Originally published in: Hawkins, Shane (Ed.): Access and Control in Digital Humanities. – Abingdon/New York: Routledge, 2021. Pp. 233-248. (Digital Research in the Arts and Humanities)
DOI: *https://doi.org/10.4324/9780429259616*

POSTPRINT

Erik Ketzan and Paweł Kamocki

Digital humanities research under United States and European copyright laws

Evolving frameworks

Copyright issues affect numerous avenues of digital humanities (DH) research, including the acquisition of copyright-protected text, audiovisual material and other types of data as corpora, the processing of that data in the course of research, and the sharing of research results (both for the peer-review process and publication of findings, including under open-access conditions, and archiving). DH researchers have relatively few resources to turn to when faced with copyright questions. Lawyers may be hired, but many of the questions DH researchers face can fall into legal gray areas, so that even when projects budget for specialist legal advice, they may often obtain uncertain guidance. Dealing with legal issues in DH research is often a question of risk minimization: adherence to best practices and acceptance of some degree of legal uncertainty. In practice, it is an open secret that many researchers simply copy texts, online material, and other data without permission, given the reality that publishers have very rarely sued researchers in court, but DH researchers should be aware of copyright laws, follow them to the best of their ability, and engage in advocacy for law reform, especially regarding the circumvention of technological protection measures (TPM).

This chapter summarizes the current state of copyright laws in the United States and European Union that most affect DH research, namely the fair use doctrine in the United States and research exceptions in Europe, including the Directive on Copyright in the Digital Single Market, which was finally adopted in 2019. While focusing on DH research, we do not generally address the related yet separate legal issues that affect teaching, libraries and archives. Naturally, any summary must omit some detail, and although we discuss legal issues in a format hopefully appropriate for the expected academic audience of this volume, we note that many of the laws and cases discussed have enough caveats, commentary, and case history to fill many further articles or even volumes. This summary begins with a description of recent copyright advances most relevant to DH research, and finishes with an analysis of a significant

remaining legal hurdle which DH researchers face: how do fair use and research exceptions deal with the critical issue of circumventing technological protection measures (TPM, a.k.a. Digital Rights Management, DRM)? Put simply, much of the data that researchers may wish to analyse via text and data mining (TDM) is protected by TPM, including films, e-books, online journal articles, and other types of material. Obtaining permission for these can be a lengthy, expensive, and inconclusive process for DH researchers, depending on their country of residence, but we hope that our discussion of the lawful means of obtaining TPM-protected material may contribute to both current DH research and planning decisions and inform future stakeholders and lawmakers of the urgent need to allow TPM circumvention for academic research. In particular, we highlight the recent practice of the German National Library in providing TPM-free copyrighted material to researchers for DH research, and propose that, especially in Europe, this may serve as a model for other countries. As for the United States, we highlight the helpful step of the Librarian of Congress and US Copyright Office in allowing researchers to circumvent the protection measures of DVDs, Blu-Rays, and digitally transmitted videos in certain cases. These demonstrate that pragmatic exceptions to TPM circumvention can be made, and hopefully broadened in the future.

Evolving copyright frameworks: United States fair use and changing European laws

Copyright law is especially relevant to DH, as the subject matter of copyright—text, images, audiovisual recordings—typically comprises the corpora of DH projects. Obtaining this data when it is protected by copyright is a key issue in DH that this chapter will address.

United States: fair use and "transformative" use

The fair use doctrine in US law is a limitation on copyright which allows certain uses of copyrighted material without the permission of the rightholder, for purposes such as *inter alia* criticism, commentary, news reporting, scholarship, or research.[1] For DH researchers using copyrighted material, determining whether such use qualifies as fair use is guided by existing statutory law and evolving case law. To determine whether a use is fair, courts look to the so-called four factor test, which examines: "(1) the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes; (2) the nature of the copyrighted work; (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole; and (4) the effect of the use upon the potential market for or value of the copyrighted work".[2] Following the landmark 1994 Supreme Court case *Campbell v. Acuff-Rose Music, Inc.*, this framework has incorporated an analysis of how "transformative" a use is. In essence, the more transformative the use, the more likely it is a fair use.[3] In *Campbell*, the Supreme Court defined transformativeness as "altering the original [work] with new expression, meaning, or message". In the ensuing decades, courts have interpreted transformativeness to include transformative *purpose*; as

the appeals court held in 2015 in the Google Books case, "Google's making of a digital copy of [. . .] books for the purpose of enabling a search for identification of books containing a term of interest to the searcher involves a highly transformative purpose", and that Google's "Snippet view [. . .] adds importantly to the highly transformative purpose of identifying books of interest to the searcher".[4] The adoption of the transformative purpose doctrine is vitally important for DH, as much DH research involves activities analogous to those in Google Books, such as the scanning of text and creation of corpora, as well as tools which provide snippets of this data to researchers or the public. While a full discussion of fair use and transformativeness is naturally more complicated than this one-paragraph summary, at first glance much DH research emerges favourably from this framework. DH research is typically noncommercial, educational, and much of it involves the arguably "transformative" TDM of works, corpora creation, and the publication of experimental results, and arguably has little effect on the market for, for example, books, DVDs, and the Internet. But a more detailed discussion of transformativeness and its role in recent cases is necessary.

Courts have increasingly placed ever greater emphasis on transformativeness in the determination of fair use. In a 2019 survey, Jiarui Liu states that transformativeness "has been gradually approaching total dominance in fair use jurisprudence, involved in 90% of all fair use decisions in recent years. More importantly, of all the dispositive decisions that upheld transformative use, 94% eventually led to a finding of fair use" (2019, p. 163). Liu (ibid.) discusses how transformativeness is changing the traditional four factor analysis:

> A finding of transformative use overrides findings of commercial purpose and bad faith under factor one, renders irrelevant the issue of whether the original work is unpublished or creative under factor two, stretches the extent of copying permitted under factor three towards 100% verbatim reproduction, and precludes the evidence on damage to the primary or derivative market under factor four even though there exists a well-functioning market for the use.

In an extensive statistical analysis of how often courts discuss transformative use in fair use cases (and which of the four factors transformative use is discussed amongst), Clark D. Asay et al. (2020, p. 967) write that "transformative use is eating the fair use world. First, modern courts take transformative use into account more than ever before. And second, this study's data show that the vast majority of modern courts use the transformative use concept throughout the fair use inquiry as the dominant means of resolving various fair use questions".

One notable example of a broad interpretation of transformativeness in fair use, as well as the increasing dominance of transformativeness in fair use analysis, is *Cariou v. Prince* (2013), in which the court held that a number of artworks by "appropriation artist" Richard Prince, which copied large sections of art photographs by Patrick Cariou and altered them to greater or lesser degrees, could

constitute fair use, and that a number of his works were transformative fair uses.[5] The photographer Cariou claimed that Prince's works both negatively impacted the market for his photographs and deeply offended him; his appellant brief states that "Prince desecrates Cariou's reverential portraits by defacing them, cutting them up, and splicing them together with erotic nudes, electric guitars and other detritus of our tawdry pop culture".[6] Yet the Second Circuit held that 25 out of 30 artworks by Prince which incorporate Cariou's photographs were fair use, and were "transformative" insofar as Prince's works "manifest an entirely different aesthetic from Cariou's photographs", and that "Prince's composition, presentation, scale, colour palette, and media are fundamentally different and new compared to the photographs, as is the expressive nature of Prince's work". This decision has generated controversy not only because of the extent of the photography copied by Prince but also through the court's emphasis on transformativeness far beyond the traditional four factors of fair use. Jane Ginsburg (2014) writes that "the Cariou court emphasized the 'transformative' nature of [Prince's] use at each stage, effectively collapsing the four-factor test into an extended discussion of transformative use, while decreasing the significance of the other factors". Kim Landsman (2015, p. 324) also discusses how the Prince case and earlier "appropriation art cases" have contributed to how "transformativeness has become virtually a one-factor test by over-powering factors that would otherwise weigh against fair use". In the case of Prince, the court focused on the transformativeness of Prince's work, while giving little weight to other fair use factors, including the fact that some of Prince's artworks sold for 2 million dollars or more and may have affected the market for Cariou's art.

Moving more specifically to fair use and scholarship, this can be summarized by some recurring themes: scholarship tends to fare well in courts, but the case-by-case nature of fair use creates uncertainty that chills scholarly activities. Writing in 2010, Randall P. Bezanson and Joseph M. Miller provide a thorough history of scholarship in copyright law: in 36 fair use cases since 1976 involving scholarship or claimed scholarship, fair use was found in 27 of these, while the "remaining nine cases involved wholesale copying, bad faith or disproportionate use in light of the scholarly purpose" (2010, p. 446). Bezanson and Miller state that courts "easily manage the borderline between scholarship and other genres" and "almost always deem scholarly use to be fair use", and that "the dominant theme in the true scholarship cases is that scholarship is transformative and thus, as a matter of presumption or even of law, courts should not deem scholarly use to infringe the copyright owner's economic interests, or even any copyright interest" (pp. 449–450). Writing in 2019, Matthew D. Bunker looked more broadly at "quasi-scholarly appropriation, such as borrowings for popular biographies, biographical films, and reference works that relate to popular culture but that may lack significant academic gravitas" (Bunker 2019, p. 8), and states that the "rise of the transformative use doctrine over the last several decades [. . .] seems a positive development for scholarly borrowers of other's expression" (p. 20). Despite this generally positive situation for scholars, Bezanson and Miller

emphasize how the case-by-case nature of fair use creates legal uncertainty for scholars, a situation we see widespread in the DH community.

While fair use is often readily acknowledged by US courts when it comes to scholarly activities, finding fair use for much DH work could be more problematic, due to the large volume of copying which DH sometimes performs, and the fact that whole works (not only sections) are often copied for TDM. In these regards, DH research in the United States was greatly aided by related cases involving HathiTrust, the global partnership of research institutions and libraries which archives media online, and Google Books. The Authors Guild, along with a number of authors and non-U.S. writers' unions and organizations, sued Hathitrust and several of its university partners for copyright infringement in a case first decided in 2012, by which point HathiTrust's Digital Library contained digitized copies of almost 10 million books, approximately 73% of them copyrighted.[7] At the same time, the Authors Guild and a number of authors separately sued Google over Google Books, which at that time had scanned 20 million books, delivered digital copies to participating libraries (who then "contributed" them to HathiTrust), and provided online text search of many titles in the form of snippets or limited amounts of pages.[8] Hathitrust provided similar, but somewhat different services than Google Books, including access for people with certified print disabilities and more limited search of copyrighted works (Google Books displayed snippets in response to queries, but HathiTrust limited these to page numbers and how many times the queried term appeared on the page). In both of these cases, the Authors Guild and its fellow plaintiffs asserted that this unauthorized reproduction and distribution of their copyrighted texts constituted copyright infringement. In separate decisions, a federal district court applied the four factors of fair use, including the transformative nature of archiving and allowing full-text search, and held that HathiTrust and Google Books' activities were both fair use. Both cases were appealed to federal circuit courts, who again handed overall victories to HathiTrust, the libraries, and Google. The US Court of Appeals for the Second Circuit held in 2014 that fair use allows the libraries to digitize copyrighted works for the purpose of permitting full-text searches, stating that "the creation of a full-text searchable database is a quintessentially transformative use",[9] and held in 2015 that "Google's unauthorized digitizing of copyright-protected works, creation of a search functionality, and display of snippets from those works are non-infringing fair uses. The purpose of the copying is highly transformative, the public display of text is limited, and the revelations do not provide a significant market substitute for the protected aspects of the originals. Google's commercial nature and profit motivation do not justify denial of fair use", and that "Google's provision of digitized copies to the libraries that supplied the books, on the understanding that the libraries will use the copies in a manner consistent with the copyright law, also does not constitute infringement".[10] The Authors Guild appealed the Google decision to the Supreme Court, who declined to review the case, effectively establishing these decisions as forceful precedent in United States courts.

The importance of the HathiTrust and Google Books decisions for DH research in the United States (and to an extent, globally, insofar as HathiTrust and Google Books' global operations establish a de facto norm) is difficult to overstate. These are two of the largest (if not *the* largest) DH projects in history, scanning entire books, by the millions, and providing some level of public access via query, and were held to be fair use by two federal district court and two federal appellate court decisions. The courts reiterated that common DH activities, such as the creation of searchable databases, are transformative uses, which, to remind the reader, is tending to overshadow the rest of the historical four factor fair use test. Samuelson notes that in the HathiTrust decision, "the court viewed the full-text search use as transformative because it considered the HathiTrust database itself as a new work that had a different purpose than the books in it" and that "the scale and scope of copying [. . .] was beyond what had been deemed fair in other cases" (2015, p. 836). Compared to this scale, scope, and level of transformativeness, many DH projects would be viewed favourably following these fair use analyses.

It should also be noted that in both of these cases, DH scholars and librarians engaged in successful public advocacy. In court cases, concerned outside parties may file amicus briefs with the court (advocacy memoranda that the court may, but are not obligated to, consider). Amicus briefs filed by a number of digital humanities professors, law professors, and associations, as well as separate briefs filed by library associations, argued the benefits of scanning, TDM, and archiving.[11] These efforts were successful insofar as the district court decisions in both the Google and HathiTrust cases cited these amicus briefs favourably, and should serve as a reminder for the DH field to remain engaged in legal advocacy which benefits the field.

DH in the United States has thus made important legal gains in recent years, thanks especially to the HathiTrust and Google Books cases, in terms of increasing importance of transformativeness in the fair use analysis, and the DH community's willingness to engage in legal advocacy. None of these, however, account for a remaining unsolved problem: how DH scholars may make use of works protected by TPM, which will be discussed below.

Fair use, meanwhile, has been growing outside of the United States. The Philippines, Israel, and South Korea have adopted US-style fair use systems, and there have been debates for proposed changes in Canada, Japan, and Australia (Decherney 2014). While various commentators over the years have advocated for the adoption of fair use principles into EU copyright law (e.g. Hargreaves 2011; Hugenholtz & Senftleben 2011), recent decisions by the Court of Justice of the European Union (CJEU) suggest that fair use will not enter EU jurisprudence through judicial opinions, but only through possible future legislation, which is currently not on the horizon.[12]

Fair use versus TPM in the United States

Much of the text and data that DH researchers may wish to use is protected by TPM: e-books, DVDs, and Blu-Rays contain software which prevents the

copying of their content, media streaming websites encrypt stream content to disable downloading, and even password protection on a website could be considered a TPM. While fair use has expanded in the United States in ways that benefit DH, and new laws across the EU are introducing more progressive exceptions for research, none of the legal frameworks fully offer researchers a clear and simple path for obtaining TPM-protected data. On the contrary, both European Union and US frameworks provide difficult or vague legal rules. Laws against the circumvention of TPM derive from two related international treaties designed to combat digital piracy and protect copyright-protected content in the age of the Internet: the WIPO Copyright Treaty and the WIPO Performances and Phonograms Treaty (both 1996), which required Contracting Party nations to "provide adequate legal protection and effective legal remedies against the circumvention of effective technological measures" that are used by authors, performers and producers of phonograms.[13]

The United States implemented these treaties via the Digital Millennium Copyright Act (DMCA) (1998), which states in §1201 that "No person shall circumvent a technological measure that effectively controls access to a work protected under this title" and that to "'circumvent a technological measure' means to descramble a scrambled work, to decrypt an encrypted work, or otherwise to avoid, bypass, remove, deactivate, or impair a technological measure, without the authority of the copyright owner", and provides substantial civil remedies and criminal penalties.[14] Bizarrely, even over 20 years after the passage of the DMCA, the basic question of whether fair use can apply to TPM circumvention remains unclear, a situation complicated by conflicting case law, voluminous scholarly commentary, and unsuccessful Congressional bills which sought to clarify the matter. As summarized by Samuelson (2015, p. 859), "The law review literature about fair use and [. . .] technical restrictions is ample, and most commentators favor fair use as a policy that should trump [. . .] technical restrictions under some conditions. However, the case law on these issues is at best mixed". Some courts and commentators have stated that fair use cannot apply to the circumvention (and trafficking in circumvention software) of TPMs; in *Universal City Studios Inc. v. Corley*, the Second Circuit Court of Appeals stated that fair use does not apply to acts of circumvention, and that Congress did not intend for it to do so.[15] Legal scholar David Nimmer (2000) argued the same conclusions. The Ninth Circuit held, as in *Corley*, that circumvention does not even require an act of copyright infringement—circumvention is enough to violate the statute—although the Ninth Circuit left "whether fair use might serve as an affirmative defense to a prima facie violation of §1201" as an open question.[16] Conversely, the Court of Appeals for the Federal Circuit concluded that fair use *can* apply to acts of circumvention, and additionally, DMCA prohibition against circumvention of TPMs only applies in cases of copyright infringement.[17] To illustrate this legal confusion with a hypothetical scenario: a US citizen, Anna Reader, purchases an ebook of a public domain novel, for example, *Moby Dick*, and uses a widely available program to strip the file of its TPM so that she may read the novel on various electronic devices and share the text of *Moby Dick* on

the Internet. Under current case law, some courts would find the circumvention (the removal of software which prevents copying) illegal under the DMCA, while at least one other might not.

DH scholars dealing with textual interpretation may be interested to know that one of the main problems in deciding whether fair use can apply to TPM circumvention rests on how one reads a single sentence in the text of the DMCA: "Nothing in this section shall affect rights, remedies, limitations, or defenses to copyright infringement, including fair use, under this title".[18] One can read this as a savings clause (Liu 2019) for fair use under the DMCA, and indeed the Court of Appeals for the Federal Circuit in *Chamberlain* notes the possibility to read it this way, as have scholars like Samuelson (2015, pp. 860–861). However, scholars such as Nimmer and certain courts have interpreted this differently, in part by turning to the legislative history of the DMCA, which saw competing bills in the House and Senate and eventually a joint legislative compromise; under this reading, as articulated by the Second Circuit, this sentence "clearly and simply clarifies that the DMCA targets the circumvention of digital walls guarding copyrighted material (and trafficking in circumvention tools), but does not concern itself with the use of those materials after circumvention has occurred. Subsection 1201(c)(1) ensures that the DMCA is not read to prohibit the 'fair use' of information just because that information was obtained in a manner made illegal by the DMCA".[19] Applying this reading to our hypothetical, Anna Reader would be liable for circumventing the TPM on her ebook of *Moby Dick*, but not liable for subsequently sharing the text of *Moby Dick* online.

Congressional attempts to comprehensively reform the DMCA, including the thorny issue of fair use versus TPM circumvention, have been unsuccessful.[20] DMCA reform has, however, taken small steps. Congress passed a law in 2014 which allows consumers to legally unlock their phones (a circumvention of TPM),[21] while additional exceptions promulgated by the US Copyright Office every three years have come into effect, allowing, for instance, the circumvention of TPM on e-books for use with assistive technologies for the blind and visually impaired.[22] These changes to the DMCA have had little relation to the majority of DH research questions, but they at least demonstrate that a research exception or even fair use exception might not be inconsistent with WIPO treaty obligations. The Electronic Frontier Foundation (EFF) has been perhaps the most vocal and active critic of the DMCA, arguing that the DMCA chills free expression, scientific research, and fair use, and that "By banning all acts of circumvention, and all technologies and tools that can be used for circumvention, section 1201 grants to copyright owners the power to unilaterally eliminate the public's fair use rights" (Electronic Frontier Foundation 2003). In 2016, the EFF filed a federal lawsuit challenging the constitutionality of the DMCA, representing three plaintiffs, including an academic researcher (Assistant Professor Matthew Green of the Johns Hopkins Information Security Institute), but as of June 2019, this lawsuit is still ongoing (Electronic Frontier Foundation 2016).

Every three years, the Librarian of Congress and US Copyright Office issue temporary exemptions to DMCA's broad sweep. As of late 2020, for instance, researchers are allowed to circumvent the protection measures of DVDs, Blu-Rays, and digitally transmitted videos to make fair uses in certain educational uses, for use in non-commercial videos and nonfiction multimedia e-books.[23] While this rule is a positive step, it would ideally be permanently adopted and expanded to clarify TDM as fair uses. In 2020, as part of the eight triennial rulemaking process, a number of organisations—Authors Alliance, the Library Copyright Alliance, and the American Association of University Professors—are petitioning the US Copyright Office to grant a new 3-year rule allowing researchers to circumvent TPM for TDM of literary works published electronically, as well as for motion pictures. We encourage readers to follow this news, as such a rule, if approved, would be a very important step for DH research in the United States.

European Union: research exceptions and technical protection measures

Research in the European Union is governed by copyright limitations and exceptions in EU and national laws, which allow researchers to lawfully perform certain unauthorised uses (exceptions) or limit the scope of the rightholders' monopoly (limitations). The 2001 Directive on the harmonization of certain aspects of copyright and related rights in the information society ("InfoSoc Directive") contains a fairly limited list of exceptions and limitations, including quotations for criticism or review, non-commercial educational or scientific research, and private copies.[24] This list was, from a DH researcher's perspective, very narrow. The limitation to only non-commercial research restricts certain DH projects, especially as there is widespread confusion over what "non-commercial" means; for instance, public–private partnerships in DH research might remove such research from the scope of protection.

European Union law was made compatible with the WIPO treaties by the 2001 InfoSoc Directive, in which the issue of the research exception vs. TPM is addressed in Article 6(4), which states, "Member States shall take appropriate measures to ensure that rightholders make available to the beneficiary of an exception or limitation provided for in national law [. . .] the means of benefiting from that exception or limitation". For example, if a researcher wished to obtain TPM-protected material for non-commercial scientific research, they should contact the rightholder and simply ask for it. In the event that these "voluntary measures" fail, various Member States have, as Gwen Hinze summarizes (2016, p. 810), "followed four main approaches: some countries have created a right of self-help that entitles exception beneficiaries to circumvent; a second set of countries have created an arbitration proceeding; a third set of countries have created a direct right of action to enforce privileged exceptions in a court of law; and a fourth set of countries have created an administrative or executive authority to

regulate the use of TPMs". In the United Kingdom, for instance, a researcher is obliged to file a notice of complaint with the UK Secretary of State, who then considers the legality of the request, and then either turns down the applicant or approaches the owner of the copyrighted work to facilitate a data exchange.[25]

So far, the method for obtaining TPM-protected material under InfoSoc sounds reasonable: if a researcher wished to obtain TPM-protected material for research, they have such a right and should simply ask, and if the rightholder is not forthcoming, go through an administrative process. There are two fundamental problems with this, however. First, the right to obtain TPM-protected material for research does not apply to content made available on the Internet under contract or license agreement. This is accomplished in the Directive via the phrase "shall not apply to works or other subject-matter made available to the public on agreed contractual terms in such a way that members of the public may access them from a place and at a time individually chosen by them".[26] As Urs Gasser & Michael Girsberger (2004, p. 17) state, "It remains unclear what such interactive on-demand services include". As Hinze (2016, p. 807) states, the exclusion of vast swaths of Internet material "arguably makes the whole countermeasure framework of little practical relevance to the digital environment". This might strike the reader in 2021 as exceedingly odd, as readers today perhaps mostly associate TPM with online material.

In addition to the scope of the right to TPM-protected material for research, as Lucie Guibault et al. (2012, p. 110) stated, "the exercise of this facility is not always easy in practice". The administrative process for obtaining TPM-protected material can be a lengthy and complicated one, as one UK researcher discovered. In 2015, the Libraries and Archives Copyright Alliance (LACA) instituted a test case on behalf of a UK university researcher by filing a complaint with the Secretary of State to obtain access to data that was protected from downloading by a CAPTCHA TPM on a website (Chartered Institute of Library and Information Professionals 2016). Over two months after filing the complaint, LACA received the decision of the Secretary of State's office, which denied the request, reportedly stating that "the complaint falls outside of the scope [as the] section does not apply to copyright works made available to the public on agreed contractual terms in such a way that members of the public may access them from a place and at a time individually chosen by them. As users of the website are able to download the databases on-line, then the terms and conditions governing access to those works will prevail, and it will be necessary to approach the owner of the website to request permission to copy/extract the data" (ibid.). Consistent with the InfoSoc Directive, TPM-protected online content falls outside the scope, which is frustrating for DH, as much of the TPM-protected content that DH researchers wish to analyse is precisely such online material.[27]

Again, however, the process for obtaining TPM-protected material for research varies by Member State, and the ability for researchers in Germany to obtain TPM-protected material was greatly aided by a 2005 agreement between the German National Library and the German Federation of the Phonographic Industry and

the German Booksellers and Publishers Association, whereby publishers deposit TPM-free copies of their CDs, CD-ROMs, and e-books with the library, who then provide these to legitimate beneficiaries such as researchers (Gasser 2005). In the realm of DH, in 2019, the German National Library provided thousands of TPM-free ebooks, many of them protected by copyright, to DH researchers in Germany for a study of thematic complexity in literature (Jannidis et al. 2019). This was the first major instance of the German National Library aiding DH researchers in obtaining TPM-free ebooks in such large quantities, and in principle would apply to the Library's entire collection.[28] This is a very positive development for DH researchers in Germany and could serve as a model for DH researchers to legally obtain TPM-protected material in all countries, especially as obtaining material from the Library seems far more straightforward than, for instance, petitioning the Secretary of State in the United Kingdom. The German model does not, however, address TPM-protected web content.

Apart from these cases, however, we are aware of few attempts by DH researchers to make use of their right to TPM-protected material under national implementations of Article 6(4), continuing the state of uncertainty (and which would greatly benefit from its own dedicated study or official report). More broadly, after the InfoSoc Directive there was great uncertainty as to whether researchers could lawfully perform TDM (the core activity of much DH research) on content such as the open Internet, as well as content protected by TPM, with some voices advocating "the right to read is the right to mine" (Mounce 2014), and some publishers and rightholder associations pushing back against a general right to TDM, proposing various licenses and platforms for supporting research TDM.[29] This confusion regarding TDM was addressed first by a number of Member State laws, and then the long-awaited Directive on Copyright in the Digital Single Market ("2019 Copyright Directive").

In 2014, the United Kingdom was the first Member State to pass a major TDM law, which allows the making of a copy for TDM under numerous provisions derived from the InfoSoc Directive: non-commercial research only, contracts may not restrict the copying, and the requirement of lawful access (called "legal access" in the English version of InfoSoc). This law provided *some* certainty for DH research in the United Kingdom regarding TDM, but also much (continued) uncertainty. For instance, the law prevents the sharing of data with another person, for example a research team member or peer reviewers to verify results. In 2016, France introduced a TDM exception (Art. L. 122–5, 10° of the French Intellectual Property Code), but its implementing decree was not passed and remains effectively dead letter (see Kamocki et al. 2018). In 2017, Germany passed a TDM law which improved upon the UK attempt, but remains constricted by the InfoSoc Directive, dubbed the "copyright in the knowledge economy" reforms, which include research exceptions for scientific research and TDM.[30] Under these new German laws, which entered into force in March 2018, for the purposes of non-commercial scientific research, up to 15% of a work may be reproduced, disseminated and made publicly accessible for a defined circle of

persons for their own scientific research and for individual third parties, to the extent that this serves to verify the quality of scientific research; for personal scientific research (presumably, research by a single researcher), up to 75% of a work may be reproduced.[31] For TDM, the new German laws are broader than in the United Kingdom: for non-commercial scientific research, it is permitted to text and data mine to create a corpus (a word the statute employs), to share the corpus with a defined circle of persons for joint scientific research as well as to individual third parties to check the quality of the scientific research.[32] The laws require that the corpus shall be deleted after the completion of research and access to the public terminated, but permits the transmission and reproduction of the material at libraries and archives for permanent storage.[33] Following the InfoSoc Directive, the German exception is expressly non-overridable by contractual clauses, but does also require flat-rate equitable remuneration be paid to a copyright collecting society.[34]

In 2019, the long-awaited Directive on Copyright in the Digital Single Market ("2019 Copyright Directive") was finally passed, providing a common framework for TDM in Europe, and which Member-States must implement within two years. The 2019 Copyright Directive creates an exception for scientific research by research organizations and cultural heritage institutions, who may perform DRM on protected works "to which they have lawful access" (Art. 3(1)). The Directive provides a partial definition of "lawful access": access based on an open-access policy, contractual means, subscriptions, and freely available online content (Recital 14). Copies of works "shall be stored with an appropriate level of security and may be retained for the purposes of scientific research, including for the verification of research results" (Art. 3(2)). Rightholders are allowed to "apply measures to ensure the security and integrity of the networks and databases where the works or other subject matter are hosted" (Art. 3(3)); presumably, rightholders might gain some kind of right to audit the security and stability of DH researcher's servers, or otherwise request certification or accreditation that computer systems are secure. As in the UK and German exceptions, contractual provisions may not override researchers' right to text and data mine (Art. 7). For much DH research in the EU, much of the 2019 Copyright Directive will be helpful: clear(er) guidelines, confirmation that the open Internet may be freely mined, and the expansion of the contractual override to all of the EU. But the Directive is one more missed opportunity to clarify whether researchers may circumvent TPM to perform TDM or provide stronger solutions to obtaining TPM-protected material than the InfoSoc Directive provides, as discussed below. It also remains to be seen to what extent the United Kingdom will implement this Directive post-Brexit; if the United Kingdom TDM law remains unrevised, it will soon become the least comprehensive TDM law in Europe (although it is also possible that once separated from the EU, the UK might adopt laws more in line with their historical legal doctrine of fair dealing). It also remains to be seen whether and how Germany will alter its TDM exception to harmonize with the Directive.

In summary, DH researchers in the United States and the European Union face legal uncertainty and significant hurdles to overcome in obtaining TPM-protected material. The German model, under which researchers obtain TPM-protected material such as DVDs and e-books through the national library, is a positive if incomplete solution that should be emulated (and expanded, if possible, to TPM-protected Web content). While the InfoSoc Directive contained language directing Member States to solve the problem of TPM circumvention for research, the national solutions have proven (at least sometimes) unwieldy, and require researchers to demand access from rightholders, who may be difficult to contact or unwilling to comply, potentially necessitate the hiring of lawyers, filing complaints with authorities, and waiting for months at a minimum. In the United States, the rulemaking of the US Copyright Office may be able to provide the best solution for the thorny question of TPM circumvention, but this process remains incomplete and is limited by the need to renew and review rules every three years. Ideally, the preferred solution in all jurisdictions would be statutory authorization for legitimate researchers to circumvent TPM on common media: e-books, DVDs/Blu-Rays, and websites in certain circumstances. Not only would this effect the goals of most copyright laws to promote science and the arts, but acknowledge the simple reality that most TPM is easily circumventable by tools found readily online; a bright teenager could remove the TPM from most of these within minutes. Naturally, many researchers simply do this already, either in ignorance of the law or trusting that researchers are rarely sued, but as DH projects grow larger, involving ever-increasing budgets, government funding, requisite audits, and overall increased scrutiny, minimizing legal risk in DH research is imperative, and will be greatly benefited by a proper legal solution in all jurisdictions.

Notes

1 17 U.S.C. §107.
2 Ibid.
3 510 U.S. 569 (1994).
4 *Authors Guild v. Google, Inc.*, 804 F.3d 202 (2d Cir. 2015).
5 714 F. 3d 694 (2d Cir. 2013).
6 Joint Brief and Special Appendix for Defendants-Appellants at 5, *Cariou v. Prince*, 784 F. Supp. 2d 337 (S.D.N.Y. 2011) (No. 11–1197), 2011 WL 5325288.
7 *Authors Guild, Inc. v. Hathitrust*, 902 F. Supp. 2d 445 (S.D.N.Y. 2012).
8 *Authors Guild, Inc. v. Google Inc.*, 954 F. Supp. 2d 282 (S.D.N.Y. 2013).
9 *Authors Guild, Inc. v. HathiTrust*, 755 F.3d 87 (2d. Cir. 2014).
10 *Authors Guild v. Google Inc.*, 804 F.3d 202 (2d. Cir. 2015).
11 The appellate amicus brief in HathiTrust was signed by over 100 professors and scholars, as well as the Association for Computers and the Humanities and Canadian Society for Digital Humanities; see the Brief of Digital Humanities and Law Scholars as Amici Curiae in *Authors Guild v. HathiTrust*, 12–4547-CV (2d Cir. 2013). Library associations included the American Library Association, Association of College and Research Libraries, and Association of Research Libraries.
12 Cf. CJEU, three judgements passed on 29 July 2019 in cases C-469/17 (*Funke Medien*), C-476/17 (*Pelham*) and C-516/17 (*Spiegel online*).

13 WIPO Copyright Treaty (WCT), Art. 11. WIPO Performances and Phonograms Treaty (WPPT), Art. 18.
14 17 U.S.C. §1201—§1205.
15 *Universal City Studios, Inc. v. Corley*, 273 F.3d 429, 443–44 (2d Cir. 2001), affirming *Universal City Studios v. Reimerdes*, 111 F. Supp. 2d 294 (S.D.N.Y. 2000).
16 *MDY Indus., LLC v. Blizzard Entm't, Inc.*, 629 F.3d 928, 950 n.12 (9th Cir. 2010).
17 *Chamberlain Grp., Inc. v. Skylink Techs.*, Inc., 381 F.3d 1178, 1202–03 (Fed. Cir. 2004).
18 17 U.S.C. § 2011(c)(3).
19 *Universal City Studios, Inc. v. Corley*, 273 F.3d 429, 443–44 (2d Cir. 2001).
20 Unpassed Congressional DMCA reform bills include the Digital Media Consumers' Rights Act (2003) and "Freedom and Innovation Revitalizing United States Entrepreneurship Act of 2007" (FAIR USE Act, 2007).
21 The Unlocking Consumer Choice and Wireless Competition Act, amending §1201.
22 The most recent being United States Copyright Office, *Exemption to Prohibition on Circumvention of Copyright Protection Systems for Access Control Technologies* (2018).
23 www.copyright.gov/title37/201/37cfr201-40.html
24 Directive 2001/29/EC of 22 May 2001, Article 5(2)(b).
25 Copyright and Rights in Performances (Personal Copies for Private Use) Regulations 2014, 296ZEA.
26 Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonization of certain aspects of copyright and related rights in the information society, Article 6(4). See also Recital 53.
27 Anecdotally, we are also aware of another UK researcher who, attempting to exercise their right to TPM-protected material, directly approached a bestselling author's agent to attempt to obtain e-book texts for research. The author's agent was sympathetic, and supportive of the researcher's goals (computational literary analysis), but stated that the costs of providing the texts was simply too high: it would have to involve a contract, different novels had been published through different publishers, and lawyers would have to be paid to draft a complex document, which was simply not worth the author's time and money. This researcher has not yet attempted a complaint with the Secretary of State, but this anecdote illustrates the potentially high monetary cost for rightholders to supply TPM-free material to researchers, even when all parties are sympathetic.
28 Correspondence with Dr. Peter Leinen, German National Library, 23 July 2019.
29 E.g., "A statement of commitment by STM publishers to a roadmap to enable text and data mining (TDM) for non-commercial scientific research in the European Union" (International Association of Scientific, Technical and Medical Publishers 2013). See further the European Commission stakeholder dialogue (13 November 2013), "Licences for Europe", on Content in the Digital Single Market (https://ec.europa.eu/digital-single-market/en/news/licences-europe-stakeholder-dialogue).
30 Urheberrechts-Wissensgesellschafts-Gesetz (UrhWissG, 2017).
31 Urheberrechtsgesetz § 60c (revised).
32 Urheberrechtsgesetz § 60d (revised).
33 Urheberrechtsgesetz § 60d (revised).
34 Urheberrechtsgesetz § 60g, 60h (revised).

References

Case law

Authors Guild, Inc. v. Google Inc., 954 F. Supp. 2d 282 (S.D.N.Y. 2013).
Authors Guild, Inc. v. HathiTrust, 902 F. Supp. 2d 445 (S.D.N.Y. 2012).
Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2d Cir. 2014).
Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015).

Brief of Digital Humanities and Law Scholars as Amici Curiae in *Authors Guild v. HathiTrust*, 12–4547-CV (2d Cir. 2013).

Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994).

Cariou v. Prince, 714 F. 3d 694 (2d Cir. 2013).

Chamberlain Group, Inc. v. Skylink Techs., Inc., 381 F.3d 1178 (Fed. Cir. 2004).

CJEU judgements 29 July 2019: Cases C-469/17 (*Funke Medien*), C-476/17 (*Pelham*) and C-516/17 (*Spiegel online*).

Joint Brief and Special Appendix for Defendants-Appellants, *Cariou v. Prince*, 784 F. Supp. 2d 337 (S.D.N.Y. 2011) (No. 11–1197), 2011 WL 5325288.

MDY Indus., LLC v. Blizzard Entm't, Inc., 629 F.3d 928 (9th Cir. 2010).

Universal City Studios, Inc. v. Corley, 273 F.3d 429 (2d Cir. 2001), affirming *Universal City Studios v. Reimerdes*, 111 F. Supp. 2d 294 (S.D.N.Y. 2000).

Publications

Asay, C. D., Sloan, A., Sobczak, D., (2020). Is transformative use eating the world? *Boston College Law Review*. 61(3), 905–970. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3332682

Bezanson, R. P. and Miller, J. M., (2010). Scholarship and fair use. *Columbia Journal of Law and the Arts*. 33, 409–470. Available from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1561846

Bunker, M. D., (2019). Decoding academic fair use: Transformative use and the fair use doctrine in scholarship. *Journal of Copyright in Education and Librarianship*. 3(1), 1–24. https://doi.org/10.17161/jcel.v3i1.6481

Chartered Institute of Library and Information Professionals (CILIP)., (2016). *Libraries and archives copyright alliance test case*. London: CILIP. Available from: http://uklaca.org/wp-content/uploads/2019/01/notice_of_complaint_to_the_secretary_of_state_-_test_case_1_0.pdf

Decherney, P., (2014). Fair use goes global. *Critical Studies in Media Communication*. 31(2), 146–152.

Digital Millennium Copyright Act (DMCA)., (1998). 17 U.S.C. §§ 512, 1201–1205. Available from: https://www.govinfo.gov/content/pkg/PLAW-105publ304/pdf/PLAW-105publ304.pdf

Electronic Frontier Foundation., (2003). Unintended consequences: Five years under the DMC [online]. *Electronic Frontier Foundation*. Available from: www.eff.org/wp/unintended-consequences-five-years-under-the-dmca

Electronic Frontier Foundation., (2016). EFF lawsuit takes on DMCA section 1201: Research and technology restrictions violate the first amendment. *Electronic Frontier Foundation* [online]. Available from: www.eff.org/press/releases/eff-lawsuit-takes-dmca-section-1201-research-and-technology-restrictions-violate

Gasser, U., (2005). German National Library: License to circumvent DRM. *Law and Information Blog* [online]. 26 January. Available from: http://blogs.harvard.edu/ugasser/2005/01/26/german-national-library-license-to-circumvent-drm/

Gasser, U. and Girsberger, M., (2004). Transposing the Copyright Directive: Legal protection of technological measures in EU-Member States: A genie stuck in the bottle? *The Berkman Centre for Internet and Society at Harvard University, Berkman Publication Series No. 2004–10*. Available from: https://cyber.harvard.edu/media/files/eucd.pdf

Ginsburg, J., (2014). Letter from the US: Exclusive rights, exceptions, and uncertain compliance with international norms: Part II (fair use). *Revue Internationale du Droit*

d'Auteur (RIDA), 242, 175–279; *Columbia Law and Economics Working Paper No. 503*. Available from: https://scholarship.law.columbia.edu/faculty_scholarship/1897

Guibault, L., Westkamp, G. and Rieber-Mohn, T., (2012). Study on the implementation and effect in Member States' laws of Directive 2001/29/EC on the harmonisation of certain aspects of copyright and related rights in the information society. *Report to the European Commission, DG Internal Market, February 2007: Amsterdam Law School Legal Studies Research Paper No. 2012–28: Institute for Information Law Research Paper No. 2012–23*. Available from: https://ssrn.com/abstract=2006358

Hargreaves, I., (2011). Digital opportunity: A review of intellectual property and growth. *United Kingdom Intellectual Property Office, Independent Review* [online]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/32563/ipreview-finalreport.pdf

Hinze, G., (2016). Brave New World, ten years later: Reviewing the impact of policy choices in the implementation of the WIPO internet treaties' technological protection measure provisions. *Case Western Reserve Law Review*. 57(4), 779–821. Available from: https://scholarlycommons.law.case.edu/caselrev/vol57/iss4/7

Hugenholtz, P. B. and Senftleben, M., (2011). *Fair use in Europe: In search of flexibilities*. Institute for Information Law, Amsterdam: IViR & Vrije Universiteit Amsterdam. http://dx.doi.org/10.2139/ssrn.1959554

International Association of Scientific, Technical and Medical Publishers (STM)., (2013). *Text and data mining for non-commercial scientific research: A statement of commitment by STM publishers to a roadmap to enable text and data mining (TDM) for non-commercial scientific research in the European Union*. The International Association of Scientific, Technical and Medical Publishers (STM) [online]. Available from: www.stm-assoc.org/2017_05_10_Text_and_Data_Mining_Declaration.pdf

Jannidis, F., Konle, L. and Leinen, P., (2019). Thematic complexity. *DH2019 Conference Book of Abstracts, 9–12 July 2019, Utrecht, Netherlands*. Available from: https://dev.clariah.nl/files/dh2019/boa/0504.html

Kamocki, P., Ketzan, E., Wildgans, J. and Witt, A., (2018). New exceptions for text and data mining and their possible impact on the CLARIN infrastructure. In: I. Skadina and M. Eskevich, eds. *Selected papers from the CLARIN Annual Conference 2018, Pisa, 8–10 October 2018*. pp. 66–71. Available from: www.ep.liu.se/ecp/159/ecp18159.pdf

Landsman, K. J., (2015). Does Cariou v. Prince represent the apogee or burn-out of transformativeness in fair use jurisprudence? A plea for a neo-traditional approach. *Fordham Intellectual Property, Media and Entertainment Law Journal*. 24(2), 321–379. Available from: https://ir.lawnet.fordham.edu/iplj/vol24/iss2/1

Liu, J., (2019). An empirical study of transformative use in copyright law. *Stanford Technology Law Review*. 22, 163–241. Available from: https://law.stanford.edu/wp-content/uploads/2019/02/Liu_20190203.pdf

Mounce, R., (2014). The right to read is the right to mine: Text and data mining copyright exceptions introduced in the UK. *Impact of Social Sciences Blog* [online]. 4 June. Available from: https://blogs.lse.ac.uk/impactofsocialsciences/2014/06/04/the-right-to-read-is-the-right-to-mine-tdm/

Nimmer, D., (2000). A riff on fair use in the Digital Millennium Copyright Act. *University of Pennsylvania Law Review*. 148(3), 673–742. Available from: https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=3355&context=penn_law_review

Samuelson, P., (2015). Possible futures of fair use. *Washington Law Review*. 90(2), 815–868. Available from: https://digitalcommons.law.uw.edu/wlr/vol90/iss2/9