Digital Proceedings of the Lawrence J. Schoenberg Symposium on Manuscript Studies in the Digital Age

Volume 2
Issue 1 *Lex scripta: The Manuscript as Witness to the History of Law*

Article 7

4-9-2010

# Digitized Manuscripts and Open Licensing

Hugh Cayless
*New York University*, hugh.cayless@nyu.edu

# Digitized Manuscripts and Open Licensing

**Abstract**
This report examines the issue of copyright law in digitizing manuscripts and making images available online. Specifically, it looks at the possible solutions provided by Creative Commons.

**Keywords**
Digitization, medieval manuscripts, open licensing

# Digitized Manuscripts and Open Licensing

Hugh A. Cayless
*New York University*

## The Basics

Copyright law in the United States changed considerably in the last quarter of the 20th century.[1] Prior to 1978, works could be copyrighted (and didn't have to be) for a term of 28 years. This term could be renewed for a second 28-year period, for a maximum of 56 years. The Copyright Act of 1976[2] (which took effect in 1978) made copyright automatic (i.e. no registration was required) and extended the term to the life of the author plus 50 years. The Copyright Term Extension Act of 1998[3] extended the term by 20 years. This means any original work you produce today will be subject to copyright restrictions until 70 years after your death.

This situation is ideal for the owners of intellectual property that makes a lot of money, but it causes significant problems in the case of works that either were never intended to make money, or that have stopped doing so, or whose authors have vanished, because it is illegal to reproduce them. It is not unlikely that a significant portion of the intellectual output of the 20th century will disappear as a result.[4]

None of this matters especially when using copyrighted material for personal research, since this kind of use is almost certain to be covered by the Fair Use exemption, but it does become relevant when it comes to publishing that research, either via traditional avenues or via the internet. While it is easy enough for an individual to ignore intellectual property rights when posting content to his own website, organizations you may be working with to publish your work, such as libraries and presses, are likely to be very reluctant to publish material without copyright clearances. Moreover, unlicensed material posted on a website hosted by an internet service provider may be subject to takedown notices (even spurious ones, since the host is not likely to bother researching the complaint). Licensing content is a hygienic measure that makes it clear how it may be reused without the author's personal permission.

Any original work that is not produced as a work-for-hire is the property of the creator, and is automatically subject to copyright law. No deployment of the copyright symbol or assertion of the author's rights is necessary. The practical implications of this are that material posted on the web, such as scholarly papers, photographs, syllabi, etc. may not legally be reused

1 See http://www.copyright.gov/title17/ for the current U.S. copyright statute.
2 See http://en.wikipedia.org/wiki/Copyright_Act_of_1976
3 See http://en.wikipedia.org/wiki/Copyright_Term_Extension_Act
4 http://www.thepublicdomain.org/2009/12/31/fahrenheit-451-book-burning-as-done-by-lawyers/

without the permission of the author, even if the intent of the author in providing free access to the material was for it to be taken and used by anyone. This fact is one of the main reasons open licensing schemes, such as those provided by the Creative Commons, exist.[5] A license provides a way for an author to legally grant permission to use her work in ways that would otherwise break copyright law. Licensing régimes like Creative Commons are not anti-copyright. In fact, they depend upon copyright law in order to function. It is only because the creator of a work automatically possesses these rights that she is able to give blanket permission to certain kinds of reuse.

Such licensing schemes appeared first on software source code in the 1980's. They had a variety of uses, but in general, they stated that the code so licensed came without any warranty and that the copyright holders were not liable for any defects. They might also add provisions, specifying (for example) that the license had to be displayed on any distribution of the software, or that copies of the source code had to be made available alongside compiled (executable) distributions. The motivation for this kind of licensing was both economic and ideological. Companies would be reluctant to reuse source code without a license, because they could later be sued for copyright infringement. All commercial software comes with a license stating what can be done with it, because installing and using the software nearly always entails making copies of it, so copyright law imposes a need for licensing. The ideological motivation came from advocates for Free Software, such as Richard Stallman,[6] who argue that software should be able to be run for any purpose, studied and changed to make it better suit your purposes, and freely redistributed, including the right to redistribute your changes.[7]

**Types of licensing**

Creative Commons (CC) has four basic provisions, which may be combined in various ways: Attribution, No Derivative Works, Sharealike, and Non-Commercial.[8] Attribution by itself allows the licensed content to be reused in any way as long as the author is credited. No Derivative Works means what it sounds like: the content can be redistributed in its published form, but not in an altered form. Sharealike means that the content can be altered and redistributed, but that the derived content must carry the same license going forward. Non-Commercial means that the content can be redistributed, but not sold for monetary gain. If another party wishes to use your content in a way that does not conform to the license, then normal copyright law applies, and they have to obtain your permission or they are in violation of the copyright.

The Creative Commons website provides a variety of licenses using the provisions listed above, customized for the copyright codes of several countries, as well as in general versions. The licenses are legally binding, though I am unaware of any legal cases to date involving CC licensing. Any violation of the terms of a CC license would be covered under copyright law.

5 The Creative Commons website may be found at http://creativecommons.org/.
6 See http://en.wikipedia.org/wiki/Richard_Stallman for some background on Richard Stallman. See also the website of the Free Software Foundation: http://www.fsf.org/.
7 See the Free Software Defintion at http://www.gnu.org/philosophy/free-sw.html.
8 http://creativecommons.org/about/licenses/.

**What license should I use?**

A lot of scholars are tempted at first by the Non-Commercial provision. The idea of someone making money off your hard work may be distasteful, but this reaction is worth examining. As Kenneth Pennington remarked during the final panel, most of what he produces has no monetary value. Given that it may be impractical or undesirable to monetize your own scholarly creations, it is worth examining some of the implications of preventing others from doing so freely. Permitting commercial reuse means:

- When someone wants to use your (seminal) article on digital manuscript editing that you posted on your (now disappeared, but rescued from the Internet Archive or its descendant) weblog in a new compilation in 40 years time, he will not have to obtain your estate's permission.
- A scholar can use your photograph of a manuscript page in a journal article without needing to ask permission.
- An entrepreneur can print t-shirts using your digital photograph of a nice initial from a manuscript page, plus text of their own creation.

Some might object to "frivolous" uses like the third example, but it gets right at the heart of questions about cultural patrimony and transmission. Should reproductions of cultural objects that have never been subject to copyright (and that would no longer be, even if they once had) themselves be subject to copyright? The fact is that they are, and some uses of the copyright on photographs may be laudable, for example a museum or library funding its ongoing maintenance costs by selling digital or physical images of objects in its collection, but the existence of such examples does not provide an answer to the question: as an individual copyright owner, do you wish to exert control how other people use a photograph of something hundreds or thousands or years old?

No Derivative Works is similarly tempting, since it demands that the work remain unchanged, even though it may be freely redistributed. But this also means that mistakes cannot be corrected, new discoveries cannot be incorporated, and improvements cannot be made by anyone other than the original creator. Depending on the circumstances, this restriction may make sense, but it does mean that the work is essentially frozen.

If CC in general was inspired by the Free Software movement, the Sharealike provision in particular is inspired by the provisions of the GNU General Public License, which state that all software that incorporates GPL'ed code must adopt the GPL license and must make available the source code of any programs that are released which fall under the GPL.[9] The GPL is often described as a "viral" license because it transmits itself to other code that it touches. This property is sometimes controversial. There are a number of companies and projects that refuse to employ GPL'ed code because of its restrictions. In addition, projects that build on a variety of codebases may have difficulty with incompatible licenses. Sharealike works in a similar way: the licensed work can be modified, combined with other works, etc., but the result must be redistributed under a Sharealike license.

After sorting through these questions, a number of projects have determined that the correct answer is to require attribution only (which is simply responsible scholarly practice

9 http://www.gnu.org/licenses/gpl.html.

anyway) and not to make any other demands on the consumers of the information they publish. Two examples I would like to highlight are the Integrating Digital Papyrology (IDP)[10] and the Inscriptions of Aphrodisias (InsAph) projects, both of which make their data available under CC-by licenses. IDP documents are encoded in EpiDoc[11] XML and contain text like the following snippet:

```
<publicationStmt>
   <authority>NYU Digital Library Technology Services</authority>
   <idno type="filename">sb.1.4821</idno>
   <idno type="ddb-perseus-style">0239;1;4821</idno>
   <idno type="ddb-hybrid">sb;1;4821</idno>
   <availability>
      <p>© Duke Databank of Documentary Papyri. This work is licensed under a <ref
type="license" target="http://creativecommons.org/licenses/by/3.0/">Creative Commons
Attribution 3.0 License</ref>.</p>
   </availability>
</publicationStmt>
```

The InsAph project makes all of its texts downloadable as a single zip archive and explains the philosophy behind sharing them thus:

> Part of the philosophy of the Inscriptions of Aphrodisias publications—and the EpiDoc conventions that we use to encode the inscriptions—is that we are able to share the XML files (which are the source code behind the online editions, the indices and other reference materials) with other projects and colleagues. Accordingly the EpiDoc XML files have been released under Creative Commons Attribution 2.0 UK. Scholars are free to download these files, modify them, process them, recirculate them, publish the results in a database, website, or other environment, so long as they give us credit for the original files, and do not attribute to us anything that they have added or modified. They must also make clear to others the license terms of this work.[12]

This kind of publication, with a CC license attached, has a number of important implications:

1. The texts are easy to find, because they have been published on the open internet, with indices that link to them, so that search engines can easily discover them.
2. There is no impediment to using the texts, even if that use should turn into a commercial publishing opportunity.
3. The texts may be copied, archived, and re-hosted by a third party, such as a library without any need for permission.
4. These texts will be used more than closed-off texts are, for the reasons listed above.[13]

---

10 See http://papyri.info for data from the IDP and the related APIS (Advanced Papyrological Information System) projects.
11 EpiDoc (http://epidoc.sf.net) is a subset of the Text Encoding Initiative (http://www.tei-c.org) used to encode ancient texts.
12 http://insaph.kcl.ac.uk/iaph2007/inscriptions/xml-repo.html.
13 Although the precise nature of the "open access advantage" is unclear, it does seem to exist. See, e.g. Nicholas Joint in *Library Review* (2009) http://dx.doi.org/10.1108/00242530910978172.

Creative Commons provides a way to foster a healthy intellectual ecosystem in the digital environment, where there is a mismatch between the ease with which information may be discovered and copied and the restrictions imposed by the laws governing intellectual property. CC works by loosening or removing those restrictions and allowing information to flow as freely as possible. For personal scholarly publishing in particular (such as on a personal website), there is no reason not to use CC licenses for your work, and a great many reasons why you should.