

INTELLECTUAL PROPERTY OFFICE: CONSULTATION ON COPYRIGHT

Response by the Wellcome Trust

March 2012

KEY POINTS

- We strongly support the proposed exception to allow text and data mining for non-commercial research. The current situation, in which researchers must gain permissions from individual publishers to mine content to which they already have legal access, prevents the effective use of these tools, and threatens to limit drastically the significant potential economic and societal benefits that could be gained.
- We agree with the proposals to establish a system to register and authorise the use of orphan works – such a system, in particular the requirement to conduct diligent searches for rights holders, must be proportionate and recognise that archival material should not be held to the same standards as published works.
- We support the proposals to extend the exceptions in relation to library preservation and on usage for research and private study.
- We consider it to be essential to ensure that copyright exceptions cannot be overridden by contract.

INTRODUCTION

1. The Wellcome Trust is pleased to respond to the Intellectual Property Office's consultation on copyright. As a global charitable foundation dedicated to achieving extraordinary improvements in human and animal health, we are committed to ensuring that the outputs of the research we fund – including both research publications and data - can be widely accessed and used in a manner that maximises the resulting benefits to society.
2. We also provide the Wellcome Library – one of the world's foremost resources for research and discovery on the history of medicine and on contemporary developments in biomedical science and health. In the past year the Wellcome Library had more than 40,000 in-person visitors and around 500,000 unique visitors to its web sites. The Library is also undertaking a major transformation project to create a ground-breaking digital library that will expand access to our collections for users the world over.
3. We agree fully with the conclusions of the Hargreaves Review in finding that UK copyright legislation has failed to keep pace with developments in the digital age: it now serves as a barrier to innovation and growth, and limits the economic and wider societal benefits that result from public and charitable investments in research. We strongly support the recommendations of the Hargreaves Review, and are pleased that the Government has accepted them in full. We believe that it is vital that these crucial reforms are implemented without delay, particularly in the context of the Government's wider commitment to stimulate growth and unlock the immense value of open data.
4. Our submission focuses on the proposals in the consultation document relating to orphan works and exceptions to copyright. It summarises our views in response to the

questions in these sections of relevance to our organisation, and includes additional case studies that illustrate our key arguments.

ORPHAN WORKS (CHAPTER 4)

Q1 Would your organisation make use of this proposed system for the use of individual orphan works? How much of the archive is your organisation likely to undertake diligent searches for under this proposed system? What would you like to do with orphan works under a scheme to authorise use of individual orphan works?

Q3 Please describe any experiences you have of using orphan works (perhaps abroad). What worked well and what could be improved? What was the end result? What lessons are there for the UK?

5. In terms of the scale of the orphan works in the Wellcome Library's collections, we would expect that a significant proportion of 20th Century works that we hold (and especially those published before 1945) will be orphan works – but this could only be validated after a diligent search has been conducted.
6. We strongly support the proposals in the consultation document to create a registry and a scheme to authorise the use of orphan works, once a diligent search has been undertaken. The Wellcome Library would certainly make use of such a service, and would seek to make designated orphan works freely available through its web site under a non-commercial licence. Collectively, these works form a major part of the record of the history of medicine, and if rendered inaccessible to all except those who visit in person, would severely limit our ability to open up the Library's collections to a worldwide audience for learning, research and discovery
7. The Wellcome Library is already undertaking diligent searches as part of its extensive digitisation programme. We have entered into a contract with the Author's Licensing and Collecting Society (ALCS) and the Publisher's Licensing Society (PLS), who are providing a paid diligent search service for around 2,000 books we wish to digitise and make freely available through our website. Further details are provided in **Box 1 (below)**.
8. The Wellcome Library also holds a large quantity of archival material (more than 500 separate collections containing over 1.5 million items). A due diligence regime appropriate for published works is not feasible for this content, which is unstructured and does not conform to any standard publishing conventions. Identification of rights holders would be such an onerous task as to render impossible any attempts to clear copyrights. By way of example, our archive of the personal papers of the serologist Arthur Mourant (1904-1994) contains in excess of 300 separate rights holders and the information about these individuals is often limited to just their name.

Box 1 – a rights-identification service provided to the Wellcome Trust by the ACLS and PLS

The Authors' Licensing and Collecting Society (ALCS) and the Publishers Licensing Society (PLS) have agreed to provide (for a fee) a "Rights Identification Service" for the Wellcome Trust as a pilot project, which involves trying to identify rights-holders for around 2,000 books that the Wellcome Library wishes to digitise and make freely available via its web site. This project will help the Wellcome Trust assess what should constitute a "diligent search" for **published works**.

For the purposes of this pilot the three partners have agreed that the following steps should be taken to identify a rights-holder:

- Determine from the ARROW database the copyright status for relevant titles (USA, UK, France, Spain and Germany);
- Determine from the ARROW database whether a title published in one of these countries is "in-commerce";
- Search ALCS/PLS member databases;
- Search ALCS/PLS works databases (e.g. COPAC, Nielsen);
- Search partner member data and external databases (e.g. SoA, PLR, WATCH);
- For non-UK titles/authors send queries to RROs and other overseas partners;
- Carry out Internet searches for titles not covered by ARROW and/or other accessible databases;
- Pursue resulting contacts (emails, letters etc. to follow up leads identified through the searching processes e.g. current/former employer) - this process will continue until all leads have been exhausted. In cases where a contact/ lead does not respond, one further follow-up communication will be sent.

As a result of this process, the ALCS/PLS will, for each title, indicate whether a work is:

- a. Out of copyright;
- b. In copyright - and the rights holder(s) are known and contactable;
- c. In copyright - but the rights-holders cannot be traced.

For items in category (b) the Wellcome Trust will contact the rightsholders to seek the necessary permissions. For items in category (c), the Wellcome Trust will treat such works as "orphans" and make them freely available on the Wellcome Library website. If a rights-holder subsequently comes forward, the Trust will, on request, remove the item from the website or seek permission from the rights-holder to keep it online.

As the agreement with the ALCS/PLS was only concluded in December 2011, it is too early to know whether this model will be effective in identifying rights-holders (and by implication, identifying which works are orphans). However, all three organisations are committed to making this project work, and we would be happy to share information with the IPO when it becomes available.

Q4 What do you consider are the constraints on the UK authorising the use of UK orphan works outside the UK? How advantageous would it be for the UK to authorise the use of such works outside the UK?

Q5 What do you consider are the constraints on the UK authorising the use of orphan works in the possession of an organisation/individual in the UK but appearing to originate from outside the UK: (a) for use in the UK only; (b) for use outside the UK? How advantageous would it be for the UK to authorise the use of such works in the UK and elsewhere?

9. We consider that it will be essential that the UK is able to authorise the use of UK orphan works by those outside of the country. It should also, if at all feasible, be able to authorise the use of orphan works originating outside of the UK that are in possession of a UK-based organisation – again both in the UK and worldwide. If authorisation for the use of orphan works was restricted to the UK, it would:

- generate additional costs – particularly if there was a need establish systems to limit access to UK internet protocol (IP) domains;
- limit the ability for the UK to leverage a global audience for this content (some of whom may be willing to pay for access);
- fail to recognise that research is conducted on a global scale (i.e. all researchers need access to relevant content);
- result in the fragmentation of research and cultural resources if other countries adopted similar processes.

Q6 If the UK scheme to authorise the use of orphan works does not include provision for circumstances when copyright status is unclear, what proportion of works in your sector (please specify) do you estimate would remain unusable? Would you prefer the UK scheme to cover these works? Please give reasons for your answer.

10. Our experience is that a large proportion of works, published in the late 19th Century and early-to-mid 20th Century could fall into this category. Preliminary investigations carried out by the Wellcome Library, based on a sample of 560 books published between 1850 and 1960, revealed that it was not possible to find the author's death date for 37 per cent of these works through searches using a wide variety of sources, including:

- Library catalogues (including authority records from the US Library of Congress);
- Dictionary of National Biography;
- The Times newspaper;
- WATCH files;
- National Register of Archives;
- Bowker Books in print;
- ACLS online search for royalties list;
- Classical Authors directory;
- Google searches.

11. As is evident from this list, determining copyright status is a time-consuming and therefore costly process, and even rigorous searches of this type will fail to uncover the status in a significant proportion of cases. For these reasons, we would argue that it is imperative that where due diligence has been undertaken, works with unknown copyright status should be included within the proposed scheme.

12. The ARROW system is designed to aid this search process, but is currently limited to a small group of western European countries. If this system was expanded, then this would have the potential to help to streamline the process.

Q7 If the UK's orphan works' scheme only included published/broadcast work what proportion of orphan works do you estimate would remain unusable? If the scheme was limited to published/broadcast works how would you define these terms?

13. We consider that the scheme should cover all works, including works which have not been formally published (e.g. archives). However, we would argue strongly that requirements for undertaking diligent searches should be proportionate to each collection, and therefore, for the reasons set out in paragraph 8, considerably lower for archives and other non-published materials than would be the case for published works. Applying a single standard to identification of rights holders for published and unpublished works is likely to mean that few, if any, archive collections from the past 100 years are made available to researchers, other than by physical consultation. The latter does not, of course, allow any greater protection for rights-holders, other than by artificially limiting the number of users available to access such material.

14. As noted above, the Wellcome Library has a vast amount of archival material. Our digitisation project includes the provision of online access to such archives; indeed, of the 4.1 million pages the Library will digitise over the next two years, 1.1 million (27%) will be of archival format. In the absence of legislation regarding orphan works, the Library is currently carrying out its own due diligence on its archival holdings, and taking a risk management approach to which materials it will make available. It is important to note that in the case of recent archival material, issues relating to sensitivity and personal data are, in our view, likely to be of greater concern, but these should not be conflated with copyright.

Q8 What would be the pros and cons of limiting the term of copyright in unpublished and in anonymous and in pseudonymous literary, dramatic and musical works to the life of the author plus 70 years or to 70 years from the date of creation, rather than to 2039 at the earliest?

15. The Wellcome Library has numerous examples – including manuscripts, recipe books and archives dating from ancient times – which, because they were never published, are still in copyright and will remain so until 2039. This naturally limits access and use of such materials. This position seems absurd and a change of the type proposed would allow libraries and archives to make this content available without potentially breaking the law.

Q9 In your view, what would be the effects of limiting an orphan works' provision to non-commercial uses? How would this affect the Government's agenda for economic growth?

16. Although not our primary area of expertise, it would seem essential not to limit the provision to non-commercial uses. Such a move would limit opportunities for innovation and economic growth from wider use of these works, and potentially serve to restrict applications that could result in genuine benefits for society.

Q12 In your view what should constitute a diligent search? Should there be mandatory elements and if so what and why?

Q13 Do you see merit in the authorising body offering a service to conduct diligent searches? Why/why not?

17. We believe it is essential that there is clear *guidance* on the key elements of a diligent search, but this should not seek to define an inflexible, mandatory list of obligations. As we have made clear, different types of works (e.g. published works and unpublished works) will require very different approaches. Specifically, the search should be proportionate to the feasibility of identifying and tracing copyright holders.
18. Our initial view is that it would not be appropriate for the authorising body to itself offer a diligent search service. There would be a risk that this could potentially give such a body an effective monopoly over the provision of such services, preventing competition and undermining the ability of users to seek the most cost-effective option.
19. Anyone should be able to provide a rights identification service, but these entities (be they private or public) should be formally authorised to undertake this activity (along the lines of a BSI certification scheme).

Q14 Are there circumstances in which you think that a diligent search could be dispensed with for the licensing of individual orphan works, such as by publishing an awaiting claim list on a central, public database?

20. As noted above, we would suggest that for published works, an obligation to undertake a due diligence search is feasible and appropriate. For unpublished works, we believe that a more flexible and less onerous approach is required. As proposed above, diligent searches should be proportionate to the feasibility of identifying and tracing copyright holders.
21. We believe that the posting of a list of works in an online, centralised "orphan repository", should be sufficient. However, to protect the reventant rights-holder, this content should be "authorised" subject to a requirement that the entity hosting such works must have a clear and explicit take-down policy.

Q15 Once a work is on an orphan works registry, following a diligent search, to what extent can that search be relied upon for further uses? Would this vary according to the type of work, the type of use etc? If so, why?

22. Once a work has been placed on the registry, we believe that this "ruling" should apply to all uses of that work (i.e. there should not be a need to carry out a second diligent search for the same item).

Q16 Are there circumstances in which market rate remuneration would not be appropriate? If so, why?

Q17 How should the authorising body determine what a market rate is for any particular work and use (if the upfront payment system is introduced)?

Q18 Do you favour an upfront payment system with an escrow account or a delayed payment system if and when a reventant copyright holder appears? Why?

23. For the books the Wellcome Library is planning to digitise, we believe that there is no market rate. By definition, all the books are "out of commerce". It should be noted that our own digitisation programme will result in in-copyright material being made available without charge for the purposes of research under a non-commercial licence. In circumstances where other institutions or individuals provide access to in-copyright material via a revenue-generating (rather than free or cost-neutral) model, this situation would be different. However a blanket assumption in favour of remuneration would be inappropriate, dissuading as it would digitisation projects such as ours undertaken without profit, for the public good.

24. The Wellcome Trust favours a delayed payment system. The experience in Canada suggests that the costs associated with setting up an escrow account are too high. Moreover, if a diligent search has been conducted then the chances of a reventant rights holder appearing are low. As such, we do not believe there is a case for tying up scarce funds in an escrow account which, in all likelihood, will rarely be needed.

Q19 What are your views about attribution in relation to use of orphan works?

Q20 What are your views about protecting the owners of moral rights in orphan works from derogatory treatment?

25. We do not believe that moral rights would be affected by the proposed changes, and hence we agree with the conclusion of the consultation document that there wouldn't be a need to alter the UK's moral rights regime. We consider that the use of orphan works, like any other works, should be attributed fully. Even if a rights-holder cannot be identified, there should be sufficient metadata available to at least attribute the work. We agree with the argument in the consultation document that the proposed system should increase the likelihood that rights holders can identify works currently believed to be orphans - reducing the risk of illegal use and enabling such rights-holders to assert their moral rights.

Q21 What are your views about what a user of orphan works can do with that work in terms of duration of the authorisation?

26. If authorisation is limited to specific time period (e.g. five years) then, by default, the costs of administering the registry would increase. We consider that authorisation should be valid for as long as its orphan work status remains.

EXCEPTIONS TO COPYRIGHT (CHAPTER 7)

Q72 Should the preservation exception be extended:

- to include more types of work?
- to allow multiple copies to be made?
- to apply to more types of cultural organisations, such as museums?

How might this be done, and what would be the costs and benefits of doing it?

Q73 Is there a case for simplifying the designation process which is part of Section 75? How might this be done and what would be the costs and benefits of doing it?

Q74 Should any other changes be made to the current exceptions relating to libraries and archives, and what would be their costs and benefits?

27. We strongly agree with the case set out in both the Hargreaves review and in the consultation document for extending the scope of the preservation exception. It seems perverse that copyright legislation should inhibit the ability of cultural organisations to preserve artistic works. Expanding the exception to cover sound recordings, film and broadcasts; allowing multiple copies to be made; and enabling museums and galleries to benefit from this exception would be concrete steps forward, and it is hard to see any significant downside to these proposals.

Q75 Would extending the copyright exception for research and private study to include sound recordings, film and broadcasts achieve the aims described above? Can you provide evidence of its costs and benefits?

Q76 Should the copyright exception for research and private study permit educational establishments, libraries, archives or museums to make works available for research or private study on their premises by electronic means? What would be the costs and benefits of doing this?

28. We also support the proposal to extend the exception for research and private study so that it is “work-neutral” and incorporates sound recordings, film and broadcasts. There seems no reason why these materials should not be available to researchers and students on the same basis as other works under copyright – and, as the consultation document suggests, this could greatly enhance the value that could be generated from these works in terms of advancing of knowledge, innovation and skills development. Although we do not have firm evidence on the costs and benefits, it seems entirely appropriate in the digital age that libraries, archives and museums should be able to make works available in electronic format.

Q77 Would an exception for text and data mining that is limited to non-commercial research be capable of delivering the intended benefits? Can you provide evidence of the costs and benefits of this measure? Are there any alternative solutions that could support the growth of text and data mining technologies and access to them?

29. We consider that an exception that allows text and data mining for non-commercial research is a crucial first step. The ultimate goal should certainly be to extend this exception to also allow commercial uses, and we would urge the Government to push for the reforms at European level needed to enable this. In the meantime, however, there is a compelling need to introduce a non-commercial research exception without delay.

30. As the volume of scientific information continues to increase at a rapid rate, the use of text and data mining will become ever more crucial to the research enterprise. These tools not only enable researchers to identify information of interest from across the scientific literature to which they have lawful access, but can also uncover new and unsuspected associations – stimulating discovery and opening up novel avenues of research and innovation. In short, these tools hold enormous potential both to increase the efficiency of the research enterprise and to enhance the economic and societal benefits resulting from public and charitable research funding.
31. Recent research commissioned by JISC explored the value and benefit of text mining and analytics to the UK further and higher education sector¹. This study highlights the significant potential benefits to the sector in terms of time savings and productivity gains. It finds, however, that the use of text mining is currently very restricted, due to the significant barriers to uptake – which include uncertainty over the current legal situation, entry costs, lack of capacity and the division of information into silos. The analysis indicates that the current situation has strong characteristics of market failure, and is preventing society from realising the full benefits of text and data mining.
32. To further investigate the barriers that constrain the adoption of text mining under current copyright rules, the Wellcome Trust asked the European Bioinformatics Institute to conduct an analysis of what would be required to gain the permissions needed to conduct a hypothetical text mining exercise across a sample of the published literature in a particular field, using the area of malaria as an example. The results of this case study are presented in **Box 2** (overleaf). It illustrates clearly that the current situation is neither scalable nor cost effective, particularly given the growing volume of published research.
33. Of course, it is also important to note that it is far from guaranteed that publishers will grant permission once it has been sought. Research commissioned by the Publishing Research Consortium² shows that in general only 35% of publishers grant permission in most or all cases. For research driven requests the figure is higher, but even here, less than two thirds of publishers routinely agree to text mining requests, and do so on a case-by-case basis.
34. The introduction of the exception would help to alleviate these barriers, and unlock the potential benefits of text and data mining tools. The publishing sector has argued that an exception is not required as individual publishers will develop text mining tools for the content they provide. But such a system would embed information silos and drastically limit the value that can be gained from the use of these tools – which depend on an ability to analyse the entire body of literature, rather than that provided by a single publisher.

Box 2 – An illustration of the resource required to undertake text and data mining

Imagine that a malaria researcher wished to use text mining tools across the current published literature in malaria, in order to identify possible new avenues for drug treatment against this devastating global disease.

In the free-to-access, UK PubMed Central repository there are 2,930 full-text articles, published since 2000, which have the word “malaria” in the title. Of these 1,818 (62%) are open access, and thus suitable for text mining without having to seek permission. However, the remaining 1,112 articles (38%) are not open access, and thus permission from the rights-holder to text-mine this content must be sought.

These 1,112 articles were spread across 187 different journals, published by 75 publishers. As publisher details are not held in the UKPMC database, the permission-seeking researcher would need to find the appropriate contact at every journal and make contact with them. Using a highly conservative estimate of one hour research per journal title (i.e. to find contact address, indicate which articles they wish to text-mine, send letters, follow-up non-responses, and record permissions etc.) this exercise will take 187 hours. Assuming that the researcher was newly qualified, earning around £30,000 pa, this single exercise would incur a cost of £3,399.

In reality, however, the researcher would not limit his or her text mining analysis to articles which contained a relevant keyword in the title. Thus, if we expand this case study to find any full-text research article in UKPMC which mentions malaria (and published since 2000) the cohort increases from 2,930 to 15,757.

Of these, some 7,759 articles (49%), published in 1,024 journals, were not open access. Consequently, in this example, a researcher would need to contact 1,024 journals at a cost (in terms of time spent) of £18,630; 62% of a working year. Clearly, it rapidly becomes completely impractical for the researcher to consider undertaking such an exercise.

Notes

1. This search was conducted on 1 November 2011, by staff at the European Bioinformatics Institute. A copy of the spread-sheet listing those journals in the non-open access set is available on request.
2. Typically around 35% of the content in UKPMC is open access. The exceptionally high figure for malaria reflects the Wellcome Trust's major funding for malaria research and its requirement that all the research it funds to be made open access.
3. The British Library estimated that it took 302 hours to just to identify 299 rights holders. See: <http://www.jisc.ac.uk/media/documents/publications/infromthecoldv1.pdf> p13 The case-study above uses an average time of 1 hour research, per journal title. This would allow time for the researcher to identify where to send the address, specify which articles they wanted to text-mine, chase non-respondents, record permission etc.
4. This calculation assume that there are 220 working days in a year, and that a researcher works 7.5 hours a day. This equates to 1,650 hours a year, of which 187 hours equals 11.3%. Starting salaries for newly qualified researchers based on UCL data, is available at: http://www.ucl.ac.uk/hr/salary_scales/final_grades.php
5. The total number of research articles in UKPMC, published between 2000 and 2011, which contain the term “malaria” totals 15,757. Of these, 7,998 articles (published in 557 journals) were Open Access, whilst the remaining articles (7,759, published in 1,024 journals) were not Open Access. This analysis was conducted on 22 Dec 2011.

35. It has been suggested that the industry could potentially develop a cross-publisher text-mining licencing solution, based around the development of a model licence. Although there may be some merit in a model licence, we would argue that it is highly unlikely such a solution would be effective in practice in facilitating text mining. The experience of JISC Collections, who sought to introduce a clause in its Model Licence to permit users in institutions to undertake text and data mining, provides one compelling illustration. Of 17 recent journal licence agreements negotiated with major journal publishers, six publishers (35%) accepted the clause, six publishers (35%) rejected this clause in its entirety, and five publishers (29%) made significant amendments to the clause³. For example, one amendment included a requirement that all text-mining requests be submitted to the publisher's Account Manager who would then consider these on a "case by case basis".
36. In a licence-based solution, the onus would still be on the individual researcher or their institution to check whether text-mining was permitted and the associated terms for each piece of content.
37. Publishers also argue that, due to limited bandwidth, their systems will be unable to cope with the widespread use of text and data mining tools. But such an argument is difficult to equate with the experiences of existing open access publishers – such as Biomed Central – who already provide access to their entire published outputs without any such problems.
38. Moreover, the introduction of a text and data mining exception would not prevent publishers from exploring appropriate subscription models to make their content available in formats suitable for text and data mining (for example, through OAI or FTP protocols).
39. Ultimately, however, this issue comes down to a question of how far the rights of publishers should reasonably extend. We recognise fully that publishers play an important role in the research process: they have the right to charge a reasonable price for the services they provide and these costs need to be met, whether that is through author fees or subscription charges. However, the contribution they make pales in comparison to that of research funders and higher education institutions⁴. It is totally unacceptable therefore that publishers should be able to control use of the outputs of research that has been supported by the taxpayer and by charities, and essentially limit the benefits for society that can result. We would emphasise once again that this is about enabling the use of material to which researchers already have lawful access, and hence for which the publisher has already received due compensation.
40. In short, we believe that the copyright system must aim to promote innovation and serve the best interests of society, and not simply serve to grant publishers a monopoly to generate further profits from research that has been funded for the public good. An exception for text and data mining for non-commercial purposes is crucial for this reason.

Q103 What are the advantages and disadvantages of allowing copyright exceptions to be overridden by contracts? Can you provide evidence of the costs or benefits of introducing a contract-override clause of the type described above?

41. It is clear that to have any real value, the exceptions that are included in copyright law cannot simply be circumvented through the use of contract terms. Therefore, alongside the introduction of these key exceptions, we believe that it is essential that the law is amended to prevent such exceptions being overridden by contract.

References

1. Analysis of the value and benefit of text mining and text analytics in UK further and higher education (JISC 2012)
2. Smit E, and van der Graff, M. Journal article mining: the scholarly publishers' perspective. *Learned Publishing*, 2012, Vol 25 (1) 35-46)
3. JISC Collections, IPO Consultation: response from JISC Collections, March 2012
4. Activities, costs and funding flows in the scholarly communications system (RIN 2008)

The Wellcome Trust is a global charitable foundation dedicated to achieving extraordinary improvements in human and animal health. We support the brightest minds in biomedical research and the medical humanities. Our breadth of support includes public engagement, education and the application of research to improve health. We are independent of both political and commercial interests