a copyright exception to cover text and data analytics for non-commercial research within certain restricted limits, in order that copying and/or any related use of works for the use of text and data analytics will not require specific additional authorisation. Any contractual provision contrary to this shall be null and void.

We have chosen option 1 because it offers a solution in a limited area where the added value of technologies is provided primarily by the user rather than the publisher/supplier, while not preventing publishers from applying technical protection measures on networks used in order to maintain security or stability. Whilst we have not been able to quantify the exact costs and benefits, we have used the research by JISC, consultation responses, and case studies provided to illustrate that the benefits of implementing a text mining exception are significant and as the exception only applies to non-commercial research the benefits considerably outweigh the potential costs.

The Government will also explore what further measures might be provided to enable publishers and other stakeholders to improve support for and deployment of these technologies, including by clarifying what can already be done under the existing law, and by supporting development of licensing process where licensing is the best solution, to ensure appropriate investment, security, and low-cost and efficient licensing processes. However, these measures would be in addition to this limited exception.

Direct Costs and Benefits to Business Calculations (following OIOO methodology)

Under the "One In, One Out" rule, whereby a measure that has a net cost to business must have a measure or measures of equivalent cost removed in order to be implemented. We are not able to quantify the costs and benefits of our chosen option but have used the research by JISC and the case studies provided in response to our consultation to illustrate that the benefits will be significant. However for the purposes of OIOO methodology we have therefore we have counted this as a zero-cost measure.

Implementation Plan: The proposal will be implemented through UK secondary legislation after full consultation, as part of a wider package, minimising additional specific costs.

Evaluation

A full evaluation strategy and Post Implementation Review is being developed for the introduction of the Hargreaves recommendations. The Post Implementation Review will detail the benefits associated with the introduction of the copyright reforms and will include input from external stakeholders. The plan will also set out how and when the benefits will be measured, which will depend on the type of benefit, as some benefits will be measured by applications and take-up that can be measured from the first year of operation, whereas others will depend on information that will take several years. The evaluation strategy will set out the activities that will be undertaken in order to evaluate the policy, drawing on management information collected through the copyright system, as well as research that is commissioned in order to measure the benefits.

The main source of data available for evaluation will be collated using industry figures. These statistics, alongside other management information on the operation of the system will be used by Government to assess the impact of the copyright reforms, including assessing whether benefits have been achieved and how policy or operations can be developed to realise benefits more effectively

_____

***Note**

This Impact Assessment (IA), which is one of a series of IAs concerning copyright exceptions, was originally published in December 2012 alongside the policy statement "Modernising Copyright". It was republished to accompany the publication of the secondary legislation that implements the chosen options set out in the IAs. Since the original publication of the IAs the Government has engaged extensively with stakeholders both formally (through a technical review of the draft legislation) and informally. No new evidence has been provided that has led the Government to alter the overall numerical assessment of the costs and benefits contained within this IA.

However, since the IA was originally published the government has become aware of some additional narrative evidence that is of relevance to arguments discussed in the IA and hence the narrative section has been updated.