

# Copyright and Humanities Research: A Global Perspective

*a DH2019 pre-conference workshop organized by  
Vanessa Hanneschläger, Pawel Kamocki, Walter Scholger, Andreas Witt*

## Program

**9.00 Opening** [chairs: Vanessa Hanneschläger & Walter Scholger]

- Welcome by the organizers, introduction to scope of the workshop

**9:30 Presentation session I** [chair: Vanessa Hanneschläger]

- *Erik Ketzan*: Should DH Consult an Oracle for the “ Copyright Case of the Decade”?
- (*Pawel Kamocki*), *Andreas Witt*: EU Exceptions for Text and Data Mining for Research Purposes : from a Dry Spell to a Crisis of Abundance?

**11.00 Break**

**11.30 Presentation session II** [chair: Vanessa Hanneschläger]

- *Kim Nayyer*: Advancing digital scholarship activities with Canadian and global comparative copyright law
- *Simone Rebora*, *Moniek Kuijpers*, *Piroska Lendvai*: Privacy and Property Issues in the Computational Study of Social Reading

**13.00 Lunch break**

**14.00 Presentation session III** [chair: Walter Scholger]

- *Nirmala Menon*: Copyright and Impact Factors in the Humanities - Decolonizing Knowledge Systems and Assessments
- *Paul Millar*: Copyright and Post-disaster Archiving
- *Kiyonori Nagasaki*: Recent situation of copyright law and other licensing issues for DH in Japan

**15:30 Break**

**16:00 Panel/Plenary Discussion: “Digital Humanities: Research with digital resources in post-anarchical times”** [chair: Walter Scholger]

Panelists: Erik Ketzan, Kim Nayyer, Andreas Witt

**Wrap-up and Take-aways**

**17.30 End**

# Abstracts

## Should DH Consult an Oracle for the “Copyright Case of the Decade”?

*Erik Ketzan*

(Birkbeck, University of London)

Google LLC v. Oracle America, Inc. is currently being appealed to the Supreme Court of the United States, and has been hyped, in Google’s own petition, as “the copyright case of the decade”.<sup>[1]</sup> It remains to be seen whether the Supreme Court will take the case, but it is worth considering the impact of a possible victory for Oracle on changes to law that may affect DH.

“The case that never ends”, as it has also been called (Finley, 2018), has wound through courts for almost a decade. In 2010, Oracle sued Google claiming that Android infringes Oracle’s Java-related copyrights. Generally, copyright protects computer code that meets the threshold of originality but not “the functional aspects of a computer program, such as the program’s algorithms, formatting, functions, logic, or system design.” (U.S. Copyright Office, 2017) While both sides agreed that Google wrote or acquired 97% of the code in its Java APIs, Oracle claimed that the structure, sequence, and organization of Java APIs were copyrightable, and that Google replicated these elements in 37 API packages. In 2012, a federal district court held that the Java APIs’ “command structure is a system or method of operation under Section 102(b) of the Copyright Act and, therefore, cannot be copyrighted”, and that for those elements of Java’s command structure included in Google code, “Duplication of the command structure is necessary for interoperability”, that “while the Android method and class names could have been different from the names of their counterparts in Java and still have worked, copyright protection never extends to names or short phrases as a matter of law.”<sup>[2]</sup> In 2014, the Federal Circuit reversed this on appeal, holding that APIs could be copyrightable, and sent the case back to determine fair use. Google petitioned the Supreme Court but was denied. In 2016, a California jury found that Google’s use of the Java APIs was a protected fair use. However, in 2018 the Federal Circuit reversed this, also, holding that Google’s use of the Java APIs was not fair use as a matter of law. Google now appeals this decision to the Supreme Court.

If Oracle prevails, APIs and their command structures will be recognized as copyrightable (by one Circuit Court), which is opposed by a wide arrange of stakeholders. A brief by 77 notable computer scientists calls this “a significant threat to the technology sector and to the public. [...] API creators would have veto rights over any developer who wants to create a compatible program – regardless of whether she copies any literal code from the original API implementation.”<sup>[3]</sup> The Computer & Communications Industry Association states that this decision “would inhibit the transfer of data between users with different computing environments.”<sup>[4]</sup> In opposition, Oracle argues that Google “inflicted incalculable market harm on Oracle” and downplays arguments fearing “the future of ‘interoperable’ software. [...] Google has conceded that it purposely made its platform incompatible with Oracle’s. [...] Google cites not a single case [...] that has ever held that copying this volume of code (or this much structure and organization) into a competing work is fair.”<sup>[5]</sup>

The effect on DH research of a Supreme Court victory by Oracle should at least be considered at this stage. Many DH projects use available APIs (e.g. Amazon API in Berry et al., 2015; Trove API in Kutty et al., 2019) or create their own (e.g. Hankinson & Craig-McFeely, 2017), but use is not copying. The extent to which DH researchers create APIs based on the command structures of existing commercial APIs should be better understood via involving stakeholders in this workshop.

Another consequence of this case could be a circuit split regarding the copyrightability of short words and phrases, in code and arguably in general. This should concern DH, as much DH research is based on snippets of texts. Fair use, however, remains a defense, and snippets in DH benefit from the liberal state of fair use and “transformative use” following especially the Google Books cases.[6]

## **References**

**Berry, D., Borra, E., Helmond, A., Plantin, J.-C., and Rettberg, J.** (2015). The Data Sprint Approach: Exploring the field of Digital Humanities through Amazon’s Application Programming Interface. *Digital Humanities Quarterly* 9(3).

**Finley, K.** (2018). The Case that Never Ends: Oracle Wins Latest Round vs. Google. *Wired.com*, available at: <https://www.wired.com/story/the-case-that-never-ends-oracle-wins-latest-round-vs-google/>

**Hankinson, A. & Craig-McFeely, J.** (2017). Building the new DIAMM: Linking and sharing data for medieval musicology, *DH2017 Conference Book of Abstracts*, pp. 71-73.

**Kutty, S., Nayak, R., Turnbull, P., Chernich, R., Kennedy, G., and Raymond, K.** (2019). PaperMine r— a real-time spatiotemporal visualization for newspaper articles, *Digital Scholarship in the Humanities*, 28 January 2019.

**United States Copyright Office (2017).** Circular 61: Copyright Registration of Computer Programs. Available at: <https://www.copyright.gov/circs/circ61.pdf>

---

[1] *Petition for Writ of Certiorari, Google LLC v. Oracle America, Inc., 886 F.3d 1179 (Fed. Cir. 2018), No.18-956.*

[2] *Oracle America, Inc. v. Google LLC, United States District Court for the Northern District of California, Case No. 3:10-cv-03561-WHA. (2012).*

[3] *Brief of Amici Curiae Computer Scientists in Support of Petitioner, Google LLC v. Oracle America, Inc., 886 F.3d 1179 (Fed. Cir. 2018), No.14-410. Pg. 3-4.*

[4] *Brief of Amici Curiae The Computer & Communications Industry Association in Support of Petitioner, Google LLC v. Oracle America, Inc., 886 F.3d 1179 (Fed. Cir. 2018), No.14-410. Pg. 4.*

[5] *Brief of Opposition on Petition for Writ of Certiorari, Google LLC v. Oracle America, Inc., 886 F.3d 1179 (Fed. Cir. 2018), No.18-956. Pg. 2.*

[6] *Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2d Cir. 2014). Authors Guild et al. v. Google Inc., 804 F.3d 202 (2d Cir. 2015)*

## **EU Exceptions for Text and Data Mining for Research Purposes : from a Dry Spell to a Crisis of Abundance?**

*Pawel Kamocki, Andreas Witt*  
(IDS Mannheim / CLARIN-D)

Already in the 1960s (Demsetz, 1967) it was predicted that new technologies will trigger the expansion of new exclusive rights. This prophecy was fulfilled and the control that right holders can now exercise over the use of their content is unprecedented. In the digital environment, almost every act of exploitation of content amounts to a restricted act (even if it is only temporary reproduction on a device's screen) and therefore normally requires permission from a right holder. In this statutory landscape, it can be said that copyright is evolving mostly through modifications of the system of exceptions and limitations (according to the « restricted by default » logic). Paradoxically enough, copyright exceptions have never before been so numerous. Teaching and research activities, which in the mid-20th century were still regarded as almost completely irrelevant from the point of view of copyright (under the *de minimis non curat praetor* principle), at the very beginning of the 21st century received their own, dedicated exception, harmonised across the EU. Article 5.3(a) of the Directive on Copyright in the Information society (InfoSoc) provide Member States with the possibility to exempt copying and sharing of works for non-commercial teaching and research purposes from the obligation to receive right holders' permission.

This exception – often implemented quite narrowly in national laws – was deemed insufficient by the scientific community. After a long debate spurred mostly by the Hargreaves Review of Intellectual Property and Growth (2010), a new generation of research exceptions appeared in some Member States to specifically address text and data mining (TDM) for research purposes. Such exceptions were adopted e.g. in the UK (in 2014), in France (in 2016) and in Germany (in 2017). Despite being based on the same article 5.3(a) of the InfoSoc Directive (and therefore necessarily limited to non-commercial purposes), these exceptions differ as to what acts can be performed, by whom and under what circumstances. They also address the issue of their relation with contractual clauses and Digital Rights Management in different ways.

In 2019, a new Directive on Copyright in the Digital Single Market (DSM) was adopted, containing not one, but two exceptions for TDM. The first one (art. 3) is mandatory (i.e. it has to be adopted by all the Member States) and it allows research organisations (as well as cultural heritage institutions) to carry out TDM of works they have lawful access to, for both commercial and non-commercial purposes. The second one (art. 4) is much broader and may potentially cover any TDM operations. According to some commentators (e.g. Rosati, 2018), the second exception could act as a 'grandfather clause' and allow Member States that already adopted TDM exceptions such as the UK, France or Germany) to largely maintain their current rules in this respect. Despite their permissive appearance, the TDM exceptions in the DSM Directive have a complex and rather unclear relation with Digital Rights Management, which may mean that the battle for TDM will continue, but will simply move to a different ground.

This presentation will briefly present the existing TDM exceptions in some EU Member States), with a particular focus on Germany, and then discuss how the entry into force of the new DSM Directive is likely to modify it. It will try to answer the question whether and to what extent the new European TDM exceptions can be beneficial to the DH community.

## **Advancing digital scholarship activities with Canadian and global comparative copyright law**

*Kim Nayyer*

(Edward Cornell Law Librarian, Cornell University)

This presentation will discuss Canada's copyright law in the context of global knowledge. Techniques for interpretation and advancement of copyright law and permissible activities for digital scholars will be demonstrated and will include (i) historical origins of Canada's copyright regime; (ii) comparative analysis with other Statute of Anne-derived laws, including the US copyright law concepts such as transformative use; (iii) the place of international and multilateral instruments in the construction of Canadian copyright law; (iv) the real if problematic application of the private law concept of license in the application of copyright law; (v) the essential role of judicial interpretation in advancement of the law to match the realities of global digital knowledge; and (vi) concepts of knowledge of Indigenous peoples locally and globally.

(i) First we will briefly review the historical origins of Canada's copyright regime. Canada is among those jurisdictions whose copyright legislation derives from the Statute of Anne. The essence of this law was the relatively simple establishment of a measure of control over the reproduction of work when availability of mass printing Kim Nayer: Advancing digital scholarship activities with Canadian and global comparative copyright law (Proposal for Copyright and Humanities Research: A Global Perspective DH2019 pre-conference workshop) widened. Canada's Copyright Act has moved well beyond that (1) early purpose and now offers a balance between protection of the rights of the copyright holder and protection of the rights of the user of the copyright material. (2) Though the Act derives from the Statute of Anne, the civil law tradition of Quebec may reflect in the French language title of the act, *Loi sur le droit d'auteur*. (3) The concept of *droit d'auteur*, or the right of the author, is on its face at odds with the concept of the right to copy; this likely is at the heart of the Canadian concept of balance of author and user rights. (4)

(ii) Next we will discuss the value of comparative analysis with other Statute of Anne-derived laws, including the US copyright law concepts such as transformative use. An important feature of Canada's copyright law is the shared historical origin with other Commonwealth systems or jurisdictions that trace their laws to the Statute of Anne. Comparative analysis is often made with copyright legislation of historically related Commonwealth jurisdictions such as the UK and New Zealand. Without a shared Commonwealth aspect and with its different constitutional framework, comparative analysis with the United States is often overlooked or dismissed as insufficient for analogy. This presentation will show such reticence should be discontinued, both because the jurisdictions share pre-constitutional origins in the Statute of Anne, and because the geographic, societal, and trade proximity of the jurisdictions make US law a helpful and influential comparator. (5)

(iii) We will follow with a look at the place of international and multilateral instruments in the construction of Canadian copyright law. Copyright law in Canada is largely reflected in federal statute, but other influences such as treaties and multilateral and international agreements inform copyright application.

(iv) Next, we will learn about and discuss the real, if troublesome, application of the private law concept of license in the application of copyright law. Private law matters can have a profound and problematic impact on copyright-related activities of digital scholars. The most prominent are licenses, an application of contract law; conflicts of applicable law is another.

(v) As a final technique in understanding the application of copyright law in the work of digital scholars, we will study the essential role of judicial interpretation in development of the law to match digital and global realities. Judicial interpretation of statutory rules create a layer of law

not readily evident in the language of the legislation alone. No interpretation of statutory provision is complete without the study of judicial discussion of it and related provisions. This is the nature of common law and is crucial to understand. The courts preceded the legislature in advancing some copyright law concepts useful in cultural and heritage institutions. This presentation will demonstrate the benefit of activities that push boundaries to clarify permissible activities and achieve a copyright law compatible with global digital needs and realities. (6) Not-for-profit digitization, preservation, transformation, and reuse and global knowledge exchange are examples of activities where legislative protection is uncertain but may be proved permissible.

(vi) We will conclude by asking questions about the impact of concepts of knowledge of Indigenous peoples locally and globally. Canadian copyright understandings are grappling with colonization and its suppression of concepts of systems of Indigenous knowledge. Indigenous peoples operated with systems of laws that continue to this day and which may cross modern political borders. (7) These systems of understanding of Indigenous knowledge protection from unauthorized reproduction, rights of reuse, and stimulations of creation are not incorporated in enacted copyright legislation in Canada, but must be addressed. International agreements contemplate Indigenous knowledge but researchers and cultural heritage and educational organizations require clarity in the scope of national adoption and the conceptualizations of protected and permitted activities. Whether such agreements are sufficient in an era of global knowledge depends on what is meant by sufficiency and through whose lens the question is explored. (8)

1 Copyright Act, RSC 1985, c C-42 <<http://canlii.ca/t/53jfk>>

2 CCH Canadian Ltd. v. Law Society of Upper Canada, [2004] 1 SCR 339, 2004 SCC 13 (CanLII), <<http://canlii.ca/t/19lpo>>.

3 N Tamaro, *The Annotated Copyright Act* (Toronto: Carswell, 2019, annual)

4 Théberge v. Galerie d'Art du Petit Champlain inc., [2002] 2 SCR 336, 2002 SCC 34 (CanLII), <<http://canlii.ca/t/51tn>>

5 Ariel Katz. "Fair Use 2.0: The Rebirth of Fair Dealing in Canada" Ottawa, ON in *The Copyright Pentology: How the Supreme Court of Canada Shook the Foundations of Canadian Copyright Law* (2013) p. 93 - 156 Available at: [http://works.bepress.com/ariel\\_katz/22/](http://works.bepress.com/ariel_katz/22/)

6 See CCH Canadian Ltd. v. Law Society of Upper Canada, [2004] 1 SCR 339, 2004 SCC 13 (CanLII), <<http://canlii.ca/t/19lpo>>; *Entertainment Software Association v. Society of Composers, Authors and Music Publishers of Canada*, [2012] 2 SCR 231, 2012 SCC 34 (CanLII), <<http://canlii.ca/t/fsov7>>; *Rogers Communications Inc. v. Society of Composers, Authors and Music Publishers of Canada*, [2012] 2 SCR 283, 2012 SCC 35 (CanLII), <<http://canlii.ca/t/fsov9>>

7 F Gaudreault-Desbiens, « La critique autochtone de l'appropriation artistique ou le droit d'auteur dans la tourmente identitaire », (2000) 40(2) *Droit et Cultures* 133-152 (France)

8 Nayyer, K. (2002). *Globalization of Information: Intellectual Property Law Implications*. *First Monday*, 7(1). doi:<https://doi.org/10.5210/fm.v7i1.925>

## Privacy and Property Issues in the Computational Study of Social Reading

Simone Rebora, Moniek Kuijpers, Piroska Lendvai  
(DH Lab, University of Basel, Switzerland)

Our project *Mining Goodreads: a text similarity-based approach to measure reader absorption* aims at studying the phenomenon of "social reading" with the support of computational methods. The main goal of the project is measuring different levels of absorption, i.e., the physical/cognitive/emotional involvement in narrative worlds, in different books and genres, by analyzing reader reviews published on the online platform *Goodreads*. Preliminary analysis showed that multiple sentences in *Goodreads* reviews semantically overlap with the items of the *Story World Absorption Scale*, a questionnaire developed with the specific goal of measuring four different aspects of absorption: attention, narrative transportation, emotional engagement, and mental imagery.

Our project is funded by the Swiss National Science Foundation (SNSF), with an 18-month research programme that started in December 2018. The main focus of the programme is the manual annotation of a selection of *Goodreads* reviews (with five annotators working in parallel), in order to train machine learning algorithms to automatically recognize signs of absorption.

With reference to the subject of the workshop, we are currently dealing with the following issues:

- *Goodreads* terms of use do not allow copying and reproducing the data hosted on the website. However, these restrictions refer clearly to commercial re-use and do not state anything about data collection for research purposes;
- According to *Goodreads* terms of use, each piece of *Goodreads* content is owned by the individual users themselves, thus authorization for the use of such data cannot be granted by *Goodreads*;
- Data might be sensitive for privacy reasons even if the reviews focus on reading experiences with (almost) no references to private life and identity (most of the users adopt nicknames or pseudo-profiles).

Currently, SNSF cannot guarantee legal protection in case of issues related to property and privacy, while the University of Basel's legal department advised against downloading or sharing any corpora derived from *Goodreads*. While workarounds for enriching the raw data for our purposes could be found e.g. by producing online annotations, machine learning training cannot be performed without data copying. Additionally, sharing of the dataset that our project focuses on would be best practice for transparency and reproducibility, and SNSF strongly advises this. It should be noted that multiple studies have already been published involving *Goodreads* data, while the new EU Copyright Directive suggested the introduction of significant copyright exceptions for research purposes.

Because these problems seem to be shared by all projects dealing with the emerging phenomenon of social reading, we think that a detailed discussion of such issues in the context of the *Copyright and Humanities Research: A Global Perspective* workshop may prove extremely relevant and productive.

## Copyright and Impact Factors in the Humanities - Decolonizing Knowledge Systems and Assessments

*Nirmala Menon*

(Indian Institute of Technology Indore)

Scholarly Publishing in India is at an uncertain but crucial moment- on the one hand, scholars from India are publishing more and more (across disciplines) in prestigious journals and publishing outfits around the world; on the other, the economic premise and objective of the same prestigious publishing houses is increasingly under scrutiny and there are attempts by scholars worldwide to untangle the stranglehold of publishers on access to research and costs of evaluation. However, every other day you also come across articles that allege, "India is home to some of the largest predatory scholarly publishing ventures". The devil and deep sea, if you will. It is in this context that we need to expand scholarly publishing avenues that are born digital, emphasize rigorous peer review and multilingual forms. What research tools and infrastructures are needed to study, document, annotate, analyze, archive, cite, and work with (in general) digital objects, especially those in Indic languages? Before we get to this core question in the research arena we need to confront the reality- that there are as yet very few digital objects that are created, developed and preserved in Indian languages. Shodhganga, a digital repository for Dissertation and Theses from universities in India is a national digitization project that has now grown substantially as a scholarly resource and the National Digital Library a project that plans to make curricular access electronic- both extraordinary projects- are still limited to works that are primarily written in English.

This is especially unfortunate in Humanities research as universities in India have a robust scholarship agenda and have some niche publications that focus on the works in different languages as recognized in the 8th schedule of the constitution of India.

1. That broad adoption of Open Research Practices (ORP) by the global research community is, in most circumstances, a public good and has the capacity to improve the practice and use of research.
2. That the transformative change described in P1 will require broad popular and practical acceptance by individual researchers parallel to that largely already achieved among funders, governments, universities, libraries, and policy makers.
3. That developing the consensus described in P2 requires us to learn across and transcend disciplinary, regional, sectoral, and other boundaries that characterise work in this area.

On the basis of these premises, I consider four main research questions:

1. Why and to what degree has the adoption of ORP by individual members of the research community lagged transformative change in other industries?
2. What differences exist in the adoption of ORP within different sectors, regions, and disciplines? What factors (cultural, linguistic, disciplinary, ethical, political, practical) have contributed to a greater or lesser adoption in each case.
3. What strategies have proved best to mobilise ORP within the research community? To what degree do these translate across research cultures, whether disciplinary or other.

I will discuss the Open Publishing Project at IIT Indore and a brief presentation about developing metrics for Humanities scholars across India that factor in the intersections but also divergences from the production and evaluation of publications in the Global North.



## Copyright and Post-disaster Archiving

*Paul Millar*

(University of Canterbury)

Beginning on 4 September 2010, a sequence of major earthquakes hit New Zealand's Canterbury province with devastating effect. The most disastrous quake – on 22 February 2011 – claimed 185 lives, destroyed much of the centre of the city of Christchurch, and made large swathes of quake-prone suburbs uninhabitable. One of many responses to the disaster was the establishment of a Digital Humanities cultural heritage memory project led by the University of Canterbury – the UC CEISMIC Canterbury Earthquakes Digital Archive (CEISMIC) – available at <http://www.ceismic.org.nz>. Initiated in May 2011 CEISMIC set out to collect images, stories, and media related to the earthquakes and their aftermath, for the purposes of commemoration, teaching, and research.

In this workshop session Paul Millar discusses the experience of the CEISMIC project in dealing with copyright issues. He outlines the relevant law it stands and is applied in New Zealand, and discusses some of the unique situations the project encountered as the team tried to ensure openness, inclusivity and rigour within a copyright compliant framework during a period of trauma and transition.

### Biography

Paul Millar is Deputy Pro-Vice-Chancellor of the College of Arts and Professor of English and Digital Humanities at the University of Canterbury / Te Whare Wānanga o Waitaha in Christchurch, New Zealand. He led the establishment of New Zealand's first Digital Humanities teaching programme and developed the CEISMIC Canterbury Earthquakes Digital Archive. He is currently the president of the Australasian Association of Digital Humanities (aaDH).