Appendix A

International baseline of text mining and related activities

This appendix presents the outputs of the international baseline of text mining and related activities. This includes an overview of the position on copyright exceptions across a number of developed countries, including some key G20 countries as well as additional Scandinavian examples. It is accompanied by Appendix B: Copyright Baseline Comparisons Table and Appendix C: Innovative Country Comparison Table.

It has been suggested that the UK may be at a disadvantage in the global economy if restrictions on the use of text mining technology prevent its exploitation and application in the UK. This would be particularly the case if other countries (such as the USA or Japan) have a more liberal approach to copyright as it relates to text mining; companies pursuing the applications of text mining would be likely to move their operations to such countries.²

Therefore, as part of this project, the brief included a requirement to assess and compare the situation on copyright exceptions across other leading developed nations and competitor economies, as far as this was possible. This is presented below, with a brief consideration of comparator international copyright exception data followed by comparative international innovation data.

A1: Comparator International Copyright Exception Data

Copyright is a complex area, which is governed by a number of key international conventions and treaties, including the Berne Convention [99], the WIPO Copyright Treaty (1996) [100] and the TRIPS (Trade Agreement on Aspects of Intellectual Property Rights) 1994 [101]. These conventions and treaties made provision for a number of exceptions and limitations to copyright. In particular, the Berne Convention ‘3 step test’ defined three principles against which any potential copyright limitation or exception should be judged. These were that any copyright exception should:

(i) Be limited to special cases
(ii) Not conflict with a normal exploitation of the work, and
(iii) Not unreasonably prejudice the legitimate interest of the author

The 3-step test is now a common international ‘benchmark’ for proposed copyright limitations or exceptions, against which any proposals for exceptions should be tested. However, the 3-step test notwithstanding, exceptions can be treated and applied differently in different countries depending on the cultural and legal tradition of that country. In the EU, the Information Society Directive (2001)[102] sought to promote harmonisation of copyright across member states. However, there remains, even within the EU, a range of differences in the relevant copyright legislation and in its interpretation and application and nuance.

² The case study of Mendeley (section 4) is a good example, showing that they are considering a need to base part of their operations elsewhere as a result of copyright restrictions in the UK.
Recent studies of copyright legislation across the globe, conducted by the World Intellectual Property Organisation, have described the position as ‘fragmented’. [103]

As part of this project we sought to take an overview of copyright exceptions. The most up to date information on the position of copyright exceptions and limitations appears that which is available through the World Intellectual Property Organisation (WIPO) Standing Committee on Copyright and Related Rights who conducted a 2010 survey across all member states to ascertain individual country positions regarding limitations and exceptions [103]. Replies were received from 61 member states; 103 questions were included in the survey.

While text mining technology has been in existence for some time, the application of text mining tools to scholarly journals and related copyright material appears relatively recent and therefore the full impact of copyright restrictions on text mining usage is still emerging. This means that in the overview of copyright exceptions there is as yet little evidence of explicit allowance for, or reference to, text mining or related activity. Of the 103 questions asked in the WIPO survey, none of these related specifically to exceptions that would enable text mining.

This may tend to give a competitive advantage to those countries that have a more liberal or flexible approach to copyright (such as those with a ‘Fair Use’ approach such as the USA [104]), which could enable text mining usage in non-commercial research to take place under a ‘Fair Use’ defence rather than needing explicit permissions. In an overview of copyright exceptions globally, only one example was found of explicit reference to text mining, and it is worth noting that this example comes from one of the leading innovation countries and a major UK competitor – Japan. The Japan Copyright Act (2011) [105] makes explicit provision to allow text mining, with Article 47 making a limitation to copyright:

“For the purpose of information analysis (‘information analysis’ means to extract information, concerned with languages, sounds, images or other elements constituting such information, from many works or other much information, and to make a comparison, a classification or other statistical analysis of such information; the same shall apply hereinafter in this Article) by using a computer, it shall be permissible to make recording on a memory, or to make adaptation (including a recording of a derivative work created by such adaptation), of a work, to the extent deemed necessary’ [105].

While the WIPO survey cannot provide information explicitly related to text mining, the survey gives some insight into the current international comparator position on copyright exceptions and the issues being raised by different countries. For instance, it is interesting to note that many countries are still seeking to consider how to deal with copyright issues relating to Distance Learning; given that the technologies involved and their applications are now widespread and relatively mature, it is not surprising that text mining – the potential of which is just beginning to be more widely realised – does not explicitly feature as an issue in this survey and related WIPO studies. The responses to selected questions for the G20, as well as some specific European economies (eg Norway, which is not a member of the G20) are presented in the Baseline Comparison table, which is included in this report as Appendix B.
The data selected for inclusion in the comparison table relate to the questions that were considered most likely to have a bearing on activity that could be related to text mining (e.g., if there are exceptions for education or research purposes, or exceptions relating to digital networks, if other law is permitted to override copyright law or the country’s position in relation to Digital Rights Management). It also indicates whether the country has an ‘open’ approach to copyright exceptions (such as ‘Fair Use’) or if (as in the UK) exceptions are more tightly defined or ‘specific’, which in some cases is seen as limiting the capacity for a flexible response to technological change.

**A2: Comparative International Innovation Data**

For information, a second comparison table Appendix C contains selected data on a range of ‘innovation’ indicators, which have been compiled for NESTA (the UK National Endowment for Science, Technology & the Arts), drawing on both UK and international data. There is a wide range of international data sources that can be used to reflect aspects of a country’s economic composition and standing. These include, for example, the OECD Science and Technology Scoreboard and the Community Innovation Survey as well as data from statistical agencies such as Eurostat. NESTA has been leading the development in the UK of an ‘innovation index’ to enable analysis of the UK position in terms of its innovation capacity. A 2011 NESTA Innovation Index report [15] reviewed and assessed the wide range of international data available and drew on a number of sources (including the OECD, Eurostat etc) to compile sets of indicators that show the UK’s relative standing in its innovation capacity and performance. A number of these indicators have been included in the baseline table because one of the key points of discussion around the use of text mining is that it is a tool that has the potential to support innovation, and enabling its use would provide new opportunities for discovery and learning that would generate related economic and social gains.

This second comparison table includes only (a) those countries that were identified as ‘leading innovation nations’ in the BIS economics analysis [88] underpinning the Innovation and Research Strategy for Growth, i.e., the USA, Japan, Sweden and Germany, together with (b) Finland, the Netherlands and Norway, as these countries featured strongly in the NESTA innovation report.

There is currently insufficient data to draw strong inferences regarding the impact of copyright law restrictions on innovation or to make links between the copyright position and a country’s economic performance. This could be a fruitful field for further study. Indeed, a number of researchers have commented that there is insufficient empirical data overall in regard to the impact of copyright on innovation – including a lack of evidence as to where copyright supports innovation.[106] New literature investigating evidence from China suggests that there is potentially greater economic gain to be had from a relaxation of copyright. [107]

However, the evidence in the present study shows that if the current situation or ‘status quo’ in the UK regarding copyright law – insofar as it affects text mining of scholarly journals and non-commercial research – is maintained, the UK will not be able to take full advantage of its considerable past and ongoing investment in the public research base. As the public research base is one of the key framework factors in supporting the country’s innovation
capacity [15], being unable to make full use of the research base would clearly be to the UK’s economic disadvantage.