

Conference Presentation
Unlocking IP Conference “National and Global Dimensions of the Public Domain”
University of New South Wales 17 April 2009

By Sally McCausland^φ

**‘Googling the archives:
– ideas from the Google Books Settlement on solving the orphan works problem
in digital archive projects**

*[*this paper is a draft subject to further revision before publication. Please contact the author for quote verification]*

Abstract

The Google Books Settlement is the provisional settlement, reached in late 2008, of a copyright infringement class action brought against Google by the American Authors Guild and the Association of American Publishers. The case concerned the legality of Google’s “Library Project” through which, in partnership with various US libraries, Google digitised vast quantities of books without copyright permission and made searchable snippets of them available online. Claiming that most of the copyright books it digitised were out of print and not otherwise widely accessible, Google famously defended itself from claims of copyright infringement by citing “fair use” under US Copyright law.

While some observers have criticised the settlement for its essentially commercial nature and in particular, its potential anti-competitive effects, its undeniable effect is to increase public and academic online access to in copyright, out of print books. In this sense it is a rare breakthrough for online access to copyright archival material. This paper considers whether this model, or one very roughly based on it, could be applied to better facilitate public online access to other forms of cultural archival material such as broadcast archives. In particular, it considers whether new forms of statutory licensing could be introduced along the lines of an opt-out scheme.

1. Introduction

1.1 The legal problem with putting twentieth century archives online

“Perhaps the single most important feature of the digital revolution is that for the first time since the Library of Alexandria, it is feasible to imagine constructing archives that hold all culture produced or distributed publicly...The scale of this potential archive is something we’ve never imagined before...but we are for the first time at a point where that dream is possible..”¹

^φ Senior Lawyer, Special Broadcasting Service (SBS) and board member, Audio-Visual Copyright Society Ltd (Screenrights). The views expressed in this paper are personal and do not represent those of SBS or Screenrights.

¹ Lawrence Lessig, *Free Culture*, ch 9, available at www.authorama.com/free-culture-9.html

Digital technologies now offer opportunities for searching and accessing cultural archives on an immense scale. Governments, commercial enterprises and cultural institutions worldwide are testing the possibilities. But copyright has emerged as a major barrier, at least in relation to the creative output of the twentieth century. Most works made in the twentieth century, and particularly in the second half, remain under copyright. The international Berne Convention copyright treaty, to which most countries are now signatories, is based on the key principles of automatic protection and absolute discretionary permission. In other words, the default settings of the international copyright system are that copyright owners need not register their creations to be protected, and that any new (say – digital) use of their creations requires the owner’s individual permission. In the case of “orphan works”, where an owner cannot be found, the risk of using the work remains with the user.

In the era of mass digitisation the effect of this copyright system is the “gridlocking”² of vast areas of cultural knowledge and production. With the gradual expansion of copyright law virtually everything, from books to bus timetables, is now copyright for terms which may now far exceed 100 years. The need to find and seek permission from each copyright owner before a new use of the archival material can confidently be made adds significant cost and risk to online archive projects. The problem of orphan works is a particularly vexing issue, and one which is currently the subject of various legislative schemes and policy proposals in different jurisdictions.

Copyright barriers to online use are particularly high in relation to recorded media such as television archives, which have multiple layers of copyright owners, and which have been neglected by legal deposit schemes.³ The complex underlying copyright interests in many television programs, and the lack of any searchable ownership register, universal identification system, or comprehensive public catalogues, can make it very difficult to research, commercially licence or repurpose old television programs. Ironically, it is now easier to search and access cultural material created in the nineteenth century than it is the twentieth:

“How is it that we have created a world where researchers trying to understand the effects of media on nineteenth-century America will have an easier time than researchers trying to understand the effect of media on twentieth-century America?”⁴

Orphan works are a significant part of this conundrum.

1.2 The particular problem of orphan works in mass digitisation projects

In this paper I focus on the problem of orphan works in mass digitisation projects – being projects where a large amount of archival material is digitised for online access. Access may be made free for subsidised public interest purposes, or be part of a business model based on self-generated or aggregated content offerings and

² The term references Michael Heller’s book, *The Gridlock Economy*.

³ For a detailed study of the difficulties of clearing public broadcaster archives for online access, see Sally McCausland, “Getting public broadcaster archives online: orphan works and other copyright challenges of clearing old (but in copyright) cultural material for digital use” (forthcoming, *Media Arts Law Review*).

⁴ Lawrence Lessig, *Free Culture*, ch 13.

remunerated via subscription, advertising revenue or other means. The largest of these, of course, is the Google Library Project which to date has digitised some seven million library books. Others, such as the European Film Gateway project⁵ and projects undertaken by some public broadcasters,⁶ are relatively modest, but still of significant scale.

From the mass digital user perspective, orphan works are a grinding resource and cost issue. The administrative time consumed by searching for orphan works owners, and eventually negotiating with those found, eats up budgets and staff time. Those who can't be found present an unquantified risk of being sued – itself a lengthy, expensive and unpalatable process. This inflates transactional costs, and thus inhibits the development of new content access models in the digital environment. All of this is upfront time and cost, before use has even commenced, and even where the use is not commercial. Many interesting digital archival projects fail to launch once the copyright implications loom.

For these reasons, the most favoured legislative approach to the orphan works issue to date – the “diligent search” model is – at least from a mass digital user perspective – largely unworkable. The diligent search model requires the would-be user to attempt to locate a copyright owner by carrying out certain prescribed or “reasonable” activities such as searching industry registers and publishing notices of intended use. Delay is inherent, particularly if an application for use must be made to a regulatory body. Depending on the particular scheme, the reward for going through this process will either be a licence to use the material (if granted by the regulatory body),⁷ or some form of reduced risk of damages or injunction if the owner subsequently emerges and objects to the use.

The “reduced risk” approach is the model currently being debated in the United States.⁸ The two orphan works bills currently awaiting legislative fiat are the result of a lengthy public submissions and hearings process, and remain controversial. Under this draft legislation, a user who makes reasonable efforts to locate the owner of an orphan work in accordance with the relevant criteria will be entitled to certain legal immunities against copyright remedies and legal costs if the orphan work owner later comes forward and objects to the use. But this still means that for every orphan work search, significant resources and time are required. As one commentator puts it, these bills:

⁵This project is a consortium of European film archives seeking to give public access to their archives online. See <http://www.europeanfilmgateway.eu/>. Orphan works are a major legal challenge to the project:

http://www.europeanfilmgateway.eu/downloads/EFG_MinervaJerusalem_2008.ppt#268,11,Metadata Interoperability

⁶ Such as those undertaken by the British Broadcasting Corporation (BBC) and Japan Broadcasting Corporation (NHK), each of which involved online clearance of 1,000 hours of archival programming: see McCausland, id.

⁷ This form of scheme exists in Japan and Canada. For a review of these various schemes see Ian McDonald, *Some Thoughts on Orphan Works* (2007) Copyright Reporter 180.

⁸ *Orphan Works Act of 2008* (Bill no HR 5889) and *Shawn Bentley Orphan Works Act of 2008* (Bill no S. 2913). The bills can be found at www.copyright.gov/orphan. For an overview see Australian Copyright Council, Information Sheet G101, 1 December 2008, available at www.copyright.org.au.

... “conceive of orphans as adoptable on a case-by-case basis only, not at scale, with all the in-depth investigation that the analogy to adoption suggests.”⁹...

Even once the searches have been carried out, protection is not absolute. The copyright owner who emerges and objects to the use is still entitled to seek damages unless “reasonable” compensation is negotiated. The potential for dispute under such a complex model is obvious.

Diligent search models may work reasonably well in the case of one off “adoptions” where the user is a well resourced entity which is not in a rush, but they are clearly a compromise of interests which do not adequately address the needs of users in the digital environment. Nor do they address the concerns of many copyright owners who perceive unfair compromise to the twin bedrocks of automatic protection and absolute discretionary permissions afforded them under the general rules of the copyright system.¹⁰

The US orphan bills remain in limbo. Yet meanwhile something much more revolutionary may be about to happen to the digital world. As the US orphan works legislation was going through its lengthy public debate process, the Google Books Settlement was being negotiated in private. The Google Books Settlement’s approach to orphan works is akin to dropping an atom bomb on the copyright system. It makes the US bills look positively coddling of copyright owners. The Google Books Settlement has been described by one commentator as “*an elaborate scheme for the exploitation of orphan works*”.¹¹

This paper evaluates the orphan works provisions of the Google Books Settlement, as they await sanctification by the court after fairness hearings due in June 2009. These provisions – if endorsed by the court – will create what one commentator has called a “magic device” for solving orphan works issues. However, this device will benefit only Google – and give it a huge competitive advantage in a field it already dominates.

I argue that whatever its particular flaws, the model set up by the Google Books Settlement is a useful template for thinking more generally about the needs of mass archive digitisation projects and ultimately, the copyright balance in the digital environment. Legislatures who proactively solve these orphan works issues will boost their country’s cultural output and give it a competitive advantage in the digital economy. We should therefore think about whether Google type “magic devices” can be made accessible to other digital users, more beneficial to the public interest and more palatable to copyright owners. Existing statutory licensing schemes – particularly those in Australia – provide a useful starting point for adaptation of these ideas.

2.1 The Google Books Library Project

⁹ Georgia Harper, “Mass Digitisation”, blog post 31 May 2008 <http://blogs.tdl.org/digitize/2008/05/31/google-book-search-2/>

¹⁰ See eg

¹¹ Robert Brewster, “It’s All About the Orphans”, blog post 23 February 2009, <http://www.opencontentalliance.org/2009/02/23/its-all-about-the-orphans/>

Google has built its fortune on supplying the world's most competitive free internet search engine, supported by online advertising sales. Google's corporate mission is to "*organise the world's information and make it universally accessible and useful.*"

Some time in the early 2000s Google began its Google Book Search project, through which it licenced from numerous publishers the right to digitise and index books for online searching by the public:

"Our agreements for those books allow us to display the page containing the searched-for term and a few surrounding pages. We also show links to enable readers to buy the book and we may show advertising related to the content of the pages, from which the publisher can receive additional revenue".¹²

Reportedly there are around a million licensed books in the Google Book Search index.¹³

In 2004 Google announced an expansion to the project in the form of the Google Book Library Project. The cornerstone was a partnership with several prominent libraries including Harvard, Oxford and the New York Public Library, who granted physical access to their collections for Google to scan. In exchange for making digital copies of scanned books available back to the participating libraries (and presumably giving the libraries fulsome indemnities) Google was permitted to make available online searchable versions of books in the libraries' collections. Out of copyright, or public domain, books were made available online in full. In-copyright books were also scanned. These could not be fully accessed but online viewers could see searchable portions or snippets "*to help people decide whether to buy the book or look for it in the library*".¹⁴

The Google Library Books Project differed from the earlier version of Google Books because Google had permission to access physical copies of the books, but not copyright permission to scan those still in copyright. Google did not seek prior permission from the copyright owners of these in-copyright books. Rather, it claimed that permission was not needed because its activities were covered by the fair use doctrine of United States copyright law.¹⁵ Nevertheless, it stated that it would continue to offer agreements to those publishers who came forward to become partners in the Book Search program.

As noted above it is estimated that the total number of books now digitised by Google and included in its Book Search index is around seven million. Of these, it is reported that one million are licensed, one million are public domain, and the remainder are in copyright but allegedly out of print.¹⁶ Google claims that these out of print, in copyright books are the "long tail" of niche product which may potentially be revived

¹² Google Inc, *Response to Notice of Inquiry Regarding Orphan Works*, submission dated 25 March 2005 to United States Orphan Works Inquiry, US Copyright Office, available at www.copyright.org

¹³ Juan Perez, "Google Book Settlement, Business Trumps Ideals", *PC World*, 30 October 2008.

¹⁴ Google submission to US Orphan Works Inquiry, above.

¹⁵ See further Ruth Allen, "Google Library: Why all the Fuss?" (2005) 23(4) *Copyright Reporter* 105.

¹⁶ Robert Darnton, "Google and the Future of Books" (2009) 56(2) *New York Review of Books*, available at <http://www.nybooks.com/articles/22281>.

by their inclusion in Google Book Search. Many of these, Google claims, are orphan works. In Google's view:

*“Orphan works represent an untapped wealth of information that can and should be made accessible to the public.”*¹⁷

2.2 The Google class action and the Google Books Settlement

In 2005 the American Authors Guild and the Association of American Publishers commenced United States class copyright infringement actions¹⁸ against Google. Google maintained its defence of fair use. The merits of the litigation have received abundant academic attention, and it is not proposed to speculate here about whether or not Google would have established its defence. On 28 October 2008, the parties announced that they had reached a provisional settlement, the Google Books Settlement, subject to the ratification of the court.

The Google Books Settlement has the following main elements:

- Google must pay a large cash sum to be distributed among members of the class whose books were digitised before 5 May 2009 and who opt to stay in the settlement and register for their share of compensation by January 2010 (around \$60 for each book digitised);
- owners who stay in the Settlement will be entitled to receive ongoing revenue share from advertising associated with the indexing and display of portions of their books in Book Search and through educational subscriptions and other uses;
- Google can propose new revenue models in future and owners may choose to opt out of having any or all of their books included in those models on an ongoing basis;
- participating libraries, researchers and new educational subscribers can have access to the entire repertoire on certain terms – some paid and some not;
- A “Books Registry”, a new collecting society to be at arms length from Google and directed by author and publisher representatives, will be set up to register owners' interests, administer payments and research ownership issues; and
- standard input licence fees will apply, therefore avoiding costly fee negotiations with individual rights-holders.

The settlement will apply to the enormous class of authors and publishers whose books have been digitised by Google, *subject only to opt-out by individual rights-holders*. The class affected includes authors and publishers who are resident outside the United States, and notices have recently been published in Australian newspapers to alert affected owners.

The implications of the Settlement for orphan works significantly change the existing legal position under copyright law. For Google, the Settlement will remove the risks of using orphan works in ways which may not fall within the fair use defence. It can use orphan works for free, without having to try and find the owners first, and keep the revenue. For orphan works holders, their rights to sue will be significantly reduced should they fail to become aware of their rights and opt-out of the Settlement by the cut-off date. While they will be entitled to claim their compensation payment of

¹⁷ Google, submission to US Orphan Works Inquiry, p 2.

¹⁸ *The Authors Guild, Inc., et al. v. Google Inc.*, Case No. 05 CV 8136 (S.D.N.Y.).

around \$60, this is a non-negotiable fee and may not satisfy them as being reasonable compensation for Google's unauthorised use of their work.

The fact that orphan work holders will lose their right to sue after the cut off date is probably largely academic. Which author would want to sue Google alone? What is particularly significant is that the underlying principles of the Berne system are fundamentally altered by the Settlement on a very large scale. It may be that some orphan works holders are happy for this shift to occur, and to receive money once they become aware of the scheme's operation. The settlement does not just reflect just Google's interests as a user but the interests of large representative publisher and author bodies. It is in this sense a market solution to the issue of orphan works.

On the other hand, it is quite possible that the terms of the Settlement will be varied or even rejected outright by the court in fairness hearings scheduled for June 2009. Objections may be lodged by class members who may also seek to be heard. Given the objections of copyright owners to the much more conservative US orphan works bills, it is quite likely that the hearings will canvass similar concerns as have been raised there – although the Settlement only concerns books, and specifically excludes visual artists and photographers who have been the most vocal critics of the bills.

The Google litigation is merely one manifestation of the continuing tension between “old” content creation models, who (broadly speaking) favour the strengthening of copyright owner rights, and “new” digital enterprises which aggregate or facilitate distribution of content found online, which seek to expand user rights in the copyright material of others. The tension can be summed up by the contrast between Google's view of itself as an “archivist”¹⁹ and of the competing view that such search engines are “parasites or tech tapeworms in the intestines of the internet”.²⁰ The results of the fairness hearings, and the body count of copyright owners who decide to opt out of the Settlement, will be indicative of which view may prevail in practice, and ultimately how successful this model of compromise between content and search might be. In the meantime, it is interesting to consider how the Google scheme could be applied as a public model to orphan works in mass digitisations of other cultural material.

2.3 Possible applications of the Google Books Settlement principles

In his recent article in the New York Review of Books, author Robert Darnton suggested that Google has privatised an opportunity to create a digital Library of Alexandria which, in retrospect, should have been a public initiative:

“Looking back over the course of digitisation from the 1990s, we can now see that we missed a great opportunity.... We could have created a National Digital Library – the twentieth century equivalent of the Library of Alexandria. It is too late now. Not only have we failed to realize that possibility, but, even worse, we are allowing a question

¹⁹ Id, p 5.

²⁰ Robert Thomson, editor, Wall Street Journal, quoted in Jane Schulze, “Google dubbed internet parasite by WSJ editor”, The Australian, 6 April 2009.

of public policy – the control of access to information – to be determined by private lawsuit”.²¹

But perhaps it is not too late to claw back a public role in the debate. The Google scheme inherently relies on regulatory intervention – in this case via the class litigation rules, to underpin it. The copyright system itself is a legal intervention in the market which is in a constant feedback loop responding to new technologies and the uses they permit. The policy initiatives on orphan works so far have been slow to address the needs of the digital environment, but they can still catch up.

A basic starting point for considering a public licensing scheme is collective licensing. Collective licensing is far more efficient from a user’s point of view than individual licensing, as only one licence need be negotiated or registered for, standard fees apply, and no upfront clearances need be carried out. The Google Book Register will be a new collective licensing body underpinned by court order. Some collective licence mechanisms already deal with the issue of orphan works in various ways.

Most relevant to current purposes are *statutory, or compulsory, licences* and in particular “universal” statutory licences (eg, one permitting use of all copyright material falling within the class of rights even where the copyright owner is not a member of the relevant collecting society). An example in Australia is the educational copying licence under Part VA of the Copyright Act. This licence allows educational institutions to copy broadcasts for educational use in accordance with the relevant legislative provisions, whether or not the copyright owner is a member of the Screenrights, the declared collecting society.²² This type of universal scheme naturally provides legal use of orphan works included in the class and type of use covered by the scheme, with no need for the user to make inquiries or even realise that a particular right is orphaned, before the use occurs. Attempts to locate copyright owners whose works are included in such schemes are undertaken by the collecting society after reviewing records of reported usage. Undistributable revenue is held on trust while attempts are made to locate the owner.²³ Once a rights-holder is found, it is invited to become a member to receive payment of its statutory royalties.

These “universal” statutory licences provide a model not dissimilar to the Google Books Settlement scheme. They provide a “magic device” for solving orphan works problems by removing the need to find the owner upfront. The model is “user uses, owner registers to be paid”. However, such schemes as currently exist are limited to narrow user groups such as educational institutions. The Berne convention requires that exceptions to the absolute discretionary rights of copyright owners to be limited to “special cases”, and this is a potential obstacle to a large scale paradigm shift to Google-style opt-out schemes in the digital environment. It may be necessary that any

²¹ Robert Darnton, “Google and the Future of Books”, 12 February 2009, vol 56 *The New York Review of Books*, available at <http://www.nybooks.com/articles/22281>

²² The Attorney-General appointed Screenrights to administer this licence under section 135P of the Copyright Act. See the Simpson Report on Collecting Societies (1995) at 31 and www.screenrights.org.

²³ Methods of dealing with undistributable funds vary between collecting societies – but may include rolling such funds back into the pool of distributable monies after a certain period. See Simpson Report on Collecting Societies, ch 21.

new schemes remain limited to “special cases” and otherwise satisfy the Berne “three step test” before they can operate.

A variation on this idea is offered by Darnton. With hindsight, he suggests, a public version of the Google Books project could have been built by a coalition of libraries, supported with action by Congress, who,

... “could have done the job at a feasible cost and designed it in a manner that would put the public interest first.”

He suggests briefly that the idea of a “rental based on the amount of use of a database” could have been as one possible way the scheme could have worked. In my view, there is no reason why such a scheme could not work as compulsory licensing, if the details could be satisfactorily worked out. After all, the Google Settlement is a non-exclusive scheme. Authors and publishers remaining in the Settlement can still licence their books to other digital access schemes if they wish. Contrary to Darnton’s pessimistic view, there is no reason why a publicly accessible “magic device” to solve orphan works issues cannot now be found outside the Google Books Settlement – nor why such schemes could not be applied outside the medium of print.

Another possible mechanism was suggested by the Copyright Agency Limited (CAL) to the Fair Use Inquiry. CAL proposed that where a collecting society could establish majority member support for a new voluntary licensing scheme, such a scheme could be approved by the relevant review body under the Copyright legislation – the Copyright Tribunal – and if approved, underpinned by a statutory scheme. The statutory scheme would function to deem all members of the class to have licenced their works – not just those who actively volunteered to do so. Orphan works would therefore be included by compulsory licence. The CAL proposal is based on the philosophy of majority rule:

*“The underlying thinking for this proposal is that if a significant proportion of copyright owners in a particular class of works agree to participate in a collective management scheme, then that scheme would be consistent with the expectations of other copyright owners in that particular class”.*²⁴

Innovatively, and unlike the existing educational and government statutory licences,²⁵ CAL suggested that an owner could opt out. For example, the owner of an orphan work who later emerges could decide to opt out of the scheme by notifying the collecting society.

The CAL proposal is not dissimilar to the Google Settlement in the way it addresses orphan works. Its main flaw appears is requires a collective body to exist.²⁶ It doesn’t

²⁴ CAL submission to the Fair Use Inquiry (2005), Attorney General’s Department, para 69.

²⁵ In addition to Part VA discussed above, See also Part VB, Divs 2 and 2A of the Copyright Act, which covers educational copying and communication of print and electronic works (for which CAL is the declared collecting society), and section 183, which permits remunerated copying of all copyright material for government purposes. It is not clear whether “government” purposes includes purposes of the ABC and SBS, which are independent statutory corporations, and it is therefore unclear as to whether ABC or SBS could use this statutory licence.

²⁶ McDonald, 180.

work where collective licensing bodies have not yet emerged for a particular class of rights-holder. For those, it seems, legislative intervention may be needed to establish new declared collecting societies or “rights registries”.

Finally, in reviewing how the Google Books Settlement principles could be applied to public models, it is worth considering Google’s own submission to the public policy debate in the US Orphan Works Bill review. This submission, made in 2005 when the Google Library Books Scheme was at the height of its controversy, is interesting for two reasons – one because of the philosophy Google applies to orphan works, and secondly, for its suggestions on how legislative change might have been introduced to solve the orphan works issue without recourse to the class settlement which has since overtaken the issue.

Google’s philosophy on orphan works is to query the presumption that owners who cannot be contacted to give permission to use their works do not want any use of their work to be made. Google calls this a “conservative assumption mandated by the current state of copyright law”, and argues that the class of orphan works owners must include at least some who do not care if others use their works. Public policy demands, in Google’s view, that the presumption of non-use be altered so that orphan works can be removed from “purgatory” for the benefit of human knowledge.

Google’s suggested solution was that a register system be used to determine which owners:

“signal a lack of interest in enforcing their copyrights by not updating the records of their ownership.”²⁷

If an owner did not update their ownership records, then, Google argued, certain uses of the work – to be determined in the legislation - should be permitted without risk. Google also suggested that any such register should be capable of automated search – a request which foreshadows the future of peer to peer machine licensing in mass digital projects.

All of the above suggestions are conceptual only at this stage, and will attract concerns that will need to be worked through. The Google Settlement is tailored to the particular interests and concerns of the parties to the litigation at hand, and as commentators have noted, the public so far has had no direct role in shaping it. In contrast, any new scheme developed to address orphan works in digital projects will need to satisfy a number of public interest criteria. First it will need to fill a specific perceived need and have the support of government, stakeholders and the public. It will need to exist within the international treaty framework as noted above, respect the interests of rights holders, continue incentives for creativity and uphold the public interest in access to cultural material. It would need to do this while allowing for new developments in the digital economy and maintaining the key benefits of efficiencies and risk minimisation for users.

This is no easy task, but it is not too late to watch and learn as Google continues its grand experiment with copyright law.

²⁷ Google submission to US Orphan Works inquiry, 3-4

3. Conclusion

Legislators are yet to find a satisfactory solution to the problem of orphan works in mass digitisation projects. Google meanwhile is developing its own private scheme. If Google succeeds in getting this scheme judicially sanctioned and winning over a critical majority of copyright owners, it will have created, with the court's assistance, something akin to a new statutory licence for archival copyright material. It will neatly solve the orphan works issues which plague digital archive projects. And it will unlock revenue and public access – at least to a greater extent than previously possible – in such material. On these grounds alone Google will have succeeded where many have failed, and for this reason it is worth watching to see what ideas can be used.