Open licensing

Among the arguments surrounding open access in the humanities that have caused controversy, few have been so fierce as those concerning open licensing.[1] Sceptics believe that liberal reuse rights will fuel an epidemic of plagiarism-like practices, will allow commercial re-enclosure of academic work and will fundamentally violate the moral rights of the academic author. Proponents, conversely, have cited the technological and social advances that could be made with the possibility of reusing material. In this chapter I explore these claims from both sides in order to ascertain the risks and the benefits of open licensing provisions but also to give some pragmatic information about the licenses under discussion.

AN INTRODUCTION AND CONTEXT TO FREE AND OPEN LICENSING

Picking up where Chapter 1 left off, 'open licensing' refers to conditions under which a copyright holder allows others to reuse material in ways that go beyond those specified within the fair use (or 'fair dealing') provisions of copyright law. Open and free licensing, like open access, has a history rooted in the free software movement. However, to truly get to grips with licensing, whether open or not, it is first important to understand how these phenomena sit in relation to copyright. Licensing does not override, and its goal is not to abolish, copyright. Licensing, instead, depends upon the legal provisions of copyright.

Copyright law, in the UK, was introduced under the Statute of Anne in 1709, was cemented by the Copyright Act of 1911 and is currently implemented under the Copyright, Designs and Patents

86

Act of 1988. In the United States, copyright was enshrined in the constitution in 1787 and designed 'To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries'.[2] The international enforcement mechanism for copyright is the Berne Convention for the Protection of Literary and Artistic Works of 1886 in which signatories agreed to recognise the copyright of all signatory nationals as though they were home nationals. Copyright is automatically conferred on eligible works; simply by creating a work, the author invokes the legal protections. However, some things simply cannot be copyrighted: facts and ideas being the most notable categories. Instead, in these instances, what falls under copyright is the *expression of* the facts or ideas. Because expressions of the same idea can be similar, copyright also exists on a spectrum of enforcement and strength that must be weighed by a court when a challenge is made.

Copyright is generally held, in its contemporary usage, to separate economic rights from moral rights. Economic rights (the ability to reproduce the work, to make derivative works, to perform the work publicly, to display the work publicly and to transmit the work[3]) can be sold or transferred and treated as though they were any other form of property. Moral rights (such as the ability to be named as the author of the work, the right to be published anonymously and the right to the integrity of the work) cannot usually be sold or transferred but can be waived.[4] The time period for which copyright applies is, counter-intuitively, incredibly difficult to ascertain accurately. To grossly simplify, however, in the case of 'a personal author who produced a work on or after January 1, 1978, it is covered for the life of the author plus seventy years'.[5] As an amusing aside, critics of the current copyright regime note that there appears to be a direct coincidence of legislation to extend copyright terms at the moments when the still highly lucrative works of The Walt Disney Company are to become public domain.[6]

Under the contemporary system of academic publishing as of 2014, there are a variety of approaches taken by different publishers towards licensing and copyright: most publishers ask for a copyright assignment or transfer; some ask for an exclusive license to publish; occasionally, a non-exclusive license to publish may be appropriate;

and last, but not least for the context at hand, comes a request for a non-exclusive license to publish under a Creative Commons (or other open) license.

In a **copyright assignment** or **transfer** model, the author agrees to transfer (irrevocably give) his or her economic rights to the publisher. The publisher then may, at its discretion, permit the author to perform certain activities that go beyond fair use with the material (such as deposit in an institutional repository – green OA). This is implemented so that the publisher may protect the author against copyright violations, libel or plagiarism and to facilitate requests for reprints.[7] Of course, a publisher could offer either financial or in-kind legal advice to their authors without such a provision. There is, therefore, a case for balancing the rhetoric of author protection against the economic advantages for a publisher of a copyright transfer. It seems more probable that publishers prefer copyright transfer both because it gives them full and exclusive ownership of the material for the entire term, including distribution in new geographical areas or in new forms not covered explicitly under licenses to publish, and also because it centralises their ability to protect intellectual property. The argument in favour of this is that publishers often invest substantial quantities of labour time (at a price) into taking on work and that, within the current sales/ subscription model, it may be easier to recover costs and/or make a profit with this form of ownership. Conversely, the author permanently signs away his or her economic rights to the work and has no comeback if he or she later wish to make such work open access and this was not initially agreed.

Under an **exclusive license to publish**, the author retains his or her economic rights, but signs away most of the practical benefits of so doing, usually for the entire term of the copyright. In this form of license, the publisher has the right to publish and make money from the work (and to act to legally protect those rights) and the author agrees never to give the specifically negotiated rights to anybody else. The reason that this mode has emerged is that it is marginally more beneficial to authors. If a specific type of publishing is not covered under an exclusive license to publish (for example, distribution in certain regions or formats), the author can renegotiate for new, better terms at a later date if the publisher (or another publisher) wished to

then undertake these activities. This may financially disadvantage a publisher whose contracts are not sufficiently 'future proof' or it may simply mean that an author is free to publish the work in other formats elsewhere, depending upon the contract. It also allows publishers who know that they do not wish to operate in a certain sphere to give the author the opportunity to exploit this aspect elsewhere.

With a **non-exclusive license to publish**, the author keeps his or her copyright but gives the publisher the right to publish the material. However, the author retains the right to license others to do the same (or to make the piece public on his or her own initiative). This is seen by advocates as a good step for open access as the author will retain the right to deposit his or her work and to republish it wheresoever he or she chooses without requiring publisher dispensation. Of course, it also gives less favourable terms to publishers who need to ensure their economic return on the labour invested (see the remarks in Chapter 2 on the co-existence of green open access with sales/subscriptions, though, for reasons why this may still be possible). This mode does not, however, permit reuse of material beyond fair dealing, as detailed below.

Open licenses

Open licenses, which fulfil the lowering of permission barriers enshrined in the BBB definitions of OA, come in a variety of forms, but the most common for scholarly articles and books so far have been those designed by the Creative Commons Foundation, which have proved enforceable in courts of law worldwide.[8] The second most commonly used open text license is the GFDL (the GNU Free Documentation License), which was Wikipedia's choice until May 2009 (when it was then superseded by the Creative Commons Attribution-ShareAlike License).

Those who would like to know much more about Creative Commons licensing for humanities researchers (particularly on a practical basis) may wish to consult the *Jisc Collections Guide to Creative Commons Licensing for Humanities and Social Science Researchers* from which much of the information in this section is derived, which is itself made available under a Creative Commons Attribution license.[9] The other source for the information provided

here is the Creative Commons site itself, also available under a CC Attribution license.[10]

The Creative Commons organisation provides seven mechanisms through which creators can allow others to use their work more permissively. The absolute, most liberal of these is the CC0 license, which 'enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law'.[11] Because it is unusual for scholars to wish to waive their right to demand attribution – and because, as I have already discussed, economies of prestige within the academy function as core drivers of academic output – I will primarily deal with the six other Creative Commons licenses that all carry an 'attribution' clause. That said, and as I will discuss below, academic citation norms and anti-plagiarism measures are so strong that even work under a CC0 license would probably not be subjected to misuse within the academy.

Beyond CC0, then, there are six, core, Creative Commons licenses, each with its own (at first) perplexing acronym: CC BY, CC BY-NC, CC BY-SA, CC BY-ND, CC BY-NC-SA and CC BY-NC-ND. The 'CC' clause in each case stands for 'Creative Commons', clearly enough. The wording 'BY' in each of these phrasings is not an acronym but literally means 'by'. Anybody using works licensed under these provisions with the 'BY' clause must give attribution, citing the original, and specifying *by* whom it was created. The modifiers then stand for 'Non-Commercial' (NC), 'ShareAlike' (SA) and 'NoDerivatives' (ND) respectively. As is clear from the above list, these modifiers can, in some circumstances, be compounded.

In order to explain what each of these licenses means for the licensor (the author, in this case), it is worth reproducing with minor modifications a table that can be found on page nine of the Jisc Collections guide (see Table 1).[12]

Note that the only two incompatible clauses are ND and SA; there is no way that a ShareAlike clause can apply if the NoDerivatives directive is also present as this would be nonsensical: there is *no* future *derivative* on which to compel sharing under the same license. Applying one of these licenses to a piece of work is as simple as writing a line of text

Table 1 *The Creative Commons Attribution Licenses*

License Designation	License Name	What does this mean for you as an author?
CC BY	Attribution	The most liberal of the Creative Commons licenses apart from CC0 Public Domain Dedication. This license allows others to distribute, remix, tweak, and build upon a work – even commercially – provided they credit the author for the original creation and clearly indicate that changes were made to the work.
CC BY-SA	Attribution ShareAlike	Similar to CC BY; however, others must license new creations under identical terms. Therefore, all new works based on such work will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia.
CC BY-ND	Attribution NoDerivatives	This license allows for redistribution, commercial and non-commercial, provided it is passed along unchanged and in whole, with credit to the author.
CC BY-NC	Attribution Non-Commercial	Similar to CC BY; however, others must not remix, tweak, or build upon the original work for commercial purposes. Although new works must also acknowledge the author and be non-commercial, reusers do not have to license their derivative works on the same terms.
CC BY-NC-SA	Attribution Non-Commercial ShareAlike	This license lets others remix, tweak, and build upon the author's work non-commercially, provided they credit the author and license their new creations under the identical terms.
CC BY-NC-ND	Attribution Non-Commercial NoDerivatives	This is the most restrictive of the six licenses, only allowing others to download works and share them with others as long as they credit the author, but they cannot change them in any way or use them commercially.

specifying the conditions under which the work may be reused and distributed, as can be seen on the copyright page of this book. Placing work under a Creative Commons license is an irrevocable act; one cannot rescind the rights one has bestowed on others after the fact.

It is important to reiterate, as above, that these licenses do not replace or abolish copyright because, without copyright, an author could not make any claim over the work, including the right to attribution. Furthermore, when the copyright term expires, the license will no longer hold any binding force; the material will enter the public domain. Indeed, the Creative Commons licenses 'are, in fact, built on copyright and last for the same length of term as the copyright in the work'. The Creative Commons foundation believes that this gives a sense of freedom back to authors, noting that such licenses 'enable you, as an author, to specify the conditions of re-use that best suit your needs, while ensuring that you are credited for your work'.[13] As will be seen, others disagree with such an assessment and find the rhetoric of 'enabling' reuse, couched in terms of 'freedom', to be misleading, particularly when funding agencies require that researchers apply such licenses to their work.

As noted in Chapter 1, these licenses – and particularly the clauses that allow modification of work – derive from a history in computer science and open-source programming cultures. It is worth saying, however, that the contexts are slightly different, which may have a bearing upon the rationales for open licensing in the humanities as opposed to computer science. With a piece of computer software, there are usually two different aspects: the source code (which is text that can be read) and the compiled binary (which is the version that can be run). The process of authoring a program is (usually – there are exceptions) to write code (a series of instructions that tell the computer, sequentially, what to do) in a high-level language that resembles words and instructions familiar to speakers of the English language. These instructions are then fed to a 'compiler', which translates and optimises them into an object code (usually assembly or machine code), an extremely low-level format that is difficult for people to understand, but easy and quick for machines to execute. The important point to note, however, is that it is extremely hard, albeit not impossible, to change the behaviour of the program or to understand its workings without the original source code. It is also

not easy automatically to change a program back from its compiled form to its source code.

This explains the importance of open source (or 'free software' as he prefers it to be termed) within Richard Stallman's philosophy of computer science.[14] In a world where we are surrounded by technology and somewhat at the mercy of software, he would argue, the obfuscation introduced by the compiler is a potentially powerful tool for control. This is because there is no easy way for others to audit, alter or fix the behaviour of important software (such as, for example, speed cameras that automatically fine people; bank machines that change our balances; credit scoring algorithms that determine whether one can obtain a mortgage to buy a house; life support machines; the list goes on).

In terms of academic research, however, and particularly that produced in the humanities disciplines, there is a different set of considerations as to why researchers might make derivative works, in which originals are altered or incorporated into another piece. Examples of derivative works in the humanities include any outputs featuring quotations or images from other sources, translations, scholarly editions and new presentations (such as digital humanities projects featuring XML encoding). In small quantities, some of these activities (such as limited quotation for the purposes of criticism and review) are permitted under law without any changes, the so-called 'fair use' or 'fair dealings' provisions. Other uses, such as inclusion of images, have far stricter criteria for fair use as the reproduction is often total, thereby obliterating the financial protections of copyright. Indeed, then, although open licensing may have emerged from the free culture movement in computer science, it is unclear as to how far the analogy to source code can be stretched in the applicability of open licensing to humanities work. For instance, the writing (or drawing, or performance, or whatever form the work deploys) within a piece is equivalent to its source code, for the software analogy. By altering these constituent parts and their arrangement, one can change the function and effective work of a piece of research, which may be undesirable or may hold value (and the perspective on this may be viewed differently in each case by different parties). Derivatives can be grand or minor in scale, drastically altering or only subtly recontextualising existing work. Within the academy, researchers already make use of the notion of derivative works when they cite the research of others, be this in the use of

ellipses, changes of emphasis ('emphasis mine') or in the inclusion of images. The centrality of such inclusions also varies in scope/scale and spans a range of types of producer, from other academics to artists/performers and beyond. Nonetheless, these are derivatives in copyright terms that are allowed, at a small scale, within the bounds of fair use. Other cases that are desirable for the academy, such as inclusions of larger portions of material in course packs (which is then potentially a derivative work), as set out below, may not be considered fair use.

The case for open licensing in the humanities, then, is substantially different from its historical context in computer science. For the humanities, open licensing should be less about the rhetoric of liberation of data/code and the attachment of 'the language of personal freedom ... to information', as Paul Duguid points out, and more concerned with potential use cases.[15] In other words, this should not simply follow the business mantra that 'information wants to be free' but should instead be predicated on whether existing copyright provisions are adequate easily to allow activities desired by academic researchers. Advocates of open licensing claim that they are not.

ARGUMENTS IN FAVOUR OF OPEN LICENSING

Among the first questions that must be considered are whether and why open licensing might be required or desired. As I will suggest, below, there are multiple areas in which advocates construe benefits and this section presents arguments from that perspective. The foremost of these, however, are the assertions in recent years that the current system of copyright is actively preventing scholarly research from fulfilling its potential. For instance, a report by the Ad Hoc Committee on Fair Use and Academic Freedom in 2010 for the discipline of communication studies noted, of their survey research, that:

Nearly half the respondents express a lack of confidence about their copyright knowledge in relation to their research. Nearly a third avoided research subjects or questions and a full fifth abandoned research already under way because of copyright concerns. In addition, many ICA members have faced resistance from publishers, editors, and university administrators

when seeking to include copyrighted works in their research. Scholars are sometimes forced to seek copyright holders' permission to discuss or criticize copyrighted works. Such permission seeking puts copyright holders in a position to exercise veto power over the publication of research, especially research that deals with contemporary or popular media.

These results demonstrate that scholars in communication frequently encounter confusion, fear, and frustration around the unlicensed use of copyrighted material. These problems, driven largely by misinformation and gatekeeper conservatism, inhibit researchers' ability both to conduct rigorous analyses and to develop creative methodologies for the digital age.[16]

As a simple preliminary finding, this gives just one example of a set of difficulties to which open licensing could pose an easier solution than changes to international copyright law.

Within different spheres of endeavour, open licensing is claimed to have varying degrees of potential. It seems fair to say, however, that there is not a single researcher who would not benefit in at least one fundamental way from even the more restrictive forms of open licensing (such as CC BY-ND). That is, without open licensing, even if one were to have monetarily *free access* to an article or book, this does not entail *permission to redistribute* that material beyond the basic provisions of fair use. Every year, universities pay to redistribute photocopies of critical material, produced by academics, to their students. This is because, for instance, despite the fact that this is use for the purposes of teaching in an educational establishment, in the UK 'Making copies by using a photocopier, fax, and so on, on behalf of an educational establishment for the purpose of non-commercial instruction generally requires a licence from the Copyright Licensing Agency.'[17] Organisations such as the UK's Copyright Licensing Agency and the US's Copyright Clearance Center act as mass collection agencies, requiring licensing agreements from universities in order to use, in many cases, material written by their own scholars and imposing limits on the amount that can be used for teaching in such cases. Furthermore, these agencies often require universities to re-purchase material that they already own, simply so that it can be reprographically distributed to students. The CLA's HE license states that, '[u]nless there are valid pedagogical reasons for using a superseded edition, all copies should be made from the current published edition' and economic hardship of one's library

at having to buy the latest edition of a work that one already owns is not a valid pedagogical reason.[18] This seems to affect all those who teach in higher education and provides a good rationale for at least the more restrictive forms of open licensing that permit redistribution as a minimum of open access.

Certain other fields of endeavour within the humanities benefit differently under open licensing. Peter Suber lists some of these benefits as the abilities:

- to quote long excerpts
- to distribute full-text copies to students or colleagues
- to burn copies on CDs for bandwidth-poor parts of the world
- to distribute semantically tagged or otherwise enhanced (i.e. modified) versions
- to migrate texts to new formats or media to keep them readable as technologies change
- to create and archive copies for long-term preservation
- to include works in a database or mashup
- to make an audio recording of a text
- to translate a text into another language
- to copy a text for indexing, text mining, or other kinds of processing

all of which are impossible under most 'fair use/dealings' provisions.[19] While it should be clear from even a cursory glance at this list that every single one of these items might be applicable to the humanities disciplines, some are especially pertinent. Of especial note are: the possibility to quote longer excerpts and include academic images; the ability to translate texts; and the ability to text-mine works for digital humanities projects. The remainder of this section will be dedicated to exploring these exemplar use scenarios.

Use beyond 'fair dealing'

One way among many of conceiving of research work in the humanities is as an argument/refutation dialectic between scholars. Under such a formulation, research work is supposed to be an ongoing effort of communicating in public to negotiate on areas of contention in order to reach a shared truth or understanding. Given that

this is the case, it can be surprising to see how little engagement there often is between scholars in print. This is jointly driven by incentives for originality and, on occasion, by a prohibitive copyright situation. On the first front: very few scholars get ahead by spending their time critiquing the arguments of others at length. Indeed, even more so in a culture where public 'impact' is becoming important, greater emphasis is placed on working from scratch than from pausing on the work of others.

That said, assuming that one did wish to mount a piece of substantial, engaged criticism of an academic's work, it is unlikely that the current system of copyright would be amenable. Indeed, to reproduce anything more than the bare minimum will place such use outside of 'fair', particularly if this represents a substantial portion of the original (even if required to make a point). Line-by-line critiques are, therefore, out of the question. Epigraphs from scholarly work could likewise be prohibited (this is contended but an increasing trend to prohibit epigraphs has been seen among some academic publishers in recent years, under counsel that this may not constitute fair use).

The images that academics produce are even more thorny. Anyone working in the discipline of art history will be able to attest to the enormous problems and costs in securing the rights to image reproduction. In fact, fair use provisions are applied differently and more stringently than images under some copyright jurisdictions. Because academics currently retain or transfer their copyright, the same goes for images produced by academics, even if these particular images/illustrations/photographs were given away for free. Although it is clear that the ability to relicense images produced by academics provides another revenue stream for academic publishers, advocates question whether this is desirable given the difficulties of image inclusion. Furthermore, some authors have argued against liberal open licensing (and open access more generally) of their own works on the grounds that it will make it harder for them to include images that are under copyright. In fact, this problem only applies to the dissemination of work as gratis OA (work that is free to read) and is not a problem of libre OA (work that is also openly licensed). This is because it is possible, when openly licensing work, to exclude third-party images from the license provisions, thereby allowing the original copyright holder to continue to license their work.

A final use case can be seen in the broader dissemination and amplification of research work that could be possible through resources such as Wikipedia, if open licensing provisions were in place. Regardless of whether one favours the anarchic construction of this online encyclopaedia (and irrespective of the quality controls in some areas), it is a remarkable resource and the first port of call for many lay readers who wish to learn about a topic. While it is already possible to quote portions of research works within Wikipedia under fair use provisions, to extend this reuse to include larger portions of work, or even whole articles, would give a far more visible presence to humanities research in a popular, public space. While some will remain wary of Wikipedia, the potential to incorporate research work within similar ecosystems will be far easier if compatible open licensing provisions are adopted.

Translation

At present, English dominates scholarly discourse.[20] In a networked world, this is a huge challenge as, in the quest for practical solutions to overcome language barriers, the risk of erasing cultural specificity is omnipresent. To date, the mutually exclusive options to militate against this have been authorised translations or neglect. The question then becomes one of canonisation: which forces allow authorised translation, what are their motivations and who is allowed to translate? The answer, in most cases, will contain at least some degree of commercial interest for works that are within their copyright term.

This is where advocates claim that open licensing could help. To return to my previous argument from Weber, humanities communities should be at least partially concerned with plurality and the communication of difference. By giving permission, in advance, to anybody to translate a work, through open licensing, a greater degree of plurality could emerge, it is argued.

Such arguments have emerged implicitly and explicitly from the work of John Willinsky and Kathleen Fitzpatrick. In the former of these arguments, John Willinsky proposed that, in the online environment, acts of reading should be supplemented by technological 'helpers' that provide side-by-side context: contextual reading.[21] These helpers could, in Willinsky's view, give information on

external references, allow quick lookups of words and many other features. It is not a huge leap to extend this to a mode whereby these helpers might also provide inter-lingual information and contexts. When thinking about translation, however, some have argued that it could be possible to extend this to the other extreme and to perform contextual *writing* on other documents. Arguments of this nature can be seen in Kathleen Fitzpatrick's deployment of Chris Kelty's notion of a 'recursive public' to illustrate how communities could become involved in 'working towards a common goal' that is 'focused on improving the communication systems that fosters its [that community's] work'.[22] In Kelty's terms, a recursive public is '*a public that is vitally concerned with the material and practical mainten- ance and modification of the technical, legal, practical, and conceptual means of its own existence as a public*'.[23] A recursive public would certainly be one that built a system whereby community translation efforts were both technologically and legally possible. Indeed, trans- lation that uni-directionally privileges English, it could be argued, is not truly respectful of the difference with which humanities communities can be said to be concerned and could be seen as an imposed controlled phenomenon, rather than a democratically recur- sive formation. Instead, if it is believed that there should be space in the university for critical thinking reflexively to consider academic practices (as a recursive public) and if it is thought that the preserva- tion of difference is to be valued, then, in one fashion, allowing bi-directional community translation of works through open licens- ing could begin to achieve this.

The counter-argument, of course, centres on the problem of bad translation. What is to be done in the instance of an incorrect translation? Often, the author himself or herself cannot evaluate whether the translation is correct or of quality and there is a fear of negative reputational association. Thus, while Sandy Thatcher points out that the CC licenses prohibit reuses that would 'distort, mutilate, modify or take other derogatory action in relation to the Work which would be prejudicial to the Original Author's honor or reputation', he also correctly notes that 'This provision might provide grounds for action against an intentionally bad translation, but not just a poor one innocently done.'[24] Such debates, however, already rage around professional translation of scholarly material. For just one single

instance, Mark Philp contends that there are problems in the translations of some of Michel Foucault's best-known works, noting that Alan Sheridan's translation of 'rapports de force' as 'relations of power' leads to a circular definition of 'power', perhaps far from Foucault's intention.[25]

None of these arguments should serve simply to denigrate the skills of translators or even to point out the fairly obvious fact that academics sometimes dispute such translations in the services of their own arguments. It is rather to note that there is a vast corpus of material where no professional translation exists (or will exist) and that advocates argue that some translation could be better than no translation in such cases. If coupled with technological measures to ensure that bad translations could be vetted and/ or rated, the prospect of community translation might be exciting. There are, however, reputational challenges and fears that would have to be overcome before this could realistically become a mass proposition, not to mention the problems of incentivising such activities and of ensuring fair remuneration for professional translators.

Text/content mining and experimental re-presentation

Text and content mining are computer-aided techniques for sweeping a large corpus of material and looking for links and trends (or, in fact, simply for finding relevant information). The most well-known exposure of these techniques is Google's 'ngram viewer'.[26] An 'ngram' is a series of linguistic attributes ('n' refers to an arbitrary number and 'grams' is simply a shorthand for the grammatical portion of text/speech under discussion, which can be a phoneme, a word, a letter etc.). Google's viewer presents a search interface for trawling a large section of their scanned book corpus that makes it very easy to spot the emergence and correlation of various terms. As has also been made abundantly clear, though, through the proliferation of lawsuits against the service from copyright holders and their representatives, the legality of Google's practice is hotly disputed as claimed fair use.

In the biomedical sciences and other scientific disciplines, it is clear why text and content mining is important. As the volume of

literature grows, it becomes necessary to trawl for existing research that may have a bearing on one's own work, but which lies within an entirely different sub-field. While the same is true for the humanities, it could be argued that nobody is actively harmed compared to a situation where, for instance, previous clinical data indicating a danger to life remained unearthed. That said, just because 'nobody will die if we can't adopt text mining' hardly seems a brilliant rationale, in the eyes of advocates, for leaving this route unexplored.

Text mining offers novel ways of exploring an academic corpus. For instance, should one wish to trace the historical genealogy of a specific concept, it becomes possible to see how ideas enter broader circulation. These techniques also offer the opportunity to search in ways beyond those implemented by publishers. While publishers have a good rationale for ensuring that people can find the material that they have published, it also makes sense to allow others to be able to create discoverability and readability experiments. Indeed, eLife's 'lens' software – which offers an innovative new presentation layer for already-published material – is one such example. The lens viewer is a system to re-present the same information in a published article within a new interface that focuses on aligning multimedia and reference elements alongside text so as not to distract from the reading experience in an exceptionally aesthetically pleasing fashion. In order to allow those who are thinking about such issues to have the chance to *try* new experimental technological presentations and data collection, open licensing is necessary, despite some jurisdictions loosening the prohibitions on such activities within fair use provisions. The computational techniques provided by text mining will not be of use to all humanities academics but they will be of use to some, especially if the broad field of digital humanities continues to grow at its present rate.[27]

These aspects constitute some of the reasons why advocates believe that it would be beneficial to apply open licenses to academic work in the humanities. It is not a comprehensive list but it does give a flavour. However, sceptics argue that there are risks that come with open licensing – and particularly the more liberal forms of the Creative Commons licenses – and it is to these arguments that I now turn.

THE CHALLENGES OF INTEGRATING OPEN LICENSING INTO THE ACADEMY

In terms of controversies pertaining to open access and the humanities, Creative Commons licensing has aroused fervent opposition. As outlined in Chapter 1, and beyond the economic arguments in Chapter 2, the main oppositions to CC licensing take two forms: concerns over scholarly integrity and broader worries about undesirable activities that could be enabled by such lowering of permission barriers.

In this section, I want to spend a little more time airing these claims and evaluating the dangers that could arise from the implementation of CC licenses. It is worth noting, upfront, however, that personal preference for a specific license may not, in the end, be a choice that rests with most authors. Many funding councils have mandated Creative Commons licenses for work that they fund. The EU's Horizon 2020, the UK's RCUK and the Wellcome Trust, as notable humanities funders, already have mandates for forms of open licensing.[28] Given the Australian Research Council's move towards open-access mandates, one could speculate that a licensing condition will only be a matter of time for this funder also. Authors who dissent from Creative Commons licenses are then faced with the (non-)choice of deciding whether they will accept funding from a particular source (under its rules) or spurn the funding in favour of principle. Given institutional pressure to attract funding, it seems unlikely that many will stick to their guns. That said, a recent survey by Taylor & Francis also indicated that authors remain extremely wary of CC BY, an aspect reiterated by Wiley, although advocates continue to criticise the methodology of these surveys.[29]

Concerns over scholarly integrity

The first worry regarding Creative Commons licenses concerns the integrity of academic research material and the author's moral rights. While some see the ability to rework material as a benefit, others think this a problem. Without a No-Derivatives clause, each license is *designed* to allow maximum reuse, including modifications to the language used. While this might seem strange, understanding some

of the logic around the creation of the licenses can help to provide a rationale. Lawrence Lessig, as the founder of Creative Commons, describes these provisions as circumventing what he sees as the unnatural provisions of copyright: 'The extreme of regulation that copyright law has become makes it difficult, and sometimes impossible, for a wide range of creativity that any free society – if it thought about it for just a second – would allow to exist, legally . . . I then want to spotlight the damage we're not thinking enough about – the harm to a generation from rendering criminal what comes naturally to them.'[30]

This is an interesting stance because Lessig clearly predicates his belief in free culture on the fact that creativity requires the reuse of preceding works. He also explicitly here signals that this desire to create and to build upon the work of others is, in his worldview, the *natural state* of humankind. Richard Stallman often uses exactly the same logic: 'people have been told that natural rights for authors is the accepted and unquestioned tradition of our society . . . As a matter of history, the opposite is true. The idea of natural rights of authors was proposed and decisively rejected when the US Constitution was drawn up. That's why the Constitution only permits a system of copyright and does not require one; that's why it says that copyright must be temporary.'[31] At least part of the controversy over open licensing can probably be attributed to different ideas of natural and moral rights with regard to copyright.

It is also within such contexts that claims for protection of scholarly integrity should be considered in the humanities. Clearly, it remains important that protections against libel or utterly false attribution remain and all of the Creative Commons licenses continue to provide these. Rather, however, these provisions are thought of in terms of allowing others to build upon and modify scholarly works to create new versions. The analogy that Lessig uses for this is another from technology: he suggests there is a paradigm of 'read only' (RO) culture and a coming wave of 'read/write' (RW) that is to do with democratic participation in production.[32] While critics often argue that people should work creatively from scratch, advocates would counter that most humanities work is already based upon the scholarship of others and 'derivative' readings of culture/history.[33] Furthermore, that 'critical editions' of certain texts are already

created does show, to some extent, that derivative practices already exist in the humanities. With this background context out of the way, however, it is worth turning to some of the specific concerns that fall under this heading.

Among these is the claim that open licensing promotes plagiarism.[34] This has been put forward in a number of forms. The first is an outright accusation of facilitating plagiarism and the second is a version of the argument that notes that the derivatives clause encourages activities that 'look like' plagiarism. To explore such claims, an accurate definition of 'plagiarism' is needed. Plagiarism is a specific form of copyright infringement with a legal meaning that informs, but that is separate from, the intra-academic contexts. Plagiarism usually refers 'to the subcategory of copyright infringement which involves false designations of authorship and other unattributed uses of copyrighted material. This is usually distinct from the other common subcategory of copyright infringement called "piracy," which involves the production and sale of unauthorized literal copies of a work.'[35] In other words, plagiarism involves making a claim to authorship of a work created by someone else.

Plagiarism is enforceable both in law and, also, within academic institutional contexts, where the penalties can be severe. The intra-institutional context is slightly different from the purely legal realm. Because the academy often uses fair dealing, or fair use, provisions to allow the reproduction of small portions of works upon which it performs analysis, there is a strict requirement to delineate quoted words from the author's words, which would not apply in other contexts where no degree of copying would have been permissible. There is also a requirement to represent accurately the original source (and its author) in both specific quotations and broader semantic terms.

The Creative Commons Attribution licenses grant specific additional rights in the legal realm but leave the academic institutional context unchanged. Just because additional reuse is possible under the law does not mean that academic citation practices will necessarily change. Whether legally allowing additional activities might eventually change the norms of the academic institutional context remains impossible to predict, however. These licenses allow anybody (except in the case of the Non-Commercial clause, which

restricts any non-personal reuse) to copy any quantity of the text that they would like, even beyond those allowed under fair use. They are also allowed to change the contents as they see fit. These licenses do not, however, allow this unconditionally.

Any user of a Creative Commons work must, 'expressly', retain an 'identification of the creator(s) of the Licensed Material and any others designated to receive attribution, in any reasonable manner requested by the Licensor (including by pseudonym if designated)'. In other words, a reuser *must* identify the original creator under legal penalty *in the manner chosen by the author* (within reason). This is designed to protect the moral right to be named and seems to avoid the risk of outright plagiarism. After all, by crediting the original author, one disclaims authorship. As an end reuser, it is also imperative under the terms of the license, if you have changed the work, that you 'indicate if You modified the Licensed Material and retain an indication of any previous modifications'. This means that it is crucial that an end-user indicates whether the work has been changed, a provision that is designed to protect the moral right to integrity of the original. It does not, however, mean, as Peter Mandler has rightly taken pains to reiterate on many occasions, that the author must legally show *what* changes have been made. That said, Mandler concurrently acknowledges that academics have their 'own norms of how best to incorporate one work within another', which he claims 'derivative use [under a CC BY license] denies'.[36] Mandler's concern seems to be that the CC BY license will interfere with academic norms because it allows (or perhaps even endorses) activities that the academy would not permit within its intra-institutional rules. What remains unclear within such an argument, at least to me, is why academic practice would change, simply because it is legally permissible to use the work differently; these two contexts can be different yet co-exist. After all, works whose copyright has expired (therefore holding none of these protections and which the law explicitly permits anyone to use in any fashion) are still subject to these intra-academic norms. Conversely, others have sometimes built valuable, digital, scholarly projects around such works; enterprises that would be practically impossible without permission to modify the original. For just one example of this, one could consider the Nietzsche Source project,

which aims to create a freely accessible, citable, reliable XML-encoded version of Nietzsche's works.[37] As another instance, one might also consider the liberal rewriting of academic material into a Wikipedia article, always with citation of course. A more extensive list of such projects can be found on the website of the Text Encoding Initiative.[38]

It is also important to note that nothing in the CC BY licenses 'constitutes or may be construed as permission to assert or imply that You [as a reuser] are, or that Your use of the Licensed Material is, connected with, or sponsored, endorsed, or granted official status by, the Licensor'. In other words, a reuser cannot imply that the original author (the licensor) condones the subsequent reuse of the work. Finally, as a reuser, 'if requested by the Licensor, You must remove any of the information required [to attribute the work] to the extent reasonably practicable'.[39] This means that the licensor (the original author) may request that their attribution be removed and thus they be disaffiliated from the work. The licensor (the original author) cannot, however, request that the work be taken down.

For the academy, this is interesting. Broadly speaking, the requirements of the CC BY licenses are: (1) attribution (without implying endorsement); (2) indication of modification; (3) the right to removal of attribution. This sounds similar to the definition of the needs of a researcher. Researchers need to be able to reproduce material and they need to attribute it. They also need, through the intra-institutional context, to specify any changes, including any ellipses, changes of emphasis etc. On this last point, the legal aspects of the Creative Commons license are not wholly in alignment with the social needs of the academy since the CC BY licenses do not require the modifier to say *how* they have changed the material. The licenses also do not allow for the material to be removed if the academy objects (although the alternative to this situation is one in which an author could censor critique through legal copyright mechanisms). It is unclear, however, whether this matters if the social mechanisms of the academy could protect against such behaviour. It is also uncertain what the likelihood and impact of such actions might be when counterpoised with potential benefits.

Concerns over undesirable reuse

The second broad category of concern over CC licensing pertains to undesirable reuse. This generally takes two sub-categories of its own: politically undesirable reuse and commercially undesirable reuse, although in the case of the latter, particularly, this is a case of: 'undesirable for whom?' Let us begin with politically undesirable reuse.

In an article at the extreme fringe of such claims, but one that can be applied more moderately, Robert Dingwall makes the accusation that 'open access is good news for neo-Nazis'.[40] In a fine instance of Godwin's law – a humorous axiom of the internet that 'as an online discussion grows longer, the probability of a comparison involving Nazis or Hitler approaches one'[41] – Dingwall hypothesises a paper about a (fictional) group of neo-Nazi racists:

> This might well result in a journal paper which demonstrates that the group's members are not demons but ordinary men and women responding to economic and social challenges with strategies that seem reasonable to them, even if based on partial information or analysis by others' standards . . . For the author, the paper presents evidence that it is unhelpful to dismiss these people as bigots: the political system needs to recognize and address their grievances, without adopting their racist solutions. With a CC-BY [*sic*] licence, however, nothing stops the group taking hold of the paper, editing it down and using it as a recruitment tool: 'Famous professor says we are just ordinary people responding in a reasonable way to the problems of our community . . .'[42]

This argument does not seem particularly solid. For one, such a group could likely attempt a claim under 'criticism and review' fair dealing provisions anyway, as do news outlets (it would be libel for which they could be taken to court, not for their actual use of the material). While Dingwall's claims are perhaps too extreme to be credible, one might consider undesirable use by more mainstream parties, whether fringe or even moderate, and the problems that this could have for the neutrality of research work.

The second concern over undesirable reuse surrounds commercial appropriation. Building on the remarks in Chapter 2, this requires a little more time to unpack as it is, itself, split across two different axes: a wholesale rejection of any kind of utilitarian appropriation of humanities work, or an antipathy towards specific commercial

entities using university research. To deal with the first of these objections, it is worth noting, as Helen Small points out, that 'the spectre of trial by proven utility [which is necessary for any commercial appropriation] has accompanied universities in one version or another since they came into existence'.[43] Indeed, Small also ably points out that while the reductive language of utilitarian thinking ought to be resisted, the humanities contribute to society through: a 'distinctive relation to the idea of knowledge as being inextricable from human subjectivity'; in a provision 'of the skills for interpreting and reinterpreting . . . culture to meet the needs and interests of the present'; in their 'vital contribution to individual happiness and the happiness of large groups'; in the role of a clinician ensuring 'the health of the democracy'; but most importantly that 'none of these arguments is sufficient without a supporting claim that the value of the objects and cultural practices the humanities study and the kinds of scholarship they cultivate have value "for their own sake" – that they are good in themselves'.[44] In this light, it is certainly true that attempts to defend the humanities need not take a purely anti-utilitarian turn, even when that utility is within industry, so long as this does not remain the driving force for the investigation.

With this in mind, it is important to acknowledge that Creative Commons licenses do facilitate commercial reuse of academic work if the Non-Commercial clause is not present. Those who wish to counter this aspect of Creative Commons licensing have asked whether appending the NC clause to mandated licenses might pose a solution. This list of anti-NC advocates includes a set of twenty learned societies who form the Arts and Humanities User Group, whose license recommendation was CC BY-NC-ND.[45] Based on recent court decisions, however, I would suggest that this may not be the most helpful approach. As Erik Möller points out, the definitions of 'commercial' are unclear.[46] This confusion has been made worse by a recent German ruling that NC licenses must refer *only to private uses*, thereby disallowing universities, charities and other organisations from redistributing and using such material.[47] Given that many universities charge student fees and make money off research patents, they are very much commercial in their own rights, even if this is not how many of those within the walls of the academy would wish it. By appending the NC clause, universities potentially lock their own

communities out of the benefits of open licensing in a baby-with-the-bathwater situation. Furthermore, if derivatives are prohibited, so too are the benefits of extensive quotation, academic image inclusion, rewriting for popular consumption, community translation and longer excerpts for course pack distribution, as well as text mining for the digital humanities in jurisdictions where there are no legal exemptions.

Given these controversies, it is unsurprising that it has been suggested that Creative Commons licenses are unsuitable for academic research and that a new set of licenses is needed instead.[48] However, writing licenses from scratch is a difficult business as they require court precedents to be trustworthy and also have jurisdictional specificity that needs expert legal counsel worldwide. Furthermore, new licenses may be incompatible with existing CC licenses used elsewhere, which could, for just one example, make the inclusion of material within Wikipedia impossible. It is also clear that a proliferation of licenses comes with problems. For example, when the International Association of Scientific, Technical & Medical Publishers recently drafted its own set of licenses, Andrés Guadamuz noted that, aside from being non-compliant with BBB definitions of open access, these new licenses served to 'taint the open access licensing environment by generating more licence complexity and more confusion [for] the academic authors'.[49] A potentially better solution for those who would like the humanities to be less utilitarian, but who also recognise both that open licensing comes with some benefits and that potential industry collaborations need not always be resisted, could be to impose a ShareAlike (SA) clause on material. This would mean that industry would be under the same obligation to give back to the community as the original academic. Indeed, any new derivatives that were made by industry as a result of using academic research licensed under CC BY-SA would also have to carry that license, meaning that the academy could benefit in turn from any transformation of the work, if valuable.

This brings me, finally, to consider the role that CC licensing could play in the broader structural changes pertaining to the marketisation of higher education, a concern that has been mentioned several times already. Some figures, most notably John Holmwood, have argued that Creative Commons licensing of academic research

will fuel this process by allowing bodies that do no teaching, but which hold degree awarding powers, to alter the material for their own ends (compiling it into a sort of online anthology or textbook) and to thereby undercut the research university on cost, leading to its extinction. This represents a serious danger to academic autonomy and the financial cross-subsidisation of research from teaching. The standardisation and mass production of academic degrees, set by an external entity, is just another factor that would take the university ever further from its ideal as a community of self-organising scholars.[50] Although Holmwood does not specify exactly how he envisages that such providers could not already use such material under the guise of 'educational' fair use, or by simply buying the material – which they could afford anyway – his argument is that there have been progressively more aggressive policy moves towards commercialisation of the university, particularly in the UK (with which is hard to disagree). Indeed, he notes that 'universities are also enjoined to increase value for money for students through efficiency savings. Here the model is one of the "unbundling" of different activities, to identify those which can be taken to market by "outsourcing" and made subject to the proper rigours of the profit-motive.' However, in a parallel to education, Holmwood then goes on to note that 'Open Access under CC BY is one of the measures designed to speed up commercialization, by making scientific innovations more immediately accessible, especially to small and medium-sized enterprises.'[51] An analogous argument could certainly be made between the cultural industries and the humanities.

I do not dispute that this seems to tally with governmental agendas. For those who support open licensing but wish to counter such approaches, however, there are a variety of responses. Holmwood clearly supports the addition of the NC clause to licenses (his own book is CC BY-NC[52]), while I believe that ShareAlike would be a better solution to the same problem: ensuring that if others benefit from the public work of academia, it remains a public good.

In this section, I have appraised some of the most common objections to the CC BY licenses: plagiarism and undesirable reuse. I remain unconvinced that plagiarism is a strong argument, but this is not a view universally held. The latter arguments about undesirable

reuses, both political and damaging to the university as it currently stands, conversely, certainly have traction in some specific areas.

*

Open licensing is contentious and, in some senses, it is possible to have open access without open licensing ('gratis' versus 'libre'). However, advocates point out that many other benefits are available under open licensing, including: dissemination for teaching, extensive quotation, text/content mining and community translation. Advocates also claim that the copyright system as it currently stands often does not fulfil the needs of academics, acting as a barrier to research, rather than a protection. Conversely, others feel that the dangers of relaxing these provisions are too great. A variety of suggestions, from new licenses through to NC-ND clauses, have been suggested as compromises. However, the Non-Commercial (NC) clause of the Creative Commons licensing scheme might be less useful than it sounds as this will exclude the university from redistribution and use. The No-Derivatives (ND) clause allows redistribution but prohibits more radical experimentation. Writing new licenses is a difficult task that complicates the licensing environment and should not be undertaken lightly.

The final point that we might wish to consider before moving on is the place of economics within the contexts of open licensing. While it should be clear that the open licensing agenda intersects with my earlier remarks on the commodity form of research work and sites of use-value that are distant from the academy, a more immediate concern comes from a particular stakeholder group: publishers. There is very little reliable evidence to know, one way or another, whether open licensing damages the sales prospects of scholarly material in print form. As I shall show when I now move to monographs, this could pose problems for models in which print subsidy forms part of those particular economic sustainability arrangements.