Law: Friend or Foe in Scientific Internet Use?

Michael Beurskens

Faculty of Law, Heinrich Heine University Düsseldorf

While legal rules provide necessary protection to scientists on one hand, they impose significant restrictions on the other. Furthermore, science is governed by its own set of ethical rules. Many assumptions about the rules imposed from the outside are actually based on the more-or-less intuitive understanding of such ethics. This is a hit-and-miss-affair: Sometimes those rules meet, while they may deviate just as often.

Introduction

Lawyers and judges often seem to act as direct successors to priests, translating the utterings of the oracle of Delphi by explaining equally mysterious legal rules (Amsterdam, 1970). In contrast, academics serve a different religion based on other rites but with a similar goal of improving society. Often, law seems to provide unnecessary double precautions, as rules of ethical research seemingly already offer sufficient protection. Indeed, such interference in internal affairs by outsiders is intuitively frowned upon. Thus, it is unsurprising that scientists blame legal rules and their implementers for imposing unnecessary and annoying limitations on essential academic activities (Green, 1992). Because it is easy to shoot the messenger (Shakespeare, 1598), asking to kill all of the lawyers to achieve utopia is certainly not far-fetched (Shakespeare, 1623). However, further examination shows that those objections go both ways. Scientists complain about both “overregulation” and “underregulation” (Sunstein, 1990): What is there is not enough—and still too much.

This can be easily exemplified by looking at popular beliefs about intellectual property law. One tends to apply a strict perspective against others seeking to reuse the data gathered, content created, or inventions made by oneself (see the heated debate in the U.S. on granting copyright to the lecturer instead of the employing university; Triggs, 2005; Holmes & Levin, 2000; Townsend, 2003). In contrast, reuse of content made by others in teaching is considered a good practice (see aptly Lessig, 2001). To many scholars, it is counterintuitive to lose each and every right to their work when signing a publishing contract. Consequently, written agreements with publishers were and are often ignored when sending papers to colleagues or even putting them on Web sites (before self-archiving clauses became popular).

The internal system of ethical research practices at universities, in contrast, is well established and thoroughly administered by university committees and even includes written commitments and frameworks, which look and feel much like “laws.” Such rules are not limited to research on human or animal subjects; they also cover practices such as the attribution of works and data or the transparency and reproducibility of research (cf. Hudson et al., 2005 on an empirical approach to ethics in Internet Research).

When trying to determine the appropriate external (legal) rules, informal internal (ethical) codes of conduct are often confused with actual requirements imposed by law (for the lawyer’s perspective, cf. Cohen, 1934). In general, ethical means should be allowed without restrictions, whereas unethical practices should not only be frowned upon by colleagues but also sanctioned by law. However, legal rules and ethical beliefs do not necessarily go hand in hand. This may be due to questions still being debated in the scientific community or by the general public but decided by the legislators (e.g., stem cell research) or due to legal rules usually not being limited to scientists but having to cover a multitude of different cases.

Ethical rules are not the only guideline applied by laypersons trying to conform to legal requirements. Because the law itself is usually inaccessible, or at least incomprehensible, to the average researcher, they tend to fill gaps with assumptions based on information gathered from unreliable sources, including colleagues, public media, or the Internet. Again, this provides a broad fertile soil for errors—even though intuition is certainly an appropriate approach even in judicial decision-making (Wright, 2006).

If the applicable law were clear and certain and remained generally unchanged, then such confusion could be eliminated with mere education or information. However, especially with regard to the Internet, law is in a constant state of flux. Copyright cases produce unpredictable results, and fundamental differences across the national legal regimes provide an unstable basis
for the increasingly international research community (cf. Dellapenna, 2000). Indeed, unlike real property and contracts, neither copyright nor privacy law (or “data protection” law) is a required subject in the study of law, nor are they part of any bar exam. Thus, even most law professors and law students can shed little light on these topics. Similarly, it is in the best interest of a university’s administrative staff to eliminate any risk of possible liability by suggesting the most restrictive approach.

This paper will examine a few common misconceptions regarding legal rules, explain the current legal framework by giving examples from the U.S. and Germany, and provide a brief look into current developments. While the field is extremely broad, we will focus on copyright and privacy laws as the most fundamental issues involved in scientific use of the Internet.

**The Copyright Dilemma**

Copyright is one of the fundamental pillars of research. Specifically, the requirement of attribution serves a dual purpose. By prohibiting plagiarism, scholars can only acquire a reputation by publishing their own ideas. Furthermore, a good reputation may form a basis for reliance, thus avoiding the need to verify information as long as the source is known to do proper research (this is an example of Akerlof’s [1970] “market for lemons”).

However, the copyright granted to scientific articles is, in most parts, identical to the copyright granted to the latest hit single, blockbuster movie, or even copy of Microsoft Word running on your computer. Whereas composers, software developers, and movie directors usually seek monetary benefits from their works, scholars generally will not be able to make a living from the articles or books they write. In contrast, if only money is at stake, there might be a good argument against perpetual property in creative products. If the public has paid enough for the cultural enrichment provided by *Jurassic Park IV*, then it should go into the public domain and be available to everyone for free. Attribution of original works, in contrast, should be perpetual—because the ability to build upon another’s research is essential for scientific progress. This simple economic difference is not embraced by the legal system, which treats popular and scholarly works largely in an identical manner. Similarly, scientific use, as such, is not generally privileged in comparison to mere private use. Furthermore, economic reasoning, or even the ethical rules of science, does not necessarily lead to a proper assumption of what is allowed or what is not allowed in copyright law.

However, there is also some good news. Unlike many areas of law (including privacy law, which we will examine below), copyright law is subject
to a rather strict framework of international treaties and conventions, which are applied uniformly across most states around the world. Still, differences persist, not only in details but also in the fundamental approach to copyright cases. This disparity favors the international character of research and is thus a useful basis for a framework for science regulation.

Plagiarism

Plagiarism has become a heavily debated topic in Germany following the discovery that the dissertation thesis of a well-known politician was, in large part, copied verbatim from other sources without attribution (on “investigative crowdsourcing,” cf. PlagDoc & Kotynek, 2012). In the aftermath, universities invested heavily into both software and personnel to ensure early detections of fraudulent dissertations and exams.

Most cases of plagiarism are not illegal. The law and the ethical rules of plagiarism intersect imperfectly (Stearns, 1992). Indeed, our culture loves repetition—from re-telling and re-re-telling of popular stories (take a look at the Top 10 movies) to copying of designs for household goods (read up on the current legal issues between Apple and Samsung). Works may be part of the public domain, e.g., with an expired copyright, or the part copied may be so tiny as not to constitute an infringement. Indeed, it would impose an all but insurmountable challenge on most scholars (and even worse on students) to re-create everything independently on their own—reinventing the wheel once again would not further science. As scary as the word plagiarism may sound (coming from the Latin word plagiarus, referring to kidnappers of another’s slaves), it is not beyond possibility that two (great) minds reach the same idea and even use the same words to express it. Just coming to identical conclusions is no misconduct at all, while rephrasing someone else’s ideas is. Copyright law also fails when plagiarism only extends to someone else’s research data. There is no legal monopoly on mere facts. Furthermore, reusing my own work without attribution will only be considered a copyright infringement if I violate the license agreement I signed with the original publisher. “Self-plagiarism,” as such, is unknown to the law.

Although the law largely ignores plagiarism, the academic penalty could be no worse—it is “a capital offense, punishable by academic death” (Onge, 1988). This is especially problematic, as the scope of unwanted practice is highly unspecific, and there are no lawyers to resolve questions of interpretation. While there may indeed be clear-cut cases, especially involving students merely copying content from readily available online sources and handing it in as their work, things become extremely difficult in professional scientific
research. In describing a case of “scientific misconduct,” Daroff (2007) aptly states:

[...] many of the problems we encountered stem from author naivety, sloppiness, and the ambiguities involved in plagiarism and self plagiarism. One doesn’t have to be a flagrant sociopath to encounter charges of misconduct or breach of ethics. (p. 532, f. 3)

Indeed, it seems that plagiarism is much like pornography—you know it when you see it (Stewart, 1964—a comparison discovered by Stearns, 1992). The issue of plagiarism is not so much the use of another’s intellectual property but the subsequent fraud—basing respect by peers and students not on hard work and intelligence but on mere copying. This puts the reputation of academia as a whole in question, which is certainly unacceptable. However, because detection is slow and imprecise and has a high rate of mistakes, plagiarism often comes as a surprise. In contrast, violations of intellectual property as such are not always frowned upon by science ethics. For instance, it is a good and well-accepted practice to send digital copies of a paper to colleagues and friends, even when the publishing contract assigns such rights exclusively to the publisher.

The value of attribution in copyright law has historically been a major controversy between the Continental European droit d’auteur and the U.S. “copyright” systems. While the European approach is largely based on the author’s natural right to his or her creation and, therefore, emphasizes attribution, U.S. copyright law is more focused on a utilitarian perspective and thus does not provide in its rules on fair use a requirement for attribution. Nevertheless, by adopting the Berne Convention for the Protection of Literary and Artistic Works (in 1989!), even the U.S. is required to ensure that every author has

the right to claim authorship of the work and to object to any distortion, mutilation or other modification of, or other derogatory action in relation to, the said work, which would be prejudicial to his honor or reputation. (Art. 6 till Sec. 1)

Nevertheless, lack of proper citations should constitute copyright infringement, even though U.S. law remains silent on the requirement of proper attribution.

Sometimes, offenders try to provide a defense of “good faith plagiarism.” One reason often expressed is that the omitted sources are not worthy of a citation, as they are not sufficiently “scientific.” Such “non-academic” sources may include anything from blog entries, tweets, comments posted in a forum, text from Wikipedia, PowerPoint presentations, or lecture materials found on the Internet to popular fiction or even texts from a different disci-
pline. Inversely, authors defend their actions by pointing out that their work was not academic, e.g., a mere slide show, some lecture notes passed out to students, or a blog entry. The reasoning again does not refer to a legal defense but shows that the ethical standard is highly imprecise. While such a close definition of quotable sources is certainly erroneous, the other variant of good faith is far more compelling: Every scholar writing a paper (such as this one) has certainly read many prior publications, most of them not in parallel to writing but beforehand. Unless one has eidetic memory, it is highly unlikely that one will be able to attribute each and every idea to a specific paper, especially when many ideas are shared among most authors (and might thus be considered “public knowledge”). Still, that defense will not be accepted because it would otherwise create a loophole that proves difficult, if not impossible, to close.

The consequences of plagiarism are extremely harsh. From a legal viewpoint, it is surprising that even a tiny bit of plagiarism will spoil a complete paper, article, or book. While the unattributed contents may only constitute a minor part of the actual product (and are often only tangential to the core theory presented), evidence of a violation of the rules of good scientific practices allegedly eliminates any benefits the work might have had for the scientific community. The law might grant damages or (in rare cases) an injunction preventing distribution of the book in its current form—but it would still respect the author’s efforts. Copying only minor parts would almost never provide cause for such an injunction. The oft-feared criminal sanctions are limited to even more extreme cases (but cf. Green, 2002, who attempts to analyze the social rules of plagiarism by comparing them to the requirements of the legal rules on theft).

Occasionally, someone tries to summarize the rules of plagiarism in simple, clear “Dos” and “Don’ts” (e.g., Gerhardt, 2006). While requiring attribution of “borrowed content” and asking for “quotation marks” seem simple enough, cases of plagiarism are not determined based on the actual behavior of the suspected author (which cannot be reproduced) but merely on the written results. The aforementioned well-publicized scandal in Germany caused many scientists to apply an extremely stringent standard of care in writing articles and books. This fear is further enforced by academic organizations, universities, and non-governmental organizations, which have designed guidelines for scholarly practices that tend to take an extremely strict stance on plagiarism. Indeed, overly careful authors on legal faculties (especially graduate students) seem to be keener to find references to add to their footnotes than to develop innovative theories or conduct their own research. Much time is wasted ensuring that everything that might have been said by
someone else before is attributed to that person—even if the author did not copy that source but merely made a “parallel invention.” It may well be possible that some authors simply choose not to publish to avoid even the appearance of unprofessionalism.

Such over-care is as worrisome as ignorance of the issue. The search for plagiarism must not turn into a witch hunt. The core issue seems to be the lack of a proper code on what constitutes plagiarism and what does not, as well as the lack of a centralized instance to interpret such rules. Furthermore, the analysis sadly always happens ex post and is therefore inevitably subject to hindsight bias: It seems to be rather easy to find someone who had comparable or even identical theories on a certain subject, but proving that one copied from the other is only possible with certainty if the text or large parts have been duplicated verbatim. Because only very naïve students (much less full-fledged scholars) would go that far, there is a significant margin for error in determining plagiarism. Thus, an open debate is inevitable, which should mainly be focused on the specific disciplines and not on the science community as a whole. A mere technological approach is useless except for detecting the most ignorant of copyists. Generally, preventive measures (such as an ongoing dialogue during the writing of a thesis or seminar paper) are preferable to ex-post sanctions.

Re-use of Content in Teaching

In the good old days, everyone bought textbooks and read them. When Xerox invented the photocopier (in 1959), teachers (both in schools and universities) suddenly had the opportunity to create a specific selection of readings specifically catering to their lectures. Such “readers” quickly became a popular practice, and suddenly students received huge stacks of printed paper to read at home. When the Internet became widely accepted, teachers once again made use of the medium and “made available” recommended materials for download and printing (Lan & Dagley, 1999). And they lived happily ever after.

Sadly, things are never that simple in real life. Supporting self-learning by providing students with a broad selection of materials might well be a good or even commendable practice from an educational point of view. Indeed, even the Berne Convention contains a provision on the specific exceptions related to teaching:

It shall be a matter for legislation in the countries of the Union, and for special agreements existing or to be concluded between them, to permit the utilization, to the extent justified by the purpose, of [...] works by way of illustration [...] for teaching, provided such utilization is compatible with fair practice. (Article 10)
That exception has its roots in the original agreement of 1886. Still, the U.S. Copyright Act does not include a specific provision allowing for the reuse of protected works in teaching, although it covers such use under the general terms of fair use (ALA). EU-Directive 2001/29/EC expressly allows for exceptions for the sole purpose of illustration for teaching or scientific research, as long as the source, including the author’s name, is indicated, unless this turns out to be impossible and to the extent justified by the non-commercial purpose to be achieved; [and for] making available, for the purpose of research or private study, to individual members of the public by dedicated terminals […] of works […] contained in their collections. (Article 5)

The domestic laws implementing these exemptions vary widely (Xalabarder, 2009; Xalabarder, 2004; Ernst & Haeusermann, 2004). Furthermore, their wording is surprisingly unspecific, referring to “small parts” and “necessity.” Thus, teachers, both in schools and universities, are largely left out in the rain. Because actual cases in court are rare, and procedures usually take many years, uncertainty prevails. Nevertheless, the number of documents made available to students has been on a steady increase. For example, the University Library in Düsseldorf currently scans approximately 5,000 excerpts (mainly from books) on behalf of lecturers for roughly 600 lectures. These numbers do not take into account documents scanned directly by lecturers and their staff, as well as materials taken from online sources and converted into PDF—the University’s central e-learning-platform receives more than 30,000 new documents per year (which include slides and lecture notes created by the professors themselves). Similarly, there is an extremely high demand by students for relevant materials beyond PowerPoint slides; in a recent survey, more than 80% of the students at the faculty of law giving an answer considered the direct availability of such materials as “very important” for their studies.

Still, the mist of uncertainty remains—so how do lecturers and university or the library administration cope with the risk of infringing on someone’s copyright? Discussions with colleagues at Düsseldorf University show that the answers seem to be highly dependent on the actual field of research. Surprisingly, teachers at the faculty of law show the least concern and the strongest belief that their actions are covered by exceptions to copyright law. In contrast, teachers at the medical faculty are very reluctant even to talk about the use of content created by others, as there is a strong belief that any such reuse would be deemed illegal. In general, most university educators assume some kind of liability shield, even though they are unable to specify
the scope. Still, the mere belief is sufficient to cause them to provide quite a lot of reading to their students.

The policy conflict underlying content reuse in teaching is evident in states in which education is predominantly funded by the government, as in Germany (see generally Chon, 2007). Thus, states are required to pay copyright owners (or more precisely the privately held publishers) to ensure appropriate means for education. By granting exceptions to such rights, they save real cash. In contrast, they also fear that too broad exceptions might be detrimental to scientists working in their institutions, as they would be sanctioned by being unable to present their ideas in printed publications.

German law provides an excellent illustration of the issues involved. Under current law, it is illegal to provide printed copies of another’s text to students (apart from exams) or even to send them texts by e-mail (whereas an exception allowing such use exists for schools, see 53 Section 3 of UrhG, i.e., the German Copyright Law). However, a well-meaning legislator created a specific exemption to the act for making short texts or short excerpts available to a specific group of students (identified by individual passwords) that actually participate (i.e., physically sit) in a course, as long as the provision of such texts is “necessary” for the purpose of teaching (see 52a Section 1 of the German Copyright Law). The only court decision available assumes that neither downloading nor printing the document is “necessary,” as reading it online on screen is deemed sufficient. Furthermore, German law requires “fair compensation” for any use made; this, in turn, requires universities and libraries to take into account any use made as a flat fee agreement being declined by the relevant collection societies. Further limitations are laid down in a 2003 “Charter” of the German Library Association and the Publisher’s Association (Börsenverein), which excludes the use of any content made available on the university network by the publisher itself, imposes a requirement to delete any content as soon as the course ends, limits the content to be made available to texts actually used in class (and thus excluding any articles meant to allow a student to expand his understanding), and only allows libraries to scan books that are available locally.

Still, most university professors in Germany make available any texts they consider useful. As mentioned before, legal cases are extremely rare and are most often filed against the university and not the individual teacher. The reason is twofold: Universities provide deeper pockets, and the state or the university will be required to indemnify anyone who did not act with “gross negligence.” Such extreme lack of care will almost never be provable. Although the available texts go beyond the allowed threshold, there is no evidence that the introduction of the exception in 2002 has actually caused
any loss of income to the relevant publishers (indeed, their profits increased by 20%, cf. Bibliotheksverband, 2012).

As in the case of plagiarism, the law again fails to fulfill the hopes or even assumptions of its academic addressees. While the issue of plagiarism is simply ignored by law and left to science ethics, the rules allowing lecturers to make texts available to their students are unable to provide for useful learning experiences. Thus, legal rules would actually limit science in one of its core purposes—if scholars knew about those rules and obeyed them. Although some professors are evidently concerned about possible violations of copyright, many still infringe and hope for the best, or at least for secrecy. Indeed, granting students access to important source texts may well be “ethical,” even though it is still “illegal.”

**Open Access and Creative Commons**

In recent years, the debate on “Open Access” to academic articles has gathered significant impact. This is largely due to some state and private funding requiring the publication of results under Open Access licenses. Nevertheless, traditional journals not only prevail but even thrive. Open Access is remarkable in the context of this paper for two reasons. First, its supporters try to adapt copyright to fit the needs imposed by academia. Second, they do so by creating another framework that sits somewhere between “ethics” and “law.” This becomes clear when one examines the licenses discussed for “Open Data.” As mentioned before, data as such is not protected by law; nevertheless, the proponents of such licenses try to create an enforceable framework for their needs.

Just as many authors never read their publishing agreements, few people actually spend the time reading the “legal code” of a Creative Commons license. Fewer still know that beyond the “Unported” version, there are indeed “ported” versions adapted to the needs of a specific jurisdiction (e.g., there is a “CC3.0-BY-DE” containing a version of the Creative Commons license adapted to the specific needs of Germany). These go beyond mere translation but try to adapt the rules to the requirements of the respective copyright laws—thus leading not to a common German Language CC-Version but requiring German, Austrian, and Swiss versions of each license. These “ported” versions take into account the specific requirements of each domestic law (e.g., regarding limitations of liability, protection of “non-creative” databases under the European Union Database Directive, and so on).

Many authors still believe that publishing their work on the Internet (whether as part of a blog, a “working paper,” or whatever) without any li-
license is sufficient. Others refer to one of the Creative Commons licenses available. Still, even with a system as simplified as Creative Commons, misinterpretations abound—especially regarding the “NonCommercial”-variants.

Indeed, publication practice seems to be governed mainly by outside influences (such as journals requiring new articles not published before) and unwritten ethical rules (mainly referring to attribution). While Creative Commons Licenses give their users peace of mind and generally perform their task extremely well, they suffer from the same issues as copyright law itself: Because it has to cover all kinds of works, it goes beyond what is needed, while leaving questions specific to academia unanswered.

The Privacy Debate

The “right to be let alone” (Warren and Brandeis, 1890), or at least ensure that personal data is not made available to everyone else in the world, seems quite intuitive at first glance. However, issues arise when one tries to apply that rule to real-life situations (see, e.g., Bruckman, 2002, who rejects analogies to practices in traditional media). This specifically relates to research on social networks, such as Twitter or Facebook. This is also where intuition begins to fail—because users make their information freely available to just about anyone, a right to privacy seems illusionary. Similarly, ethics provide little certainty or even guidance (Burk, 2008).

No modern scholar would believe that any data made available on the Internet (and specifically on social networks) are public and thus not subject to legal protection. Nevertheless, the Dos and Don’ts regarding the use of such data are not only subject to laws but also to an increasing ethical debate. The relevant European Union Directive 95/46/EC refers to anonymization of data in its recital 26:

> whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable; whereas codes of conduct within the meaning of Article 27 may be a useful instrument for providing guidance as to the ways in which data may be rendered anonymous and retained in a form in which identification of the data subject is no longer possible.

While this sounds fine in theory, we know that “deanonymization” is possible based on a very small amount of separate identifiers (cf. Ohm, 2010, also covering the relevant exceptions in the U.S.). The law remains ignorant of this possibility, which may in turn make any publication of data gathered on social media (even if much attention is paid to anonymization) possibly illegal. For the future, Ohm suggested several possible paths—strictly punish-
ing those who cause the actual harm, waiting for some (unlikely) technology to allow for perfect automatic anonymization or banning any re-identification technologies (much like technologies used to circumvent copy protection measures). The privately held social network providers have so far been unable to perform their role as gatekeeper and clearinghouse for academic use (Livingston, 2011). Still, it would be neither ethical to keep all research data secret nor would a potential breach of confidentiality regarding personal data be accepted under an ethical framework. Far more practical, yet also largely based on assumptions is the analysis by Yakowitz (2011). She correctly emphasized the value of research data and the danger of shifting to an “opt-in” model in research. Nevertheless, automatic anonymization may work in some contexts, but it is certainly impossible, e.g., in Twitter research. Thus, any authority or court eventually reviewing data publication would be required to apply a balancing test—with uncertain results.

Trying to resolve the issue in a 12-page paper is certainly impossible. Once again, a clear line rule is needed. Balancing tests are impractical in everyday social science research. Such rules should be developed and proposed by actual scientists—and not imposed upon them by a regulatory authority. There is a general danger of generalizing the issue to cover related interest groups, such as journalists (Scassa, 2010). However, much like the current rule, which is based on pre-existing academic practices, a new rule should be based on the actual needs of scientists and not exclusively on external policies, as observed in copyright law.

Summary and Outlook

Law and ethics are largely independent regimes. Rules of legal interpretation fail in determining “ethical” conduct, whereas legal rules will never be able to decide fully whether a practice is acceptable in academia. Conflicts between legal and ethical regimes are difficult to resolve. The issue is further emphasized when the legal framework lacks specificity and relevant practical experiences in its implementation (cf. Kaplow, 1992).

Intuition often provides the only, yet uncertain, guideline for determining proper scientific conduct. The situation not only causes significant risk for researchers and teachers alike, but it also leads to wasteful and often unnecessary precautions. A clear and definite framework, which can be interpreted by established methods, seems highly desirable. Developing such frameworks suffers from collective action problems, as well as divergent legal frameworks in different countries and diverging ethical standards in different disciplines. Nevertheless, such a task is not only in the interest of individual
scholars but also in the best interest of legislators and universities. Imposing outside rules on the academic process is unlikely to have a positive impact on actual practice.

Acknowledgements

This research was supported by the Düsseldorf University Strategic Research Fund.

References


Attribution 3.0 Germany (CC BY 3.0). Retrieved September 29, 2012 from http://creativecommons.org/licenses/by/3.0/de/deed.en


