Open Content Licenses

Jan Newmarch
Faculty of IT
Monash University, Australia
jan.newmarch@infotech.monash.edu.au

August 23, 2006

Abstract

Open source licenses such as the GPL are designed as licenses for software. Although they can be used for other types of content such as books, a set of licenses called *open content* licenses have been developed for documents of various kinds. This chapter examines the reasons why such licenses have been developed and looks at some typical examples.
Keywords: open content, dark matter, copyright, Creative Commons, GFDL

1 Introduction

> *Obscurity is a far greater threat to authors and creative artists than piracy* - Tim O'Reilly[11]

For many years individuals and communities have been contributing to the sum of human factual and creative knowledge. The original ideas of copyright and patents in the US were to ensure that this knowledge remained accessible to all, with strict time limits before material passed into the public domain. Of late the intention has often been perverted to ensure that certain companies can retain a long-term control over some profit-making resource. For example, commentators have remarked on the convenience of the US extending the copyright period just as Mickey Mouse was due to turn 50 years old.

The Open Source movement began to ensure that computer source code could be quarantined against unnecessary restrictions and achieved success through the use of novel copyright statements such as the GPL and Berkeley licenses. However, knowledge comes in many forms that are also covered by copyright law, such as images, novels, poems, dictionaries, computer documentation and various types of performance. The copyright statements designed for software may not always be the best when applied to these other forms of expression.

The *open content* licenses are designed to deal with some of these other content styles. Similar to open source licenses they are designed to ensure that the content remains "free" in some sense. For example, a license may ensure that it must be possible

to copy and distribute a book without having to pay a publisher. Such a license may indeed forbid publication for profit, while another license might allow such sales by a particular publisher and a third license might allow any publisher to print and sell copies.

A typical copyright notice found in many books might say

> All rights reserved. No part of this book may be reproduced in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission in writing from the publisher.

An open content license on the other hand might say

> Open Publication works may be reproduced and distributed in whole or in part, in any medium physical or electronic, ... Commercial redistribution of Open Publication-licensed material is permitted. Any publication in standard (paper) book form shall require the citation of the original publisher and author[15].

It can be seen that whereas the standard copyright notice only serves to restrict rights to the publisher, an open content license opens many of these rights to all. As with open source licenses, the primary remaining restriction is the right of the author to retain citation in the document.

2 Personal experience

I have authored three books. The first was written in traditional style over a year, with drafts sent to the publisher for review by a small number of referees and with publication always in doubt. In general, this was a fairly miserable experience. In particular, one of the referees clearly wanted me to write a different book to what I wanted, and kept leading it further and further in the direction he wanted until he finally stated that he was bored with it. During this period I felt I was largely working in the dark and I really don't know how or why I persevered.

The book–a different one to what I wanted to write–was finally published and sold about two thousand copies.

For the second one I had a more sympathetic publishing team, and it was a more pleasurable experience. During this I had a running fight with the publisher as I wanted to prepare the book myself using Interleaf and they doubted my competence. Nevertheless, the fate of the book was always up to the publisher, and I was never really in control. This one sold eight thousand copies.

By the time I got to my third book, it was clear that (like the majority of authors) I hadn't cracked the art of writing a best-seller. So instead of aiming at making money I set out to enjoy myself and keep maximum control. The book started at ten pages, and I added about ten pages per week for six months. All of this was done in public by putting all versions on the web under an open content license.

The open content license ensured that my rights as author were guaranteed. Other authors could have added content and I would have welcomed that, but back in 1999

collaborative works such as wikis were not a common model and I ended up as sole author. But I did receive substantial feedback right from the beginning and this continues even now. This to me is a significant advantage of what I might call "open development of open content", which is a common model for open content: feedback from the beginning to end.

The other significant gain to me was that I was not beholden to any publisher. The web allowed publishing without the requirement of a hard-copy publisher, and the license allowed the widest possible distribution.

The book was a success in its format. It was about a particular field of computing, and as a result of the book I became well-known in the field, from its initial inception at the beginning of 1999. Eventually, a publisher came to me, and we negotiated a deal in which a hard-copy appeared in 2001. This hard-copy sold reasonably well, but not enough to retire on.

While I did not make a fortune, as a result of the license and process, I achieved the following goals

- I got feedback right from the beginning versions of the book from a varied audience, not from a small number of publisher-nominated reviewers. As a result, the book was substantially improved by ongoing peer-review.
- Since anyone could print their own version, or even distribute copies to classes, it achieved a wide distribution from the beginning.
- I become a well-known figure in the field from the beginning of the project, not just at the end.

There is a serious question about whether sales of the hard-copy were increased or decreased as a result of this open license. The traditional means of selling books is "buy before you try". Buying from a catalogue supports this style. Bookshops give a limited form of "try before you buy" by allowing a potential purchaser to browse. Many on line book sellers will now make a sample chapter available. But in all of these cases the purchase is done under limited knowledge of the book. Making the whole content available improves the knowledge of the buyer, so that they are more likely to make an informed purchase.

On the other hand, if the content is available for free, the incentive to pay for a copy is reduced. There are significant advantages to having the material on line: for example, sections can be copied and pasted into other documents. Anecdotally, I have many emails saying that they have bought a hard-copy version because they prefer that to a screen-based version. This may come down to the "toilet" problem: would you rather balance a book on your knee, or a computer?

I have also changed presentation styles because of the design for an on line rather than paper version. Books are constrained by paper–too much information and the book becomes too heavy and expensive. This is not a limitation of web-based material. On the other hand, it is common to give a PowerPoint dot-point view of the world, and this often dumbs down the potential of the Web. I treat it as a opportunity to give as detailed a view as is necessary, free of the constraints of paper costs or of screen sizes (after all, detail can be at the end of a link). Consequently, instead of giving snippets

of programs as a book or PowerPoint often does, I am able to include fully working programs.

Cedergren[2] analyses three open content projects from a "value chain" viewpoint, to identify value factors for open content projects in general. He identifies factors such as the stimulation of cooperation, learning from others, intrinsic motivation for open content, altruism, discontent with standard media and opportunities for new business models. My own experience supports this model. Tim O'Reilly[12] says "There's an even more fundamental freedom that underlies the work of both free software advocates and the most proprietary of software developers, as well as anyone else engaged in creative work. And that is the freedom to offer your work to the world on the terms that you choose, and for the recipients to accept or reject those terms." Vint Cerf[3], when asked which technology has evolved in the most surprising way, commented: "The rate at which people are contributing information to the World Wide Web." These two offer other perspectives on why people contribute open content and the scale on which they are doing it.

3 Process and license

While open source existed before the web, the explosion of open source activity has only become possible because of the huge increase in all internet activity such as email and ftp caused by the web. Open source repositories such as SourceForge have been immensely important in allowing access to source code.

There are two issues in such repositories: one is based on *process*, the other on *license*. The process is often phrased as "release soon, release often." The process makes new versions available as soon as possible and encourages feedback. The license (if it is open source) encourages others to make their own changes. The combination of the two can lead to rapid evolution and development of software.

My own experience in the process for open content was that feedback was swift and useful. The experience of the many Web Logs (blogs) and shared content documents such as Wikipedia is that the opportunity to add or change documents also leads to rapid evolution. That is, some of the factors that make open *source* valuable also make open *content* valuable.

4 Inaccessible knowledge

The web consists of static and dynamic content. Static pages can be easily indexed by search engines and found by users. Dynamic pages are often hidden behind form parameters and are not easily found unless you enter the correct sequence of tag/information pairs. Typically, this content resides in databases and web forms are used to provide a front-end to the database content. Such content has been labelled the "dark matter" of the internet[8]. In many cases, the producers of this content will attempt to make it easier to be found by search engines, as it is in their commercial interest for it to be visible. Sharkey[14] refers to PCs as the dark matter of the internet, but here he is talking about unused processing power rather than inaccessible knowledge.)

According to the Wikipedia entry for the "Gutenberg Galaxy," the British Library holds about 50 million items while the Library of Congress holds about 119 million items. The National Library of Australia reports[9] that it increased its collection by 417,000 items in 2004-2005. The "World list of scientific periodicals published in the years 1900-1960"[1] lists 60,000 periodicals and the Web of Knowledge[] indexes 8,000 journals on an ongoing basis. There is an enormous amount of information being produced, but the majority of this content is inaccessible.

The world has seen a sea-change in how we attempt to access information. Instead of browsing through libraries, simple access to a search engine such as Google will unearth a wealth of information. Previously inaccessible material in paper form remains inaccessible and is being replaced by on line information sources. We can generalise from "dark matter of the internet" to "dark matter of information systems": the content exists but in general is not on the internet in any form.

There are many attempts to bring this "dark matter" of paper material into the electronic domain. For example, book vendors such as Amazon usually make selected parts of books available such as cover pages, tables of contents and sometimes sample chapters.

A more ambitious project that is receiving attention is Google Print, where university library collections are being scanned and indexed. Once done, searches will be able to access printed material over hundreds of years. However, this is facing intense opposition from many publishers, who claim that their copyright is being infringed by Google Print.

On another front, publishers of academic journals have been charging ever-increasing rates for their publications, costs which are typically borne by university libraries and individual academics. In most cases, the actual content is donated free by the authors, and the referees and editors also donate their services. The general public is often the endpoint for these explicit and hidden costs since many universities and research projects are (at least in part) government funded. Consequently, there is much pressure being brought on these publishers to place the content back into the public domain. Publishers are responding in a variety of ways: allowing authors to place pre-edited versions in public respositories; allowing authors to place final copies in such repositories after permission is sought (and granted) from the publisher; and the publishers themselves making the content publically available after a period of, say, six months after publication.

The key to publisher power in keeping content as "dark matter" is of course copyright. Authors have surrendered their copyright to the publisher who then has control over the visibility of the content. Under any of the open content licenses this would not be possible.

5 Meaning of copyright

Copyright is usually expressed as a "negative right". That is, rather than allowing people to do things, it restricts them. From `http://www.patents-info.com/Default.aspx?PageContentID=37&tabid=178`:

> A copyright holder typically has exclusive rights:

- to make and sell copies of the work (including, typically, electronic copies)
- to import or export the work
- to make derivative works
- to publicly perform the work
- to sell or assign these rights to others

> What is meant by the phrase ”exclusive right” is that the copyright holder and only the copyright holder is allowed to do these things; everyone else is prohibited from doing them without the copyright holder's consent.

In most countries copyright exists merely by authoring a document. That is, copyright does not have to be explicitly claimed. A large number of documents on the Web have no claim of copyright. Consequently, copyright exists and is held by the author. This limits the rights of others to copy, reuse or alter the material. This holds, for example, for the majority of Web Logs (blogs) even though they are often fairly casual documents, and the author has no intention of placing any particular restrictions on use.

Copyright may be assigned to others. This may give up all or some of the rights. In addition, copyright may be assigned on an *exclusive* or *non-exclusive* basis.

In addition to copyright, an important concept for many authors is that of the “moral right to be asserted as the author”. In many countries this needs to be asserted explicitly but once done cannot be removed or given away, unlike copyright. Most open source and open content licenses do not use this, but include the identity of the author as an explicit part of the copyright statement.

6 Copyright history

The on line encyclopedia Wikipedia has lengthy sections on copyright history in general and more details on copyright law for particular countries. According to this, copyright is a concept stretching back at least as far as the Library of Alexandria. Early developments of copyright law in the fourteenth and fifteenth centuries were primarily to give copyright to publishers in order to protect their investment.

The British Statute of Anne gave rise to the modern concept of copyright, giving copyright control to the author. However, it also brought in a time limit to this, so that works went into the public domain after a period long enough to bring benefit to the author, initially of a maximum of 28 years.

Chartrand[4] points out in his detailed history of copyright that granting rights to the author in practice just granted them to the publisher, and apart from the expiration period, strengthened the position of publishers while doing little for authors.

The US copyright law was based on the English law, but initially was explicitly aimed at improving the position of the new republic: for example, no copyright was granted to works of foreign origin, so that books from England were given no protection at all. Over the years, this has changed so that the US along with most other countries now follows the Berne Convention. One consequence of this is that it is no longer necessary to use the copyright symbol (©) to protect a work: copyright is automatic

in countries that follow the Berne Convention. However, it is still common practice to use the copyright symbol. The period for which copyright holds has been steadily increased typically to the life of the author plus fity years, which is inarguably well beyond the period of benefit to the author–but well within the period of benefit to the publisher!

7 Problems with software licenses

It is possible to license text-based work under an open software license. However, works such as books usually have a different structure and usage to software source code, and this may be enough reason to use a different license.

7.1 Where is the value?

A major difference between, say a book and a program, is the use to which the original text is put. The text in a book is interpreted directly by people, as they read it. The text itself is the primary value. A program on the other hand has primary value once it is compiled so that it can be executed. However, in that form it is "locked away" and is no longer open to inspection, changes or development by other people. Open source protection of the source code is to ensure the rights of others to produce their own compiled versions. Open content protection, on the other hand, is to ensure that it is kept open in its primary form.

7.2 What is protected?

The commercial venues for books and programs differ too. Traditionally, a program gains commercial value once it is compiled, put on a CD and made available through a software vendor. A user exploits this value by loading the program and executing it. A closed source vendor will never make the original source available. The license protects copying of the *executable* program rather than the *source* and the source is kept as a trade secret.

The traditional print model is that a book is published by a company, and what they publish is of course the text itself. This is then open to public view, and is protected by licenses that forbid copying or replication of the text itself. That is, in the closed content model the *source* is protected from copying.

7.3 What is limited?

The open source licenses for software generally have no restriction on compilation and sale of the executable and attempt to ensure that the originating source code remains open and free. On the other hand, the open content licenses attempt to ensure that the originating content remains open and free, but may limit the mechanisms by which others can sell the content–for example, by only allowing a single publisher to print a commercially viable number of copies.

7.4 Structure

Software and books have different structure. A program has structure required by the programming language, but really only requires a slot somewhere to record the copyright notice and author attribution. A book on the other hand will often have several sections

- Publisher information (including ISBN)
- Author information
- Foreword (by someone else)
- Dedications
- Table of contents
- The actual content
- References
- Index

In general, the open source licenses allow you to change any part of the the program that you want to, and only protect the license itself and the author attributions. For open content, the holder of the copyright may wish to ensure that, for example, the dedications are not removed, and the original preface must always remain. An open content license may have means of labeling and protecting sections of a book that are not possible for open source licenses.

8 Typical Open Content licenses

8.1 Creative Commons

The Creative Commons license was published in 2002[5] and is intended for use by authors and artists. As well as written material such as books it is also intended for music, images such as photographs and videos. Even for textual materials, the scope is wide: books, pamphlets, course lessons, blogs and so on.

The Creative Commons license is explicitly aimed at sharing information:

- to reproduce the work in any medium
- to make derivative works
- to distribute copies of the work
- to distribute copies of derivatives

The default license even allows you to sell copies for profit.

In its default form, the license is similar to the Berkeley software license. For most people, this is too generous. There are four optional riders that can be added to this license which restrict some of these:

- while others can use, copy or make derivatives, they must give you credit as original author
- no-one can use it for commercial purposes
- no-one can make derivative works
- derivative works must use the same license

Not all of these are possible combinations. A typical license will allow anyone else to use the work for non-commercial purposes and make derivative works (such as abstracts) under the same license, and that the original author must be acknowledged.

It should be noted what it does allow

- the original author can still negotiate with, say, a publisher to sell copies
- while the author may chose an option that forbids derivations, they can still grant this right on an individual basis
- the license allows others to make this a part of some larger work such as a collection of papers. It is not necessary for this larger work to use the Creative Commons license. In other words, using the work as a piece of larger does not "infect" the larger work (unlike the GNU software license)
- nothing takes away common law rights to use a "reasonable" sample

8.2 Gnu Free Documentation License

The Gnu Free Documentation License (GFDL) is an open content license from the Free Software Foundation[7]. The on line encyclopedia Wikipedia has adopted this license and so the many thousands of authors who are contributing to this are doing so under an open content license.

This is similar in intent to the GNU General Public License. It belongs on the extreme of open licenses in that it enforces as few restrictions as possible. You can reproduce, sell or modify content protected by this license. What you are not allowed to do is to remove the copyright or attempt to place any constraints on the freedoms it gives to users of this content.

In contrast to the GPL software license in which the content is regarded as a homogenous quantity. the GFDL pays attention to the particular structure of text. The GPDL separates content and metadata about the content, and the metadata is subject to different rules.

Metadata is called "secondary sections" and includes things like the publication history of the content, information about the author, reviews, etc. A secondary section is any section of the work that is not explicitly content. While a little loose in wording, most books (for example) are readily split into content and secondary sections.

As a special case of secondary sections are *invariant sections*. While a secondary section could be modified, an invariant section cannot. For example, the original title might be labeled as invariant so that it must appear in every version of the content.

The GFDL rules are generally reasonable if all you are doing is copying or reproducing the work. One of the less satisfactory rules is that the full text of the copyright notice must appear in every copy of the work and as this takes five pages it can often be a waste of paper. However, the license has many conditions that can become a nuisance as soon as derivative versions are made. For example, if you change the content you are required to change the title, but also reproduce the original title. Similarly, dedications must be kept. This can lead to documents with more secondary sections than actual content!

The rules for derived copies are probably designed to deal with the situation in which the content evolves in a way unacceptable to the original author. This can certainly happen and there aren't any simple and widespread ways in which any piece of text in a book or other document can be attributed to any particular author. Essentially this is a problem in version control, and this is not represented well in common publishing systems.

9 Infighting

There are many open source licenses for software and many strongly held opinions about the virtues or otherwise of particular licenses. While in many cases these may appear to be nit-picking, they can also represent fundamental differences in philosophy. These differences also appear with open content licenses.

For example, Debian is a well-known open source distribution of the Linux (or GNU/Linux system). They have defined policies about what is acceptable as open source, and enforce these rules in their software distributions[6]. Debian has a number of objections to the Creative Commons license, some of which *are* nit-picking while others are more fundamental[13].

The Creative Commons license usually allows derivative works to be made. In these, it requires that the author attribution must remain. However, an author may feel that the new work has deviated so far from their original intention that they no longer wish to be associated with it. The original author can ask for their attribution to be removed and the new author has to do this. However, suppose that the new author is pointing out errors made by the old author? Debian claim that the actual wording of the license may allow the original author to require removal of the original text (not just attribution), which would void the criticism and amount to censorship. This can be fixed by appropriate re-wording of the Creative Commons license.

I made my book available on line first under an Open Content license and later under the Creative Commons license. While I wanted it to be open, I also wanted to attract a publisher to sell printed copies from which I could get royalties. To do this, I invoked the "no commercial use" option for the on line version, to give me some control of commercially printed copies. This is against the Debian statements about the meaning of "free", and use of this option disqualifies the content from being "free" in their terminology.

Another criticism of both the Creative Commons and GFDL licenses is about their phrasing about Digital Rights Management (DRM). Both licenses ban restricting content by a DRM system. However, Debian points out that as worded, it may invalidate

legitimate uses (such as saving in Postscript format to send to a printer).

I regard this provision as particularly important: I work for an Australian university. As in many countries, the moves in Australia to a commercial model for education increases the probability that my university will consider the intellectual property produced by its academics as property to be bartered, sold and above all *restricted* to fee-paying students. I have argued elsewhere that this a bad approach for many reasons[10], but I finally hope to rely on the anti-DRM provisions of these licenses to ensure that the university will not be able to close access to my open content material.

The GFDL license is apparently incompatible with the GPL license. This means that GFDL documentation cannot be included in a GPL program and *vice versa*. This kind of hiccup can be overcome by licensing both documentation and source under both licenses, but this will rapidly become tedious.

Philosophical differences will always remain, and will always cause diverse opinions. On the other hand, legal pedanticisms can be overcome with sufficient lawyers. There are still problems with sufficient technical content to be overcome, though. One of these is how express anti-DRM concepts in a way that does not restrict ordinary users. Another is how to deal with managing licenses for content across multiple application domains. The primary example for this is software documentation. A third issue is how to manage derived copies in a simple and unobtrusive manner.

10 Conclusion

The world of software has evolved a set of licenses tuned to that environment. Any programmer given some source code will want to play with it and modify it. Given an executable module, there is little that can be done with it apart from just executing it. The open source licenses are there to ensure that the source code remains available and that programmers can scratch any itch they have.

Works that are released in primarily final form such as books, photographs, videos and music performances benefit from a license that respects the finality of that format while still leaving it open for everyone to use it. The Creative Commons license is one attempt to give a license for this use. Other works–and again books are a typical example–have more complex structures than software, and a license such as the GNU GFDL can deal with this structure.

These two licenses are not the final word on open content licenses and we can expect to see more in the future (others already exist). When licenses are tuned to particular use domains, it is inevitable that conflicts in philosophy will arise between different beliefs as to the "perfect license", and conflicts in practice will arise when a work is required to satisfy multiple roles.

References

[1] P. Brown and G. B. Stratton, editors. *World list of scientific periodicals published in the years 1900-1960*. Butterworths, 1965.

[2] M. Cedergreen. Open content and value creation. *First Monday*, 2003. `http://firstmonday.org/issues/issue8_8/cedergren/index.html`.

[3] V. G. Cerf. The view from the top. *Spectrum*, 2004.
[4] H. H. Chartrand. Copyright c.p.u. http://www.culturaleconomics.atfreeweb.com/cpu.htm.
[5] Creative Commons. Creative commons. http://creativecommons.org/.
[6] Debian. Debian free software guidelines. http://www.debian.org/social_contract#guidelines.
[7] Free Software Foundation. Gnu free documentation license. http://www.fsf.org/licensing/licenses/fdl.html.
[8] D. Hawking. Dark matter on the web. http://www9.org/final-posters/poster30.html.
[9] National Library of Australia. Annual report - operations. http://www.nla.gov.au/policy/annrep05/pdf/NLA_Annual_Report-3-Report_on_operations.pdf.
[10] J. Newmarch. Who owns university courseware? In *AusWeb 2000*. http://jan.netcomp.monash.edu.au/ausweb00/course_owner.html.
[11] T. O'Reilly. Piracy is progressive taxation, and other thoughts on the evolution of online distribution. http://www.openp2p.com/lpt/a/3015.
[12] T. O'Reilly. Open source software, 2006. http://tim.oreilly.com/opensource/index.csp.
[13] E. Prodomou. debian-legal summary of creative commons 2.0 licenses. http://rym.waglo.com/wordpress/index.php?p=201.
[14] C. Sharkey. Pcs are the dark matter of the internet. http://www.shirky.com/writings/dark_matter.html.
[15] D. Wiley. Open publication license. http://www.opencontent.org/openpub/.