RELX response to Intellectual Property Office's Consultation on Artificial Intelligence and Intellectual Property

Introduction

RELX is a UK-based provider of information and data analytics for professional and business customers across a range of sectors, including financial services, science, technology, medical, healthcare and energy.

We employ over 5,000 people in the UK and support customers in 180 countries. We utilise technology and data to help our customers improve their decision-making across the sectors we serve. We help scientists make new discoveries, doctors and nurses improve the lives of patients, lawyers win cases, prevent online fraud and money laundering, and help insurance companies evaluate and predict risk.

RELX welcomes the opportunity to respond to the government's consultation on Artificial Intelligence (AI) and Intellectual Property (IP). With an annual investment of more than £1 billion in technology and with 8,000 of our 33,000 employees being technologists, RELX is both a significant consumer and developer of the latest technologies, including techniques which fall within the broad definition of AI.

Our AI and machine-learning use cases range from discovering new therapies for rare diseases via drug repurposing using Elsevier's Entellect platform, to helping identify farms and fields at risk of pests or disease through Proagrica, part of RELX's Risk & Business Analytics division. At the same time, RELX is the rightsholder of large numbers of copyright works, most notably through our science, technology and medical journals publications within our Elsevier business. This high-quality scientific research content is in many cases the data which powers machine-learning applications.

The UK is rightly ambitious in its desire to be a world leader in AI and we welcome the government's National AI Strategy. One of RELX's key tech hubs is located in London and we support the commitments made in the National AI Strategy to invest in the UK's AI ecosystem and look forward to the White Paper due to be published on governing and regulating AI. RELX has also responded positively to the government's consultation on reforming data protection rules in the UK, another key aspect of the UK's National AI Strategy and National Data Strategy.

As the Government considers its approach to both AI and IP in the coming months, we stand ready to share our expertise in both areas.

Overall Remarks

Before answering the specific questions set out in the consultation, we would like to take this opportunity to make some broader points on the interaction between AI and IP. This section also speaks to questions 18-21 in the IPO's consultation document.

Some participants in the debate often claim that IP – and copyright in particular – represents a barrier to the development of AI. Hence, they argue the need for further copyright exceptions to allow greater or totally free access to content for the purpose of TDM and the onward development of AI systems.

However, instead of viewing AI and IP as competing priorities, we believe that AI and IP are highly complementary: copyright is vital in supporting the development of trustworthy AI. The purpose of

IP rights is to incentivise investment in innovation and new ideas through the author having limited exclusive control over the use of the work and thereby protection against infringement. With this framework in place content creators can securely invest in the production of high-quality, authoritative, verified works. Such works go to form the data and content sets relied upon by AI systems. In this context copyright therefore plays a key role in ensuring the content utilised by AI is high quality, thereby reducing the risk of a 'garbage in, garbage out' dynamic and the negative consequences that might flow for individuals and society from AI-decisions based on inferior content.

We believe it is critical, as a user of both proprietary and third-party copyright-protected content in our own AI systems and data-driven technologies, that any proposed changes to the IP regime should give careful consideration to the effect those changes might have on the incentives to create high quality content which, in turn, would have an impact on the development of AI more broadly.

To give an example of what is at stake from a public interest perspective, RELX's Elsevier business has created an information product, *ClinicalKey*, comprising standardised clinical guidelines and other medical information built through the integration and enrichment of proprietary and third-party licensed content. *ClinicalKey* is used extensively across the NHS and relied on by clinicians as a trusted source of insights and guidance when determining care plans and treatments for patients. *ClinicalKey* has TDM-powered search capabilities, but the quality of the insights surfaced by these search capabilities is heavily dependent on the quality of the underlying content. Put simply, poor content would lead to poor outcomes for patients, which is why the copyright regime must continue to incentivise investment in high-quality, authoritative, verified works.

Finally, as the National AI Strategy notes, the UK is a world leader in AI, ranking third for private investment in AI in 2020 behind only the USA and China. This success has been achieved, in part, because of the current intellectual property framework that facilitates investment in high quality content that is valuable for text and data mining. This positive relationship between IP and AI/TDM is one which should be maintained as the UK seeks to go even further with AI.

Responses to Questions

Copyright – text and data mining

1) If you license works for TDM, or purchase such licences, can you provide information on the costs and benefits of these? For example, availability, price-point, whether additional services are included or available, number and types of works covered by the licence etc.

As noted above, RELX businesses engage in activities that both license out and license in content and so we have a wide perspective on how the market for copyright-protected information currently operates.

For example, Elsevier's *Scopus*, the world's largest abstract and citation database of peer-reviewed literature, including scientific journals, books and conference proceedings, was developed and continues to be maintained by integrating content that is licensed from about 12,000 publishers. While this is by no means a simple task, it does show that it is possible to access significant amounts of works through a licensing approach and to build a new product on solid copyright-friendly foundations.

Our view is that a licensing approach provides the most effective way of delivering and accessing copyright-protected material regardless of whether consumption is via conventional means (i.e.

human reading), or for TDM analysis. Our Elsevier business has many licensing agreements in place with commercial entities that enable TDM. For example, we have commercial relationships with leading pharmaceutical companies that deliver content to them specifically to support their R&D-focussed analytics.

The benefits of the licensing approach for TDM are extensive:

i. **Enables normal exploitation.** Licensing gives rightsholders the opportunity to exploit the work in the normal fashion whilst evolving their licensing approaches to meet the changing patterns of consumption and needs of users (e.g. the trend towards greater machine consumption which in some parts of the market may become the norm).
ii. **Promotes innovation.** It gives rightsholders the ability to develop new licensable services for TDM, such as providing the content in machine digestible formats in combination with enhanced delivery mechanisms (e.g. 'Data as a Service' - DaaS)
iii. **Enables legitimate content protection.** Licensing gives rightsholders the ability to protect content by giving them clarity on which organisations or individuals have legitimate access to content, a matter of critical importance given that whole content sets or large subsets of content sets might be being licensed out.
iv. **Gives certainty to licensed users.** Licensed organisations and their users receive clarity as to what they are permitted to do with licenced content. This avoids confusion and disputes and affords greater certainty than relying on copyright exceptions (n.b. for these reason RELX's more content acquisitive businesses strongly prefer contract over copyright exceptions as their means of accessing and using third-party content).
v. **Develops the market.** The flexibility of licences helps to drive the development of a market for TDM as it enables both parties to agree on what the value of that content is from a TDM perspective and ensures there is not a compulsory free transfer of value from one party or sector to another and that the investment by rightsholders into that content is recognised.
vi. **Encourages investment.** With licensing protecting existing investments in content, rightsholders are in a position to invest further in high quality, authoritative and verified works. This has the broader benefit of ensuring high quality data and content exists for the training of AI systems.

It is worth noting that the market for content for use in TDM and other technologies is still fairly nascent, given AI and related technologies are themselves still developing phenomena. This makes it difficult to provide a large amount of evidence to assess the size of the current market. We would therefore respectfully suggest that a particular degree of caution should be exercised before significant changes are made to copyright laws and instead the market should be allowed to develop.

Given the government's desire for data insights relevant to the market in this area, we would be happy to discuss this topic in more detail with officials, if that would be useful.

2) <u>Is there a specific approach the government should adopt in relation to licensing?</u>

An approach that encourages the greater use of licensing as a basis to access more copyright-protected content would be a positive outcome of this consultation as it would support the further development of the market for content across all use cases.

The government may wish to consider the role it could play in developing model TDM licensing terms for commercial use. We believe these could particularly help smaller rightsholders, which may not have the resources or in-house legal expertise to develop licences from scratch that would give

access to third parties and yet give rightsholders the assurances and protections they need enter into a contractual agreement through which they might be allowing the entirety of their content sets to be copied and stored by a third party for the purpose of TDM.

3) Please rank the options in order of preference (most to least preferred) and explain why.

Based on an assessment of each of the option presented for TDM, RELX's order of preference is:

Option 0 – No legal change
Option 1 – Improve licensing environment for the purposes of TDM
Option 3 – Adopt a TDM exception for any use, with a rights holder opt-out

Our reason for putting Option 0 first is that we are not aware of evidence which suggests there is a problem with the current legal framework in the UK for accessing copyright protected material for the purposes of the TDM. This is based on our experience as an acquirer of content under licence, as well as rightsholder.

Option 1 is our second preference as this represents a minimal disruption to the current system. Again it is not clear from evidence that improvement in the licensing environment is urgently needed but we would support measures to improve conditions if required.

If a change to the UK's TDM exception is deemed necessary then Option 3, which retains the rightsholder's ability to opt out, is a preferable one. It would be important to ensure that the rightsholder opt-out is an express reservation against the reproduction of works for the purposes of TDM and that it is provided in machine-readable form, so as to be understood by web crawlers. However, the Digital Single Market Directive (2019) and its Article 4, which sets out a provision similar to what is floated in Option 3, is new and so it is too early to judge its impact.

For the reasons we outline further below we do not believe that Option 2 (extend the existing TDM exception to cover commercial research and databases) and Option 4 (a TDM exception for any use, which does not allow rights holders to opt out) are viable.

4) If you have experience of the EU exception with opt out for rights holders, how has this affected you?

The EU exceptions for TDM as enshrined in Articles 3 and 4 of the Directive on Copyright in the Digital Single Market have only had legal effect since June 2021. In some Member States, such as France, implementation has only happened in the past month, and in others such as Austria, there is still no national implementation. This makes it impossible to make a judgment on the impact of these measures on the operation of TDM, the behaviour of licensors and licensees, or on the market conditions generally. However, we closely monitor the market and would be pleased to provide the government with greater insight on how the EU exception is operating when more information is available.

RELX businesses alongside other organisations are involved in developing technical solutions for how rightsholders could implement the technical means to reserve their rights in respect of copyright-protected content, as permitted under the Directive. We would be happy to share more details with officials on the progress made so far on these solutions if that would be of interest. If the UK ultimately opts for an exception that allows TDM to be applied to copyright-protected content for any use provided the user has lawful access and with a rightsholder opt-out (i.e. Option 3), it may be

helpful for the government to better understand the work being undertaken on rights reservation solutions and their viability as an effective protection for rightsholders.

5) How would any of the exception options positively or negatively affect you? Please quantify this if possible.

Under Article 9 of the Berne Convention for the Protection of Literary and Artistic Works 1967, national legislation to permit reproduction of works should apply to *"certain special cases, provided that such reproduction does not conflict with a normal exploitation of the work and does not unreasonably prejudice the legitimate interests of the author".*

Options 2 and 4 fall foul of at least the second two steps of this "three-step test". As for the first step, whilst the purpose of TDM can be argued to be a "certain special case" if an exception is applied to any instance in which TDM is undertaken it begins to look like a less specific provision.

Scientific publishers have a strong and growing business in licensing of works to commercial users for the purposes of TDM – and onward usage in AI systems. In the present era, this constitutes as "normal" an exploitation of the work as the selling of a physical journal subscription would have been at any point in the last century. Any exception which rendered this commercial arrangement moot or unviable would clearly impinge upon publishers' exclusive rights. In so doing it also clearly runs directly against the legitimate interests of the author, as represented by the publisher to which the exclusive rights are transferred.

Not only would such wide exceptions have negative consequences for publishers and authors but, we would also argue, for the wider research environment. The revenues which publishers generate from TDM are the basis for incentivising investment not only in further high-quality content, but in the technological tools that make TDM and AI possible. A licensee of content receives a lot more than just access to the works. They are provided with an Application Programme Interface (API) which make the access to and ingestion of the content technically feasible. These tools are being developed and improved constantly, driven in part by the investment of publishers. Any changes to the copyright regime which imperilled these revenue streams would severely undermine the investment in TDM technology.

Furthermore, we envisage that the introduction of Options 2 or 4 would lead to significant increases in copyright infringement and the creation of derivative works that incorporate infringing content. The prospect of commercial gain would tempt many users with lawful access to misuse content in ways they do not under the existing explicitly 'non-commercial use only' TDM exception. Indeed, we would expect that some would be tempted to exploit commercial use exceptions and reverse engineer content-based products, making these or subsets of them available on the market. (Again this offends the Berne requirement that an exception should not lead to conflict with normal exploitation)

While existing legal remedies to infringing content would remain, we are concerned that these exceptions would make it more difficult for rightsholders to protect and control their portfolio of works. Rightsholders would have little knowledge of how their works are being used, the whereabouts of whole content sets that have been copied, or the extent to which users who have taken copies continue to have lawful access. Such problems when taken together might create a huge enforcement problem that does not exist today and put additional burdens on rightsholders seeking to protect their content by requiring them to chase down a much larger body of infringing works.

For further information on this submission, or for further discussions please contact:

[redacted]