

# Google Library: Why all the fuss?

---

Ruth Allen

## Introduction

Google Book Search was launched by the internet search engine company Google Inc (“Google”) in October of 2004 at the Frankfurt Book Fair.<sup>1</sup> The aim of Google Book Search is to digitise published works and allow searching of their contents online, providing access for publishers to more customers and greater access for the public to such works.

In December 2004 Google announced its latest endeavour, the Books Library Project, an extension of Google Book Search.<sup>2</sup> Google Book Search already consisted of the Books Partner Program, to which a number of publishers had already signed at the time the Books Library Project was announced.<sup>3</sup>

Publishers and authors have reacted to the Books Library Project with outrage, with two lawsuits being filed against Google in the latter half of 2005. These groups take the view that creating the project database involves massive infringement of copyright.

Google argues that its actions are justified by the fair use defence under US law.<sup>4</sup> This article examines the context in which the controversy has arisen, including an overview of Google Book Search, its extension to books in libraries, and an analysis of the issues and arguments that are being put forward by both sides. Alternative models being developed by rival companies are also outlined.

## History of the Google Print controversy

Several developments occurred in 2005 in response to the announcement of the Books Library Project. In May 2005 the Association of American University Presses (AAUP) wrote to Google, expressing its members’ concerns at the implications of the Books Library Project. The letter also posed a number of questions concerning the operation of the project and Google’s views on the application of copyright law to the project.<sup>5</sup> Firstly, the AAUP questioned whether the fair use defence could apply to reproduction on the scale carried out under the Books Library Project, both in relation to the number of works and of different classes of work.<sup>6</sup> The letter also raises concerns about the potential for misuse and future direct exploitation, and asks how Google plans to safeguard the information stored in its “dark archive”.<sup>7</sup> A possibility that worries authors and publishers is that Google might decide to make the service subscription-based without giving equitable remuneration to copyright owners in the scanned works.

In June 2005, the contract between Google and the University of Michigan (one of the libraries participating in the Books Library Project) was released following a request under Michigan’s Freedom of Information legislation. The contract makes clear two digital copies of each book are made, one copy for Google and another for the university.<sup>8</sup>

In August 2005, Google agreed to halt the process of scanning works that were still subject to copyright until the beginning of November 2005, in order to give publishers time to “opt out” of the Books Library Project. However, this was only a temporary measure and the scanning of works in which copyright still subsists has now resumed.<sup>9</sup>

In September and October 2005 the US Authors’ Guild and the Association of American Publishers filed suits against Google in relation to the Books Library Project. The suit filed by the Authors’ Guild included several named authors and was lodged as a class action on 20 September 2005 in the United States District Court Southern District of New York.<sup>10</sup> The Publishers’ suit was lodged on 19 October 2005 in the same court.<sup>11</sup> Both sets of plaintiffs seek injunctive relief against Google for infringement of their copyrights. The Authors’ Guild also seeks an award of damages or an account of profits.<sup>12</sup>

Google Book Search has also been contentious at an international level. The President of France and the head of the French National Library have argued that the project is too focused on the English language and Anglo-Saxon and American thinking, and that this will increase Anglo-American cultural dominance on the world stage.<sup>13</sup>

### **How the Google Book Search project works**

Rather than being provided as a subscription service, Google Book Search generates profits for Google through the sale of advertising space on its search pages. While Google agrees to pay a share of these profits to publishers under the Books Partner Program contracts, no remuneration to copyright owners is planned with regard to the profits derived from the Books Library Project.

Under the Books Partner Program, the full text of the authorised work is scanned, digitised and stored in a database created by Google. Users of the service will then be able to view the page on which their search query appears and a page on either side of it, and to click on links to sites where they can purchase the book. Another link directs users to the website of the publisher whose book they are viewing.

Publishers may elect to allow advertisements to be included in the page that displays the result of the search. If they do so, they will be entitled to receive a proportion of Google’s profits from the advertisements. Google generates an estimated 98 per cent of its multibillion-dollar revenues through advertising revenues.<sup>14</sup>

### **The Books Library Project**

The Books Library Project involves digitisation and online use of books without permission from the copyright owners. There are five participating libraries involved in this aspect of the project, four of which are university libraries.<sup>15</sup>

Of the participating libraries, only the American university libraries are allowing works in which copyright still subsists to be scanned into the project’s database. The project began with the University of Michigan library, which is named in the suit filed by the Author’s Guild.<sup>16</sup>

Under the Books Library Project, works in the collections of the participating libraries are scanned, digitised and stored in the database system set up by Google.<sup>17</sup> Users of the service are able to search the database and retrieve parts or all of the work matching their search. How much of the work can be accessed depends on whether or not the work that is the subject of the query

is still in copyright. For those works that are no longer in copyright the user may access the full text. For those works that are still subject to copyright, only “snippets” will be accessible, along with the bibliographic information of the work. The term “snippets” has been defined by Google to mean a couple of sentences either side of the location of the search query.<sup>18</sup> Users will only be able to view three instances of where their keyword search appears, not every instance. Users will also be able to access reviews of the works.

Publishers can direct that certain works not be scanned and digitised as part of this project. Google has stated that where copyright owners wish to exclude works that have already been scanned into the database from the Print Library Project, it will remove the works.<sup>19</sup>

## Legal arguments

Copyright owners have argued that by reproducing their copyright works in digital form without permission, Google is infringing their copyright. They argue that the “opt-out” policy announced by Google does not rectify the problem, since copyright law requires that someone wanting to use the material must get permission; there is no principle that allows this obligation to be reversed. In the words of the CEO of the Association of American Publishers:

Google’s procedure shifts the responsibility for preventing infringement to the copyright owner rather than the user, turning every principle of copyright law on its ear.<sup>20</sup>

Google’s supporters have argued that this situation arises not from Google’s actions but from the nature of the internet.<sup>21</sup>

The principal arguments that have been raised by Google and its supporters concern the “de minimis” and “fair use” principles under US law. It has also been argued that Google’s use of works is akin to the commonly accepted practice of indexing in libraries, despite the fact that indexing does not require reproduction of the work being indexed.<sup>22</sup>

### *The de minimis principle*

The *de minimis* principle requires that for an action to amount to *prima facie* infringement of the rights of the copyright owner it must amount to more than an inconsequential and insubstantial use. Determining whether an action falls within the *de minimis* principle requires a subjective review of the relevant facts. However, in this respect the end use of the material does play a part.<sup>23</sup>

One instance where the *de minimis* principle was held to apply to an alleged breach of copyright was where a copy was made but not used.<sup>24</sup> One factor to be taken into account in determining whether an alleged copyright infringement is *de minimis* is the recognisability of the portion which is copied.

The application of the principle is generally confined to instances that are deemed too insignificant for the courts to deal with. It may be that the *de minimis* principle could apply to the reproduction and communication of “snippets” in response to a searcher’s enquiry, at least in some cases.<sup>25</sup> However, the fact that entire books are being copied and that the scale of the project is immense, is likely to take Google’s actions outside the scope of the *de minimis* argument.

### ***“Fair use”***

The main argument presented in defence of Google’s actions is that it falls within the fair use defence. Four factors are to be taken into consideration when determining whether a fair use defence is applicable to a given instance. These are:

- (1) the purpose and character of the use, including whether such use is of a commercial nature or is for non-profit educational purposes;
- (2) the nature of the copyrighted work;
- (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and
- (4) the effect of the use upon the potential market for, or value of, the copyrighted work.<sup>26</sup>

The leading case which analyses these principles in relation to the actions of internet search engines is *Kelly v Arriba Soft Corporation*.<sup>27</sup> Arriba created a program that crawled the internet, found images, copied them and transformed them into low-resolution thumbnails for inclusion in the company’s database. In this case, the Ninth Circuit of the US Federal Court held that the fair use defence applied.

Whether or not the actions of Google fall within the parameters of the fair use defence depends on an analysis of all four factors in the context of the decided case law, in particular *Kelly*.<sup>28</sup>

### ***The purpose and character of the use***

As in *Kelly*, Google’s actions can be properly classified as commercial in nature. However, a major point of difference is that Google is not creating a database from material already on the internet but digitising and uploading the material itself. Another point of distinction is that Arriba was held to have made “transformative use” of the images in its database. By contrast, Google is making digital copies and retaining them. Thus in some respects the first factor in the *Kelly* analysis may weigh slightly against the actions of Google.

### ***The nature of the copyrighted work***

The Ninth Circuit Court in *Kelly* stated that:

“[w]orks that are creative in nature are closer to the core of intended copyright protection than are more fact-based works.”<sup>29</sup>

One of the issues raised in the letter of the AAUP to Google concerned the approach to applying the fair use defence to actions covering many different categories of work: yet within a single library collection works range from, for example, dictionaries to works of fiction to art books to musical scores.<sup>30</sup> It may therefore be difficult to work out how to apply this factor to reproductions of whole library collections in this context.

### ***The amount and substantiality of the portion used***

Statements by Google and its supporters tend to focus on the small amounts of digitised works that will be made available to someone searching the database. However, copyright owners

argue that the making of the digital copies in the first place infringes copyright. While the decision in *Kelly* allowed initial copying of entire works due to the character of the end use, Google's activities are distinguishable from those of Arriba. In this context, it is worth noting that Google is making extra copies of the digitised material for the use of the libraries holding the collections being used.

### *The effect of the use upon the potential market value of the copyrighted work*

This factor concerns the overall impact on the ability of the copyright owner to exploit their intellectual property. Google has argued that the Books Library Project may in fact increase the opportunities for copyright owners to extend their market.

However, this factor requires assessment of the long-term effect of the alleged infringement as well as any short-term effects. As has already been noted, the AAUP fears that at a future time Google's dark archives may be open to misuse or a different *modus operandi*. Were this fear to be realized, the market value of the copyrighted work could very well diminish and restrict the means by which the copyright owners can exploit their work.

Also, in creating the Books Library Project, Google creates and circulates digital copies made from traditional print copies. This limits the ability of the copyright owners to make decisions about digital reproductions of their work.

### *Conclusion on fair use*

While Google has argued that it can rely on the fair use defence, it is not clear that it will succeed. Further, in view of the courts' decisions in cases like *Napster*<sup>31</sup> and *Grokster*,<sup>32</sup> its reliance on the "opt-out" scheme may not be viewed favourably. These cases make it clear that copyright owners do not waive their rights by failing to exploit them in particular ways, and do not bear the onus of preventing infringements.<sup>33</sup>

### **Similar projects**

A number of projects similar to Books Library Project have been announced or have begun operation. Late in 2005 Microsoft announced a competing service in partnership with the British Library, to digitise 100,000 out-of-copyright works.

The internet book company Amazon.com has announced that it will allow its users to purchase segments of works and view them electronically on-line. Copyright holders will receive royalties for the exploitation of their works.

On 30 September 2005 the European Commission announced a plan to produce a rival to the Books Library Project and create a digital library of European historic and cultural heritage.<sup>34</sup> The details of this project are yet to be finalised. However, the project is expected to see the digitisation of books, film fragments, photographs, manuscripts, speeches and music.

Finally, in late 2005, Yahoo announced its participation in the Open Contents Alliance.<sup>35</sup> This Project is to see the digitisation of 18,000 works that should be accessible by the end of 2006. This project uses an opt-in policy for works that are still subject to copyright. Material to be digitised includes books, speeches, audio, video and music.

It is clear that full text searches created as part of the Books Library Project will continue to be available online for the time being, and will cover an increasing range of material. It also appears from the diverging legal arguments that unless a settlement can be reached between the parties, the final chapter of the Books Library Project is still a long way from being written.

## Endnotes

- 1 Google Book Search was previously known as the “Google Print Project”.
- 2 Books Library Project was previously known as the “Print Library Project”.
- 3 Books Partner Program was previously known as the “Print Publisher Program”.
- 4 For a discussion and analysis of the fair use doctrine see [www.copyright.gov](http://www.copyright.gov). The Australian Copyright Council has also published a book on this principle from an Australian perspective: Ian McDonald, *Fair Use: Issues & Perspectives*, Sydney, 2006.
- 5 For the full letter from the AAUP to Google see [www.aaupnet.org/aboutup/issues/gprint.html](http://www.aaupnet.org/aboutup/issues/gprint.html).
- 6 This argument was raised especially in questions 2, 3 and 4 of the AAUP letter; see [www.aaupnet.org/aboutup/issues/gprint.html](http://www.aaupnet.org/aboutup/issues/gprint.html).
- 7 “Dark archive” is a term used in reference to data storage. It is a type of archive in which access is either limited to set individuals or completely restricted to all. Its purpose is to serve as a repository for data that can be used as a failsafe during disaster recovery. The argument in relation to the Google “dark archive” was addressed in questions 10, 11 and 12 of the AAUP letter; see [www.aaupnet.org/aboutup/issues/gprint.html](http://www.aaupnet.org/aboutup/issues/gprint.html).
- 8 It is not known whether the same goes for other participating institutions under their contracts.
- 9 See [www.iwr.co.uk/information-world-review/news/2145276/google-library-scanners-back](http://www.iwr.co.uk/information-world-review/news/2145276/google-library-scanners-back).
- 10 *The Author’s Guild v. Google, Inc.*, 2005 WL 2463899 (S.D.N.Y.) (No. 05-CV8136). See [www.linksandlaw.com/news-update34-court-document-authors-guild.htm](http://www.linksandlaw.com/news-update34-court-document-authors-guild.htm).
- 11 To view the Association of American Publishers’ claim against Google lodged in October 2005, see [www.publishers.org/press/releases.cfm?PressReleaseArticleID=292](http://www.publishers.org/press/releases.cfm?PressReleaseArticleID=292).
- 12 For full details of the relief sought by the Author’s Guild, see [www.linksandlaw.com/news-update34-court-document-authors-guild.htm](http://www.linksandlaw.com/news-update34-court-document-authors-guild.htm).
- 13 See [news.bbc.co.uk/2/hi/technology/4512831.stm](http://news.bbc.co.uk/2/hi/technology/4512831.stm).
- 14 See paragraph 17 of the Author’s Guild claim at [www.linksandlaw.com/news-update34-court-document-authors-guild.htm](http://www.linksandlaw.com/news-update34-court-document-authors-guild.htm).
- 15 The participating Libraries are Harvard University Library, the University of Michigan Library, Stanford University Library, Oxford Bodleian Library and the New York Public Library.
- 16 See in particular paragraph 31 of the Author’s Guild claim at [www.linksandlaw.com/news-update34-court-document-authors-guild.htm](http://www.linksandlaw.com/news-update34-court-document-authors-guild.htm).
- 17 The digitised material is stored in a dark archive.
- 18 See [print.google.com/googlebooks/common.html#2](http://print.google.com/googlebooks/common.html#2).
- 19 For further details on the Books Partner Program and Books Library Project see [books.google.com/googlebooks/publisher.html](http://books.google.com/googlebooks/publisher.html).
- 20 Quoted in Jonathan Band, *The Google Print Library Project: A Copyright Analysis* at p. 2. See [www.policybandwidth.com/publications.htm](http://www.policybandwidth.com/publications.htm).
- 21 Lawrence Lessig, “Google’s Tough Call”, *Wired Magazine*, Issue 13.11 November 2005.
- 22 Lawrence Lessig, “Google’s Tough Call”, *Wired Magazine*, Issue 13.11 November 2005.
- 23 Although the application of this principle to US copyright law has some similarities to the Australian concept of a “substantial part”, it is far from an exact equivalent. For a discussion of the principle, see Melville B Nimmer and David Nimmer, “Nimmer on Copyright”, (Matthew Bender, New York), 2:8.01[G].
- 24 *Knickerbocker Toy Co v Azrak-Hamway International Inc.*, 668 F.2d 699 (2d Cir. 1982)
- 25 Note, also that a “snippet” could be a substantial part of a work such as a poem, song lyrics or an artistic work. However, Nimmer comments that “among the several potential meanings of the term de minimis, that defence should be limited largely to its role in determining either substantial similarity or fair use”: Melville B Nimmer and David Nimmer, “Nimmer on Copyright”, (Matthew Bender, New York), 2:8.01[G], 8-25.
- 26 17 U.S.C § 107.
- 27 336 F.3d 811 (9th Cir. 2003).
- 28 However, it should be noted that that decision, although persuasive, was decided in the Ninth Circuit and therefore is not binding on the court deciding the Google case, which has been filed in the Second Circuit.

- 29 See *Kelly* at 820.
- 30 See question 3 of the AAUP letter, at [www.aaupnet.org/aboutup/issues/gprint.html](http://www.aaupnet.org/aboutup/issues/gprint.html).
- 31 *A&M Records Inc v Napster* (2001) 50 IPR 232.
- 32 *Metro-Goldwyn-Mayer Studios Inc and others v Grokster Ltd and another* (2005) 3 IPR 645.
- 33 A federal court in Nevada ruled on 19 January 2006 that the Google Cache feature does not infringe copyright. Summary judgement in favour of Google was awarded on the basis that serving a webpage from the Google Cache does not directly infringe copyright as it results from the automated activity of Google servers, and that the Google Cache feature falls within the scope of fair use and the safe harbour provisions. See [www.eff.org/IP/blake\\_v\\_google/google\\_nevada\\_order.pdf](http://www.eff.org/IP/blake_v_google/google_nevada_order.pdf).
- 34 See [europa.eu.int/rapid/pressReleasesAction.do?reference=IP/05/1202&format=HTML&aged=0&language=EN&guiLanguage=fr](http://europa.eu.int/rapid/pressReleasesAction.do?reference=IP/05/1202&format=HTML&aged=0&language=EN&guiLanguage=fr).
- 35 This development was reported by the BBC; see [news.bbc.co.uk/1/hi/technology/4304192.stm](http://news.bbc.co.uk/1/hi/technology/4304192.stm).