

Creative Commons and open access to scholarly works: common myths about an imperfect match

 blogs.lse.ac.uk/impactofsocialsciences/2013/03/07/creative-commons-and-open-access-to-scholarly-works/

3/7/2013

Adequate licensing and attribution of scholarly work in the digital age have presented many issues for scholarly and publishing communities. While many open access advocates consider Creative Commons Attribution (CC-BY) to be consistent with the strongest form of open access as it includes the least restrictions, [Heather Morrison](#) compares the various CC licenses and argues that the lack of restrictions of CC-BY actually leaves open access vulnerable and other options should be explored.



Last November I completed a doctorate at the Simon Fraser University School of Communication on [Freedom for scholarship in the internet age](#). My dissertation builds on many years of open access advocacy and my career as a professional librarian. My work focuses on open access and transforming scholarly communication to a system that prioritizes what I consider the important goals of scholarly work: advancing the knowledge of humankind and supporting the work of scholars themselves. In the course of preparing my chapter on open access I undertook a preliminary mapping of open access and the Creative Commons (CC) licenses. My conclusion is that while the Creative Commons licenses are highly valuable tools for open access, the two don't really map.

For example, the concept “free of charge” is essential to any definition of open access, but none of the CC licenses are specific to free of charge. There are pros and cons to the use of any of the CC license elements for scholarship. The restrictive elements Sharealike, Noncommercial, and NoDerivs (no derivatives) limit re-use, but can also provide useful protection to scholars, research subjects, and even the open access status of the works themselves. Use of CC licenses is somewhat experimental at this point in time, and there are reasons to consider approaches that don't involve CC licenses. This is, in brief, the advice that I gave to this February's U.K. [Business, Innovation Skills Committee's inquiry into Government's Open Access Policy](#). This post presents the CC license choice options and illustrates some of the dilemmas arising from use or non-use of particular licenses, such as the common myth that CC-BY is necessary to enable data and text mining.

About Creative Commons licenses

CC licenses provide a means for creators to waive rights that are otherwise automatic under copyright. The licenses are designed for use by human readers (two versions, one simple and one legal code) and/or machines (for example, the flickr service allows users to select photos by CC license type).

The CC license choices include (from the CC [“About the Licenses”](#) site):

Attribution CC-BY

This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.

Attribution – Sharealike CC-BY-SA

This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to “copyleft” free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit

from incorporating content from Wikipedia and similarly licensed projects.

Attribution – NoDerivs CC-BY-ND (NoDerivs = no derivatives)

This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.

Attribution – Noncommercial CC-BY-NC

This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.

The elements can be combined, as is the Attribution-Noncommercial-Sharealike license, or CC-BY-NC-SA, or the Attribution-Noncommercial-NoDerivs or CC-BY-NC-ND. There is also a Public Domain (CC-0) option.

The CC licenses offer many benefits for scholarship, such as the ability to re-use works such as graphs and charts of other scholars without having to seek permission. Each element of the CC licenses adds restrictions that can either unnecessarily restrict scholarly re-use, or provide essential protection for scholars, depending on one's perspective. For example, CC-BY allows for commercial use and the creation of derivatives by any third party without permission seeking. This expands the usefulness of scholarly works, and introduces conveniences for researchers, teachers, and publishers, but may not be compatible with ethical treatment of research subjects.

At a superficial level, the CC-Attribution only license appears to embody the spirit of libre or strong open access as defined by the Budapest Open Access Initiative (BOAI):

By "open access" to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.

In spite of the superficial similarity between the BOAI definition and CC-BY, there are important distinctions, such as the "free availability" which is central to BOAI but not CC licenses as discussed earlier.

One example of how the reality of CC licenses does not necessarily match perception is the common myth that the CC-Attribution only license (CC-BY) is needed to facilitate data and text mining. Why is this a myth? Because CC-BY is not necessary, sufficient, or even desirable for data and text mining. Internet search engines routinely conduct data and text mining on a massive scale without any need for CC-BY. That's how search engines work! CC-BY can be used with works that are not at all suitable for data or text mining, such as locked-down PDFs. The Attribution element of CC-BY is problematic when a number of data / text mining sources are combined; data experts recommend CC-0 or public domain, not CC-BY.

If the goal is reusability of data, then the most direct way to achieve this may involve such means as developing standards for the data and metadata. For example, Stéphane Guidoin of Open North talked at the BC Open Data Summit in February 2013 about the rapid development and spread of transit applications, made possible in large part by the General Transit Feed Specification (GTFS) co-developed by Google and the City of Portland <https://developers.google.com/transit/gtfs/reference>. By using GTFS, a standard that helps to identify transit-related fields like bus route and stop number, a city facilitates the use of all the applications that build on this standard. The

same principles apply to data that is used by scholars: to re-use a dataset, it is important that the dataset be easily identified in a standard way and format that other scholars and their automated tools can easily ingest into systems designed for analysis.

There is a lot of heated debate about the noncommercial element of the CC license suite. One of the reasons behind this debate is that there is no common understanding of what is meant by commercial or noncommercial, as CC found in a report issued in 2009 <http://creativecommons.org/weblog/entry/17127>. The argument of those who oppose the use of the noncommercial element is that this will impede usage in cases where the licensor intended to allow usage. To some extent, this may simply reflect limitations on understanding of the nuances of copyright law among the population, as discussed in the CC report. For example, in a recent listserv discussion a noted scholar berated a publisher for using a CC NC license, saying that this meant that the scholar could not use the work in their teaching. The publisher replied that this was not implicit in the license, but was rather a matter of how the scholar chose to interpret the license.

Many open access advocates consider CC-BY to be consistent with the strongest form of open access, libre open access, as it includes the least restrictions. I argue that the lack of restrictions leaves open access vulnerable, for example vulnerable to re-enclosure for toll access dissemination downstream. For this reason, I consider CC-BY-NC-SA to be the closest choice of the CC license options for strong or libre open access, allowing a broad range of re-uses while imposing restrictions that protect the open access status of the work for the long term.

The full impact of the Creative Commons licenses at this point in time is not fully known. Allowing for the creation of derivatives could open up the potential to increase the speed of knowledge creation and/or the development of useful new tools and services, or it could slow down progress by facilitating the creation and dissemination of poor quality derivatives. For this reason, the use of particular licenses for scholarship should be considered experimental. Use of the CC licenses should be encouraged, but a particular license should not be selected as a default, and researchers should not be required to use a particular license. Perhaps the default statement that comes with [Open Journal Systems](#) journals, “This journal provides immediate open access to its content on the principle that making research freely available to the public supports a greater global exchange of knowledge” is sufficient.

Note: This article gives the views of the author(s), and not the position of the Impact of Social Sciences blog, nor of the London School of Economics.

About the author:

Heather Morrison recently completed a PhD at Simon Fraser University School of Communication on the topic *Freedom for scholarship in the internet age*. She specializes in transformative change in scholarly communication, especially open access, as well as being a professional librarian and adjunct faculty member at UBC's ischool. In July 2013 Heather will be joining the staff of the University of Ottawa's School of Information Studies as Assistant Professor. Links to most of Heather's work (published, informal and in-progress) can be found on her scholarly blog, [The Imaginary Journal of Poetic Economics](#). Or follow her on twitter [@hgmorrison](#).

- Copyright © The Author (or The Authors) - Unless otherwise stated, this work is licensed under a Creative Commons Attribution Unported 3.0 License.