UNDERSTANDING ORGANIZATIONAL RESPONSES TO INNOVATIVE DEVIANCE: A CASE STUDY OF HATHITRUST

by

Alissa Lorraine Centivany

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in The University of Michigan
2016

Doctoral Committee:

Professor Paul N. Courant, Chair
Associate Professor Steven J. Jackson, Cornell University
Associate Professor Ariel Katz, University of Toronto
Professor John L. King
Professor Jason D. Owen-Smith

DEDICATION

For Lali Bee, Mirabelle, and Bobby.

ACKNOWLEDGEMENTS

Words can only begin to hint at the depths of my gratitude and admiration for the advisors, mentors, teachers, colleagues, friends, and loved ones who have supported and inspired me on this journey. In particular, I thank: Steve Jackson, Paul Courant, John King, Tom Finholt, Jessica Hullman, Michael Sperling, Brigitte Miesch, Charles Centivany, Helen Hedden, Suzanne Smulsky, Charley Centivany, Aimee Centivany, Lali Bee Centivany-Glushko, Mirabelle Centivany-Glushko, and Bobby Glushko.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter I: Introduction

In October, 2008, thirteen universities announced the launch of HathiTrust, described at the time as a shared digital repository consisting of over two million volumes.[1] In the intervening eight years, HathiTrust has become much more than a safe, secure place to store digital materials. HathiTrust is a now a *partnership* of more than one hundred academic and research institutions united around the goal of contributing to the common good by collecting, organizing, preserving, communicating, and sharing the record of human knowledge.[2] It is a *digital library* with a fully indexed and searchable catalog; members of the general public can read and download public domain works for HathiTrust's website, hathitrust.org, and persons with certified print disabilities can get full access to HathiTrust's entire corpus; HathiTrust is "by, of, and for libraries and librarians."[3] HathiTrust is a *research center* offering computational access for text mining and other forms of non-consumptive research by nonprofit and educational researchers. HathiTrust is an *organization* with an executive director, committees and advisory boards, employees, bylaws, and a membership with voting rights. Notwithstanding he collective nature of the organization, HathiTrust is *part of a single institution*, offered as a service of the University of Michigan. In addition, HathiTrust (along with several of its member institutions and their representatives) was a *defendant* in a major copyright infringement lawsuit. As of January, 2016, HathiTrust contained nearly fourteen million volumes, six hundred twenty-two terabytes of information, thirty-nine percent of which is known to be in the public domain, which leaves the remaining sixty-one percent subject to copyright.[4] But the way HathiTrust

[1] HathiTrust Press Release, October 13, 2008, "Launch of HathiTrust – October 13, 2008," available at http://www.hathitrust.org/press_10-13-2008. HathiTrust was launched jointly by the twelve-university consortium known as the Committee on Institutional Cooperation ("CIC") and the eleven university libraries of the University of California system (the UC's ten research university libraries plus the system-wide California Digital Library ("CDL"). The CIC institutions circa 2008 included: University of Illinois at Chicago; University of Illinois at Urbana-Champaign; Indiana University; University of Iowa; University of Michigan; Michigan State University; University of Minnesota; Northwestern University; Ohio State University; Pennsylvania State University; Purdue University; University of Wisconsin-Madison.

[2] HathiTrust, "Mission and Goals," available at https://www.hathitrust.org/mission_goals

[3] Interview with an architect of HathiTrust.

[4] See HathiTrust Digital Library Statistics and Visualizations at http://www.hathitrust.org/statistics_visualizations

looks now does not reveal all that much about *how* or *why* it turned out that way. This research will describe and explain how and why HathiTrust came to be, and came to become *this* HathiTrust.

This research will trace the emergence and evolution of HathiTrust as part of a larger and more complex story about processes of sociotechnical transformation. Scholars have long-recognized the entangled and mutually constitutive relationships among technology and social practice.[5] Transformation in this context is characteristically dynamic, non-linear, multi-directional, and guided by careful deliberation and planning as well as luck and accident.[6] Technologies open new spaces for social practice and engagement and, through processes of invention, use, modification, and regulation are also themselves socially constructed.[7] The story of HathiTrust teaches important lessons about the intersections of technology and social practice, particularly as it relates to organizational and institutional behavior.

This work aims to contribute to understandings of sociotechnical transformation by adding an additional important layer of inquiry, namely how emerging technologies, social practices, *and law and policy* interact and co-evolve. Transformation is a concept of growing concern in the world of copyright. Indeed, over the last twenty years, transformation has become so central to fair use determinations that one would be hard-pressed to find a fair use discussion that didn't address, at length, questions around transformation. Transformation in the copyright context is related to the broader notions of sociotechnical transformation mentioned above, but it is also distinct from those concepts. In copyright law, transformation is made quantifiable through the application of elements, factors, comparisons, analogies and tests. Part of this research, therefore, will be to explore and import transformation in the copyright context into these broader conceptions of sociotechnical transformation.

---

[5] See e.g. Kling, R. (Ed.). (1996). *Computerization and controversy: value conflicts and social choices*. Morgan Kaufmann; Kling, R. (1991). Computerization and social transformations. *Science, Technology & Human Values*, *16*(3), 342-367; Bijker, W. E., Hughes, T. P., Pinch, T., & Douglas, D. G. (2012). *The social construction of technological systems: New directions in the sociology and history of technology*. MIT press;

[6] See e.g. Jackson, S. J., Gillespie, T., & Payette, S. (2014, February). The policy knot: Re-integrating policy, practice and design in Cscw studies of social computing. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 588-602). ACM.

[7] See e.g. Ackerman, M. S. (2000). The intellectual challenge of CSCW: the gap between social requirements and technical feasibility. *Human-computer interaction*, *15*(2), 179-203.

Ways of Understanding Sociotechnical Transformation

Developments in information technology are fundamentally altering the essential mechanisms by which information and knowledge are created, organized, shared, and disputed. The advent of the Internet, the creation of technical protocols and standards, and improvements in digitization combine to enable content to be moved rapidly, at relatively low cost, and without significant degradation. Today, anyone with an Internet connection can gain tremendous access to the world's shared cultural record, a feat unfathomable twenty years ago. And regardless of the forms in which these works are fixed, whether they are literary, audio(-visual), or pictorial, graphical, or sculptural,[8] opportunities to participate in, interact with, modify, build upon, and transform the works abound.

This is not to suggest, however, that advances in information and communication technologies have proceeded lock-step with positive sociotechnical transformations. While the retrospective conversion of twenty million analog works was unquestionably impressive as a technical matter, and as a feat of engineering throughput, the HathiTrust story (and related Google Books story which is discussed in part in this work) demonstrates that sociotechnical transformation involves much more than stewarding content from one format or technical environment to another. The technological aspects of digitization are perhaps the simplest, least controversial evidence of transformation in the context of mass digitization. Mass digitization spurred other complex, complicated, and often murkier transitions and reconfigurations with respect to the existing copyright regime and the traditions and practices of knowledge and memory institutions (universities and libraries). Changes in information and communication technologies have significantly expanded the range of means available for engaging with cultural and intellectual works, however legal and social institutions have tended to lag behind in terms of legitimating many of those means. This tension, and its relationships to sociotechnical transformation, is the chief concern of this thesis.

The analytic approach adopted in this research takes as its jumping off point the observation that a central and recurring theme in research on sociotechnical transformation is the notion that technological change puts existing institutions — legal, political, and/or social — under significant strain. The strain can be characterized in a number of ways but a fair generalization might be to say that technological change often disrupts or disturbs the relative

[8] 17 U.S.C. §102(a)

equilibria between innovation and control. For example, new technologies are often characterized as dynamic, disruptive, and/or destructive forces in comparison to the relatively conservative, slow-moving, self-preserving forms adopted by many institutions.

This theme appears in a number of works notable for their explorations of sociotechnical transformation. Elizabeth Eisenstein's work on the emergence and effects of the printing press highlighted the social, political, religious, and legal tensions raised by the transition from manuscript culture to print culture.[9] Joanne Yates explored how changes in information and communication technologies shape and are shaped by organizational and managerial contexts in the first part of the twentieth century, leading to a rise in the mechanization of work processes.[10] Richard John offered a detailed historical account of the complex ebbs and flows — particularly among innovation, federal policymaking, and social response and adoption — of telecommunications technology development in the United States.[11] Thomas Park Hughes' research on the development of electrical systems couples historical observations with new frameworks and analytic tools for describing and explaining what he calls the "social construction" of technology.[12] A common thread linking these relatively disparate moments in the history of technology is the centrality of disequilibrium and strain in processes of sociotechnical transformation.

A number of theories and analytic approaches have developed for explaining, describing, and understanding the meanings of disequilibrium and strain in processes of sociotechnical transformation. For example, the so-called Chicago School of sociologists recognized nearly a century ago that rapid and significant changes in technology can lead to social disorganization as they undermine the web of normative controls (expectations, rules, laws, etc.) that typically govern our choices and interactions.[13] In other words, significant changes in technology alter the possibility space of our activities *and* alter the context of decision-making.

In addition, in his structural strain theory, sociologist Robert K. Merton argued that a mismatch or imbalance in society between culturally accepted goals and the availability of

[9] Eisenstein, E. L. (2005). *The printing revolution in early modern Europe*: Cambridge University Press.
[10] Yates (1993) *Control Through Communication: The Rise of System in American Management.*
[11] Richard John, *Network Nation*
[12] Hughes, T. P. (1993). *Networks of power: electrification in Western society, 1880-1930*: JHU Press. Hughes, T. P. (1987). The evolution of large technological systems. *The social construction of technological systems: New directions in the sociology and history of technology*, 51-82.
[13] Thomas, W. I., & Znaniecki, F. (1918). *The Polish peasant in Europe and America: Monograph of an immigrant group* (Vol. 2). University of Chicago Press.

legitimate means to accomplish those goals can push people (and communities) toward deviance, i.e. rules, norms, laws no longer serve as effective restraints on behavior because of the strain caused by this disequilibrium. Merton's research was not particularly concerned with the role of technological change or intellectual property laws[14] but, when considered in conjunction with the previously discussed work and approaches, it seems plausible that technological change has implications for both the ways societies generate and internalize broad social or cultural goals and, perhaps more importantly, for the means society creates and/or legitimates for accomplishing those goals.

For purposes of clarity, we can break Merton's theory into two parts. The first part we can call the "end" which describes the generation and acceptance of broad social goals. Applied to the topic of this thesis, we can readily identify a widely accepted goal as being the promotion of the progress of society through the creation, communication, and use of intellectual and cultural works. In the United States, this goal is uncontroversial and clearly articulated in the U.S. Constitution which gives Congress the power to "promote the Progress of Science and useful Arts …."[15] The second part of Merton's framework deals with the "means" which are the ways (paths, tools, opportunities, etc.) society makes available and/or deems legitimate for accomplishing a given end. The Constitution gives clues here as well, saying that we can accomplish the goal "by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries …." Those means are, of course, further articulated in the Copyright Act (and the Patent Act, although that is beyond the scope of this thesis). Taken in combination, a hypothesis begins to emerge suggesting that interactions between existing legal frameworks, emerging technologies, and social practices around the creation, dissemination, and use of cultural works results in a disequilibrium or imbalance between the goals of copyright and the availability of legitimate means for accomplishing those goals.

This hypothesis seems to be somewhat borne out in the existing debates amongst rightsholders, content and media industries, educational institutions, and a variety of advocacy groups concerning networked digital technologies and mass digitization. The following brief

[14] A chief focus of Merton's work was on the often socio-political structured inequalities within society. His perspective was informed by, and perhaps can be understood as a response to, trends of urbanization, immigration, industrialization, and increased mechanization of aspects of human activity and interaction (such as work) of the late nineteenth and early twentieth centuries. While technological change played a role in this, it was not a primary focus of Merton's work. In addition, as far as I can tell, his work did not reference intellectual property law at all.
[15] U.S. Const. Art. I, § 8, cl. 8.

discussion highlights four contrasting perspectives on the relationship between existing laws and emerging technology and provides a back-of-the-napkin sketch of the disequilibrium in the copyright world. The following discussion illustrates how different, sometimes oppositional, approaches can simultaneously resonate with the facts surrounding the mass digitization of libraries' in-copyright collections.

The Copyright Disequilibrium

Discourse around copyright law and technological change is rich, diverse, and nuanced. It also tends to be fairly strident[16] and characterized by rhetorical excess.[17] I will therefore not attempt to summarize or distill the full extent of copyright debates around emerging technologies here. Rather, I will provide a simplified snapshot of the basic perspective of some of the key stakeholders in the mass digitization context.

The first perspective, held by some rightsholders, is that existing copyright laws are *under-restrictive* with respect to technological change. In the context of mass digitization, the Authors Guild and some proprietary publishers have been proponents of this viewpoint.[18] In addition, some individual authors have been outspoken about their desire to see copyright laws strengthened as a way of protecting and preserving their intellectual property and incentivizing their continued productivity. Margaret Atwood, for example, a prolific and highly esteemed Canadian author, has expressed public skepticism and criticism about digitization. Her position is neatly captured in a diptych (Fig.1) created by Atwood and presented at a 2011 conference.[19] In addition to publicly presenting this (and other) illustrations, Atwood emailed the drawing to the chief librarian of a major academic research library pursuant to some private discussions regarding her concerns about the library's digitization practices and policies.

[16] Samuelson, P. (2010). Copyright Principles Project: Directions for Reform, The. *25 Berkeley Tech. LJ 1175*.
[17] Patry, W. (2009). *Moral panics and the copyright wars*: Oxford University Press.
[18] See e.g. The Authors Guild (2015). "The Authors Guild's Top Legislative Priorities – 2015," available at https://www.authorsguild.org/wp-content/uploads/2015/05/2015-AG-Top-Legislative-Priorities.3.23.15.pdf
[19] O'Reilly Media, *Tools of Change for Publishing*, Feb. 14-16, 2011. New York, NY. Information regarding Atwood's presentation, *The Publishing Pie: An Author's View*, is available at http://www.toccon.com/toc2011/public/schedule/detail/17569, last visited 6/24/15.

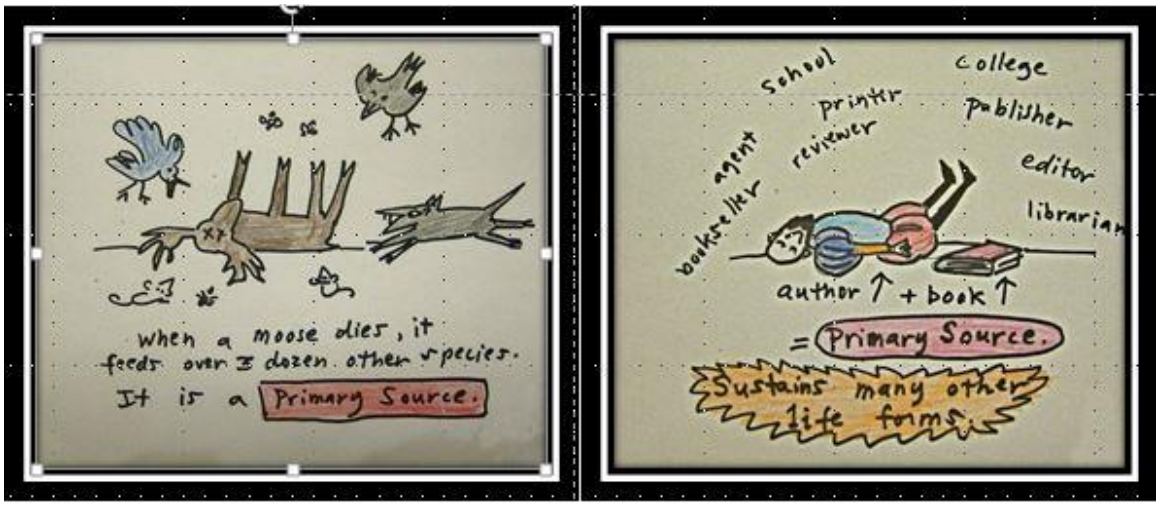


Figure 1. Margaret Atwood drawing

In this diptych, Atwood compares the ecological interconnectedness of the natural world with that of the literary world. Like the moose (another Canadian national treasure), authors are a "primary source" of life-sustaining energy for numerous others including educational institutions (colleges, universities, and other schools), libraries and librarians, and various others integrated in publication and distribution chains. In Atwood's view, "Everything else in the world o'publishing [sic] depends on them."[20] In both images, the primary source is depicted as dead although, in her presentation, Atwood is quick to recognize that unlike the moose, authors need not perish in order to sustain others on the food chain although, she notes, many dead authors been "very lucrative," generating lots of money for lots of people.[21] Under this view, which some authors share, digitization (and electronic distribution more generally) jeopardizes authors' abilities to fully control the exploitation of their creative works, thus potentially undermining their livelihoods and, by extension, the livelihoods of others within the literary ecosystem.

A second, contrasting perspective held by others including authors[22], technologists[23], and public interest advocates[24] is that existing copyright laws may be *over-restrictive* with respect to

[20] Atwood, M. (2011). *The Publishing Pie: An Author's View,* presentation at O'Reilly Media Tools of Change for Publishing conference, Fed. 15, 2011, https://www.youtube.com/watch?v=-6iMBf6Ddjk at minute 9:25-9:35.
[21] Atwood, M. (2011). *The Publishing Pie: An Author's View,* presentation at O'Reilly Media Tools of Change for Publishing conference, Fed. 15, 2011, https://www.youtube.com/watch?v=-6iMBf6Ddjk at minute 9:35 – 9:42.
[22] See, e.g. PLOS (Public Library of Science), a nonprofit organization of scientists committed to making the world's scientific and medical literature freely accessible to scientists and to the public,

emerging technologies. Proponents of this view tend to object to a maximalist or rightsholder-centered perspective of the law, focusing instead on the broader public policy-based goals of copyright. For example, those who advocate this perspective may be quick to point out the fact that readers are notably absent from Atwood's illustration even though readers, and the public more generally, are a core to copyright law. This view encourages courts and lawmakers to exercise caution in allowing rightsholders to invoke copyright laws as mechanism for preserving their dying business models particularly given the combinatorial and participatory nature of innovation and creativity across intellectual and cultural domains. The incumbent beneficiaries of an ailing regime, such as the traditional proprietary publishing industry, ought not to be able to rely on a hidebound regime that stifles rather than promotes technological innovation, creativity, and generativity. As technologist and open access advocate Brewster Kahle said: "Copyright laws are intended to promote the progress of science and useful arts for the good of society as a whole — let's let technology help us do that!"[25]

A third viewpoint might be generalized as a law and economics perspective. This viewpoint is less concerned with possible normative consequences of copyright law's means-end imbalance than the prior two viewpoints. Under this perspective, intellectual property laws are essentially trade regulations and therefore the sorts of harm generated by copyright disequilibria are characterized in terms of a copyright system that is inefficient, rife with uncertainty, and is prone to slack.[26] The law and economics perspective originated in the work of Ronald Coase and Gary Becker, and has been adopted by several distinguished jurists including Seventh Circuit judges Frank Easterbrook and Richard Posner and noteworthy intellectual property law scholars

---

https://www.plos.org/about/plos/core-principles/ . See also the Authors Alliance, "promoting the public good by supporting authors who write to be read," at http://www.authorsalliance.org/

[23] See, e.g. Brewster Kahle and the Internet Archive, focused on "providing Universal Access to All Knowledge," at https://archive.org/about/bios.php; See, e.g. Richard Stallman, originator of the "copyleft" concept and founder of the Free Software Foundation and GNU Project, works to "encourage free software to spread, replacing proprietary software that forbids cooperation, and thus make our society better," *Copyright and Community* speech delivered at Wikimania conference, Aug. 4-8, 2005, Frankfurt, Germany.

[24] See, e.g. Creative Commons, a nonprofit organization that provides free legal tools for enabling the "sharing and use of creativity and knowledge," http://creativecommons.org/about.

[25] Interview with Brewster Kahle of the Internet Archive.

[26] Coase, R. H. (1937). The nature of the firm. *economica*, *4*(16), 386-405;
Coase, R. H. (2012). *The firm, the market, and the law*. University of Chicago press; Becker, G. S. (2009). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press.

including William Landes[27] and Wendy Gordon[28]. Gordon's work in particular may be illustrative in the context of this research. She explores the potential of developing what she calls an "ideal" model of copyright law, one that balances rightsholders' incentives and interests in controlling and exploiting their works with justified personal liberties of free speech and fair use *and* the concept of "harm".[29] Fair use determinations, Gordon argued, should include consideration of whether or not a market failure has occurred; when a particular desired use is not available for purchase on the market, and the use would serve the public interest while not substantially impairing the rightsholder's incentives, the use should be permitted as a fair use.[30]

The final viewpoint I will mention is shared by those whose chief concern is the rule of law. On the surface, this perspective does not privilege either rightsholders or users; it avoids much of the rhetoric characteristic of the first two viewpoints discussed. In addition, while it shares the law and economics concern with the efficiency of the copyright system, values including credibility, legitimacy, relevance, and functionality share equal footing under the rule of law perspective. Perhaps not surprisingly, the current Register of Copyrights, Maria Pallante, tends to advocate for this perspective. In an article entitled "The Next Great Copyright Act," Pallante observed that significant changes in technology have raised fundamental questions around the continuing functionality, credibility, and relevance of the existing regime.[31] She observed that "the copyright world which once had predictable, even pristine demarcations, has morphed dramatically"[32] as a result of recent disruptive developments in information and communication technologies, and on that basis she and others have advocated for comprehensive copyright reforms to protect rightsholders, industries, the public and, perhaps most importantly, the rule of law itself.

---

[27] See, e.g. Landes, W. M., & Posner, R. A. (1989). An economic analysis of copyright law. *The Journal of Legal Studies*, 325-363. See, e.g. Landes, W. M., Posner, R. A., & Landes, W. M. (2009). *The economic structure of intellectual property law*: Harvard University Press.

[28] See, e.g. Gordon, W. J. (1982). Fair use as market failure: a structural and economic analysis of the" Betamax" case and its predecessors. *Columbia Law Review*, 1600-1657. See, e.g. Gordon, W. J. (1991). Asymmetric Market Failure and Prisoner's Dilemma in Intellectual Property. *U. Dayton L. Rev., 17*, 853. See, e.g. Gordon, W. J. (1997). Intellectual property as price discrimination: Implications for contract. *Chi.-Kent L. Rev., 73*, 1367.

[29] See, e.g. Gordon, W. J. (2008). Harmless Use: Gleaning from Fields of Copyrighted Works. *Fordham L. Rev., 77*, 2411. See, e.g. Gordon, W. J. (2002). Excuse and justification in the law of fair use: transaction costs have always been part of the story. *J. Copyright Soc'y USA, 50*, 149.

[30] Gordon, W. J. (1982). Fair use as market failure: a structural and economic analysis of the" Betamax" case and its predecessors. *Columbia Law Review*, 1600-1657.

[31] Pallante, M. A. (2012). Next Great Copyright Act, The. *Colum. JL & Arts, 36*, 315.

[32] Pallante, M. A. (2012). Next Great Copyright Act, The. *Colum. JL & Arts, 36*, at 339.

The four perspectives just discussed resonate, to varying degrees, with the stakeholders in the mass digitization project. The perspectives also highlight tensions in discourse around copyright and technological change, highlighting aspects of disequilibrium or mismatch between copyright's goals, its means, and emerging social practices. A common thematic thread connecting these otherwise oppositional perspectives is what some have called "a growing sense of disarray and disjuncture"[33] among existing legal regimes and institutions, social practices and norms, and the new and exciting, but often largely unproven, potentials of emerging technologies.

Innovative Deviance

Circling back to Merton's structural strain theory of deviance, the copyright disequilibrium can be understood within Merton's framework as evidence of a mismatch or imbalance between a culturally accepted goal (in this case the overriding goals of copyright law) and the availability of legitimate means to accomplish the goal (which are perceived differently by different stakeholders but nevertheless reflect a general consensus that copyright laws need reformation). Under Merton's framework, the consequence of this imbalance is a social strain or tension among copyright law and emerging technologies that effectively pushes people toward deviance — behavior that is illegitimate, unorthodox, and/or infringing. Indeed, much of the discourse around copyright and technological change devolves fairly quickly into discussions of enforcement, the idea being that if we only have better tools for deterring and punishing unauthorized uses of copyrighted works, that would solve most of the problem. As we continue with Merton's theory, we can see that the emphasis on enforcement provides an illusory remedy.

Merton contends that illegitimate, unorthodox, and/or infringing behavior is the natural outflow the structural strain caused by a means-end imbalance in society and this "deviance" will follow one of four possible paths: innovation, ritualism, retreatism, and rebellion (see Fig. 2).

[33] Edwards, P. N., Jackson, S. J., Chalmers, M. K., Bowker, G. C., Borgman, C. L., Ribes, D., ... & Calvert, S. (2013). Knowledge infrastructures: Intellectual frameworks and research challenges.
p. 19.

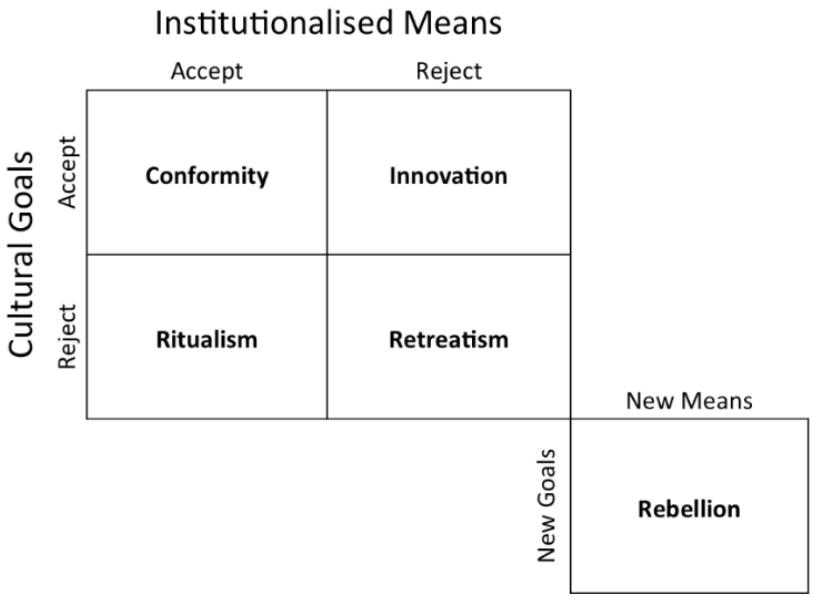


Figure 2. Robert K. Merton's Typology of Deviance

While Merton's typology will be discussed in greater detail in Chapter III, I take this opportunity to highlight one of the four paths, innovative deviance, as the path of primary interest in this research. Innovative deviance can be understood as the "creative use of illegitimate means to obtain valued legitimate ends" and "the rejection of institutional practices but the retention of cultural goals."[34] In addition, Merton describes innovative deviance as "a normal outgrowth of having accepted cultural goals without having been provided with the opportunity to legitimately achieve those goals."[35]

Note that *of the four possible paths of deviance*, innovative deviance is the only one that promotes or sustains the culturally accepted goal; the other three paths reject it. This is significant because, if we accept the premise that technological change produces social strains that lead inexorably to increased levels of deviance (or social disorganization), then rather than devoting our efforts to enhanced enforcement (through deterrence or punishment of the deviance), we might want to step back and distinguish between *good* deviance — i.e. infringement that is utilitarian, socially productive, and/or promotes the overriding goals of

[34] Merton, R. K. (1938). Social structure and anomie. *American sociological review*, *3*(5), 672-682; Merton, R. K. (1968). *Social theory and social structure*. Simon and Schuster.
[35] Merton, R. K. (1938). Social structure and anomie. *American sociological review*, *3*(5), 672-682; Merton, R. K. (1968). *Social theory and social structure*. Simon and Schuster.

copyright with relatively little harm to rightsholders — from *bad* deviance — i.e. infringement or other behaviors that do not promote the overriding goals of copyright.

Merton's structural strain theory, and his concept of innovative deviance in particular, provides the basic scaffolding of the analytic framework for this thesis. However, Merton's theory does not, on its own, enable us to fully describe or explain the complex processes of sociotechnical transformation embodied in exemplars like the HathiTrust case. Additional theories and approaches will be pulled in to help flesh out and extend Merton's contribution.

Continuing with the top-down approach of Merton's structural strain theory, this thesis will explore meaningful synergies with the work of other theorists. For example, Thomas Parke Hughes' concept of a "reverse salient" introduces the observation that processes of sociotechnical transformation may be slowed down or impeded by sticking points.[36] While Hughes' research was primarily concerned with technological reverse salients, the concept might be applied more broadly to include things like laws, policies, and organizational or institutional traditions that seem to counteract forward progression. This perspective adds an additional, potentially useful point of analytic departure to Merton's focus on a means-end disequilibrium.

In addition, while Merton's theory offers clues about the sources of strain and deviance, it does not offer much guidance on how strain might be alleviated and balance restored in the system. Joseph Schumpeter's theory of creative destruction,[37] Albert Hirschman's exit, voice, and loyalty framework, Manuel Castell's work on informationalism and the networked society,[38] and Clay Christensen's work on disruptive innovation[39] could provide some important clues and perspectives on this question.

Perhaps the biggest short-coming of Merton's theory with respect to understanding the processes of sociotechnical transformation is that, in emphasizing the structural aspects of the problem, it does not give adequate attention or consideration to the dispersed, modest, more granular and nuanced aspects of the problem that percolate from the bottom up. For example,

---

[36] Hughes, T. P. (1993). *Networks of power: electrification in Western society, 1880-1930*. JHU Press; Bijker, W. E., Hughes, T. P., Pinch, T., & Douglas, D. G. (2012). *The social construction of technological systems: New directions in the sociology and history of technology*. MIT press.

[37] Schumpeter, J. (1942). Creative destruction. *Capitalism, socialism and democracy*, 82-5.

[38] Castells, M. (2011). *The rise of the network society: The information age: Economy, society, and culture* (Vol. 1): John Wiley & Sons; Castells, M. (2011). *The power of identity: The information age: Economy, society, and culture* (Vol. 2): John Wiley & Sons. Castells, M. (2010). *End of Millennium: The Information Age: Economy, Society, and Culture* (Vol. 3): John Wiley & Sons.

[39] Christenson, C. (1997). The innovator's dilemma. *Harvard Business School Press, Cambridge, Mass*.

Merton's theory does not offer much guidance in terms of helping us understand, describe, and explain how and why a particular form of innovative deviance was selected over others (assuming there are multiple, albeit potentially infringing, ways one might promote the culturally accepted goal). It does not provide clues about why some people, organization, and institutions may choose innovative deviance over conformity. In other words, it does not offer many clues about *how* technological change affects transformation on individual or organizational levels:

> "People who study how technology affects organizational transformation increasingly recognize its dual, paradoxical nature. It is both engine and barrier for change; both customizable and rigid; both inside and outside organizational practices. It is product and process."[40]

Understanding these aspects of sociotechnical transformation requires a more granular analytic approach. To explore these questions, this research will draw upon sociologists Reckless' work on inner and outer containment factors and Hirschi's discussion of the importance of social bonds.[41] In addition, economist Albert Hirschman's exit, voice, and loyalty framework provides some useful clues about how perceptions about whether a problem is an economic one or a political one can influence behavioral outcomes.[42] These questions are also particularly informed by the work of organizational theorist Karl Weick's research on sensemaking, which offers important clues about the relationship between individuals' and organizations' meaning construction and decision-making processes.[43]

Methodology

This thesis explores sociotechnical transformation through the complicated, often messy, co-evolution of copyright law and policy, technological change, and emerging social practices. This research focuses on a particular case, the emergence and evolution of HathiTrust.

The methods used in this thesis combine traditional legal research and analysis with a qualitative case study approach.

---

[40] Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information systems research*, *7*(1), 111-134.
[41] Reckless, W. C. (1961). New Theory of Deliquency and Crime, A. *Fed. Probation*, *25*, 42; Hirschi, T. (2002). *Causes of delinquency*. Transaction publishers.
[42] Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states* (Vol. 25). Harvard university press.
[43] Weick, K. E. (1995). *Sensemaking in organizations* (Vol. 3). Sage; Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization science*, *16*(4), 409-421.

This study seeks to answer the following three research questions:

- RQ1: How and why did the University of Michigan engage in mass digitization of in-copyright works and how was its sensemaking and decision-making reflected in intra-organizational practices, processes, mechanisms, policies, and tensions?
- RQ2: How and why did these (conceptions, decisions, practices, processes, mechanisms, policies, and tensions) prompt the genesis of HathiTrust and how have they continued to evolve over time and in response to internal and external factors?
- RQ3: How might HathiTrust's emergence and evolution deepen understandings of processes of sociotechnical transformation and inform copyright policy debates around technological change?

Outline of the Dissertation

Processes of sociotechnical transformation are a source of growing interest amongst researchers. This thesis explores this topic by tracing the emergence and evolution of HathiTrust. Chapters II and III consist of a review of relevant literatures, focusing first on the copyright implications of mass digitization in effect when the mass digitization project was initially embarked upon in late 2004 before moving on to a review of relevant social science perspectives. Chapter IV discusses the research design and methods used in the empirical study. Chapter V begins the story of HathiTrust by exploring its pre-history in the partnership between the University of Michigan and Google around the mass digitization of Michigan's print collection. Chapter VI describes and explains how that partnership led, in part, to the genesis of HathiTrust. In Chapter VII, I discuss the evolution of HathiTrust from its initial inception as a shared digital repository to a semi-sovereign collectively governed organization. In each of those chapters, the innovative deviance framework is used to highlight particularly salient features of the emergence and evolution of HathiTrust. Chapter VIII focuses on the *Authors Guild v. HathiTrust* litigation and innovative deviance is drawn upon as a referent for the complex judicial interpretations and analyses of the mass digitization project. In the final chapter, Chapter IX, attention turns toward implications for the future, for HathiTrust and for society as a whole. The story of HathiTrust can facilitate deeper understanding of how institutions and organizations shape and are shaped by the interactions among copyright, technological change, and emerging social practices. In addition, the analytic framework which

expands and extends innovative deviance in combination with sensemaking approaches can provide meaningful insights into the complex, often murky, processes of sociotechnical transformation.

Chapter II: Related Legal Literature

This thesis explores processes of sociotechnical transformation through the story of HathiTrust. HathiTrust emerged, in large part, out of the mass digitization project entered into by Google and a number of research institutions, most notably the University of Michigan. That project had (and continues to have) enormous copyright implications. Therefore this chapter, the first of two literature review chapters, focuses on the copyright aspects of technological change and emerging social practices, paying particular attention to the legal posture of mass digitization. The chapter that follows, Chapter III, focuses on related social science literatures referenced in the Introduction.

The review of legal literature is organized into three main sections. The first section explores copyright and technological change, offering a perspective grounded in the history and sensitive to the themes and patterns that have emerged over time and reappear in the current discourse and debates. This section concludes with a brief discussion of large-scale digitization efforts — the precursors of the mass digitization project.

The second section of this chapter offers an in-depth doctrinal analysis of mass digitization. This work is intended to chart the legal landscape, with both its known features and its uncertainties, as it existed at the time the mass digitization project was in contemplation, late 2002-2004. Getting a firm understanding of the copyright implications of mass digitization provides important analytic scaffolding for the empirical work that follows. In order to recognize the features and function of innovative deviance and understand the sensemaking and decision-making processes of those involved, we need a window into copyright law and policy. That said, the doctrinal analysis that follows demonstrates that the law as it existed at the time did not actually provide much guidance on large-scale digitization of in-copyright works. For purposes of copyright law, the legitimacy of such a project was uncertain (as is often the case when technological change enables new forms of creation, interaction, and participation with cultural and intellectual works). Therefore, in addition to highlighting relevant copyright

doctrine and policy, this chapter sets up the social science perspectives and empirical work that follows, particularly with respect to innovative deviance and sensemaking under uncertainty.

The final section of this chapter provides a connective bridge to the social science literature that follows in Chapter III and suggests that, instead of viewing copyright law as most significantly a matter of *property*, we might instead consider it as fundamentally a matter of *social relationships.* Literature related to participatory culture and economic pragmatist perspectives of innovation provide insights into the "copyright and social relationship" perspective and offer a counterbalance to the doctrinal analysis that precedes it.

Copyright and Technological Change

Copyright law is often referred to in legal scholarship as being "a creature of statute," a system of rules, statutes, and social control algorithms designed to incentivize and optimize cultural and intellectual production. It is also part of a broader and more complex sociotechnical system that co-evolves, readjusts, and transforms with the technologies, behaviors, organizations, and institutions it is designed to mediate. Using an historical perspective, this section draws out some key themes or patterns that resonate in more recent controversies around copyright and technological change, particularly with respect to the precursors of the mass digitization project, to illustrate how copyright law shapes and is shaped by technology and social practice.

Copyright law and technological change have always had a closely linked and fairly tumultuous relationship, typically with copyright seemingly struggling to keep pace with the relatively more rapid changes in information and communication technologies. The first copyright-light privileges, royal printing licenses, arose in the fifteenth century in Venice, Italy, after the introduction of the printing press. These privileges were typically limited in duration (e.g. a monopoly set at five years) and in scope (e.g. limited to the printed reproduction of a single work or works).[44] It was not until 1491 that the Venetian Senate granted an author monopoly rights over printing and selling his own work.[45] Fifty years later, in 1545, the Venetian decree of the Council of Ten (which by this time had adopted a more general series of

[44] Patry, W. F. (2000). *Copyright Law and Practice*. The Bureau of National Affairs, Inc., available at http://digital-law-online.info/patry/patry1.html, at pg. 4.
[45] Peter of Ravenna was granted exclusive rights in his work *The Phoenix*. (Patry, 2000:4)

regulations governing printing and distribution) passed a prohibition on publishing an author's work without proof of the author's permission.[46]

Germany also had some early forays into copyright-like protections, including a grant to the widow of Albrecht Dürer of exclusive rights to publish the late artist's works. In 1511, Dürer himself crafted a copyright notice that ranks with the most aggressive of all time, declaring:

> "Hold! You crafty ones, strangers to work, and pilferers of other men's brains. Think not rashly to lay your thievish hands upon my works. Beware! Know you not that I have a grant from the most glorious Emperor Maximillian, that not one throughout the imperial dominion shall be allowed to print or sell fictitious imitations of these engravings? Listen! And bear in mind that if you do so, through spite or through covetousness, not only will your goods be confiscated, but your bodies also placed in mortal danger."

With respect to the development of copyright law in the U.S., the most relevant jurisprudential ancestor is England, the source of our common law tradition and the movable type printing press. The invention of the moveable type printing press in England in 1476 was a watershed moment triggering not only England's first copyright-like privileges, but also the tremendous social, political, and religious reconfigurations that followed.[47] "The ability of printers, via movable type, to produce large numbers of copies relatively quickly and inexpensively led to two important related developments: (1) a potentially large, new market of readers, and (2) the need to protect authors and publishers/booksellers from pirates bent on stealing that new market."[48]

The first copyright-like privileges in England arose in response to a particular technological development: the invention of the printing press. Motivated by concerns around preserving the authorial and attributional integrity of King Henry VII's statements, proclamations, and statutes (and, later, Henry VIII's), in 1504 King Henry VII appointed William Facques as the first royal printer, giving him an exclusive right to print official documents. In 1518, a printing privilege was issued to Richard Pynson, the second royal printer, in the form of a two-year prohibition against others reprinting a sermon of particular significance.[49] Pynson was regarded as, technically and typographically, the best English printer of his generation. Compared to his peers he was also prolific, publishing hundreds of books over the course of his

[46] Patry, 2000:4.
[47] William Caxton established his movable type printing press at Westminster in 1476. (Patry, 2000:5-6)
[48] Patry, 2000:6.
[49] The sermon was the Latin sermon of the dean of St. Paul's Cathedral. (Patry, 2000: 6)

lifetime. As a consequence of his prolificacy, Pynson was hugely influential in terms of the standardization of the English language and the English society more generally.

Pynson can serve as a useful figurehead or unwitting representative of a set of more generalized observations about the sociotechnical reconfigurations brought on by the invention of the printing press. In addition to issues and concerns around the negative potentials of the emerging technology (i.e. the reproduction and disseminations of errors and misrepresentation), recognition grew around the printing press' potential to both stimulate diversity and demand conformity within society. This was patently true in the sense that printing, publishing, and the concomitant increases in literacy facilitated the expression of new ideas, beliefs, and values and enabled the censorship and restraint of expressions. It was also latently true in the sense that printing, and standardized type more specifically, had an almost subliminal influence on diversity and conformity. Eisenstein argues that this technical standardization, somewhat surprisingly, gave rise to an emerging sense of individualism that had not existing in English society prior to the introduction of the press: "The more standardized the type, indeed, the more compelling the sense of an idiosyncratic personal self."[50]

In this way, the first copyright-like privileges can be understood as having developed in response and relation to this range of social potentials embedded in printing and publication technologies. As Pynson's influence spread, so did copyright-like privileges. By 1533, the first accusations of piracy surfaced and "an act was passed that ingeniously worked to the benefit of both English printers and the Crown by banning the importation of foreign books, and thereby ideas, such as those of Martin Luther."[51] By 1538, Henry VIII instigated a form of prepublication censorship, decreeing that all new books needed preapproval by the Privy Council before they could be published. In an effort to stem the flow of seditious, heretical, and/or disruptive texts, the Licensing Act, a sweeping system of pre-publication censorship, was passed into law by the British Parliament in 1557. This Act gave the Stationer's Company a perpetual monopoly over all printing and publication in England; Catholicism was a primary target of suppression.[52] But as diversity and conformity are interdependent features of society, the Licensing Act prompted its own sociotechnical reconfigurations. In particular, the Act seemed to

[50] Eisenstein, E. L. (2005). *The printing revolution in early modern Europe.* Cambridge University Press, at p.56.
[51] Patry, 2000:6.
[52] Patry, 2000:7.

cultivate a growing sense of discontent regarding the use of copyright-like laws to restrain the creation and communication of cultural and intellectual works via emerging technologies.

One of the earliest and most famous instances of pushback against the Licensing Act was proffered by John Milton in his *Aeropagitica: A speech of Mr. John Milton for the liberty of unlicensed printing to the Parliament of England* (1644).[53] In this speech, Milton offered an impassioned criticism of the licensing requirements and censorship practices of the Stationer's Company, arguing: "Give me the liberty to know, to utter, and to argue freely according to conscience, above all liberties."[54] The progress of society could be measured, in Milton's view, by the freedoms of its citizenry to create, disseminate, and use cultural and intellectual works. As vessels of free expression and therefore vital to the welfare of society, books, Milton argued, were "the precious lifeblood of a master spirit embalmed and treasured up on purpose to a life beyond life."[55]

Not only do books serve an important cultural or aesthetic function, as mirrors of nature and human society, Milton recognized that they also serve as potentially important sites of contestation whereby social values, beliefs, and norms can be promoted, traded, modified, and/or destroyed:

> "Though all the winds of doctrine were let loose to play on the earth, so Truth be in the field, we do injuriously, by licensing and prohibiting, to misdoubt her strength. Let her and Falsehood grapple; who ever knew Truth put to the worse, in a free and open encounter?"[56]

Laws, in Milton's view, should not restrict the progress of society by censoring and restraining the creation and communication of multiple and varied viewpoints and expression. On this basis, he opposed the Licensing Act.

The flipside of Milton's critique was advanced a few years later by Thomas Hobbes. Hobbes did not share Milton's optimism about the unfettered flow of information and social progress that would inevitably result from a deregulated printing and publishing industry.

---

[53]Milton, J. (1976). 1644.". *Areopagitica: A Speech of Mr. John Milton for the Liberty of Unlicensed Printing, to the Parliament of England*.
[54] Milton, J. (1976). 1644.". *Areopagitica: A Speech of Mr. John Milton for the Liberty of Unlicensed Printing, to the Parliament of England*.
[55] Milton, J. (1976). 1644.". *Areopagitica: A Speech of Mr. John Milton for the Liberty of Unlicensed Printing, to the Parliament of England*.
[56] Milton, J. (1976). 1644.". *Areopagitica: A Speech of Mr. John Milton for the Liberty of Unlicensed Printing, to the Parliament of England*.

Instead, Hobbes offered a compelling rationale for supporting copyright-like laws (and other legal protections).

In *Leviathan* (1651), Hobbes famously argued that individuals' mutual consent to abide by laws, rules, and other restraints is necessary "to prevent the certain collapse of humanity into an ongoing war between people with conflicting self-interests."[57] With respect to cultural and intellectual works in particular, Hobbes arguments would suggest that, absent copyright-like laws, "there is no place for industry because the fruit thereof is uncertain: and consequently no Culture of the Earth; … no account of Time; no Arts; no Letters; no Society."[58] It follows to reasons that Hobbes might have viewed intellectual property laws as a necessary precondition of cultural and intellectual production.

When the Stationer's Company's monopoly over printing ultimately lapsed in 1694, the copyright regime that grew to replace it encapsulated elements of both Milton's *Aeropagitica* (adopting the goal of learning and social progress) and Hobbes' *Leviathan* (adopting the means of legal incentives to stimulate and protect the creation of cultural and intellectual works). The balance between the two perspectives was codified in the Statute of Anne, the first copyright act, and continues to resonate today in the basic means-end formulation of the copyright system. [59]

The Statute of Anne, enacted in 1710, was a "watershed event in Anglo-American copyright history."[60] Construed in terms of the sociotechnical reconfigurations it signaled, the Statute of Anne diverged from the previous Licensing Act in four key ways. First, it vested rights in the author of new works, rather than publishers, thereby shifting focus from *dissemination* of existing works to the *creation* of new works. Second, while the Licensing Act authorized perpetual privileges, the Statute of Anne *limited the term* of protection to fourteen years, with the possibility of one fourteen-year renewal term. Third, as a precondition of receiving copyright rights, works had to be registered and deposited in an official repository. The combined effect of the limited term and depository requirement was the creation of a public domain consisting of works that either had not met the formalities requirements or whose

[57] Hobbes, T. (1969). *Leviathan, 1651*. Scolar Press.
[58] Hobbes, T. (1969). *Leviathan, 1651*. Scolar Press.
[59] The full official title of the Act is "An Act for the Encouragement of Learning, by Vesting the Copies of Printed Books in the Authors or Purchasers of such Copies, during the Times therein mentioned." 8 Anne, c. 19 (1710). "The Statute of Anne," April 10, 1710. Available at http://avalon.law.yale.edu/18th_century/anne_1710.asp
[60] Joyce, C., & Patterson, L. (2003). Copyright in 1791: An essay concerning the founders' view of the copyright power granted to Congress in Article I, Section 8, Clause 8 of the US Constitution. *Emory Law Journal*, *52*(909), At p.916.

copyright had lapsed. Finally, the Statute of Anne abandoned the practice of pre-publication censorship that Milton had objected to and which had been central to the Stationer's Company's licensing regime.

In addition, and more generally, the Statute of Anne established the basic means-end formulation (largely derived from the just-discussed concepts articulated by Milton and Hobbes) which continues to operate, over three hundred years later, as the essential structure copyright law in England and other common law jurisdictions, including the United States. The formulation, plainly evident from its official title — *An Act for the Encouragement of Learning, by Vesting the Copies of Printed Books in the Authors or Purchasers of such Copies, during the Times therein mentioned* — describes copyright law as "not an inevitable, divine, or natural right that confers on authors the absolute ownership of their creations" but rather as a system designed to "stimulate activity and progress in the arts for the intellectual enrichment of the public" by granting authors a limited monopoly over their original creations.[61]

The historical and philosophical foundations of the current system of copyright law reveal a few key observations. First, copyright is, and always has been, intricately tied to technological change. Second, copyright law's means-end formulation, which continues to serve as the essential structural foundation of the current common law copyright regime, has deep historical and philosophical roots that can be traced back to Milton's theories on freedom of expression and Hobbes' theories about the important of property laws. In addition, its goal — achieving social progress through enrichment of the cultural and intellectual record — operates in tension with its means — state-sanctioned monopoly rights in the exploitation of the fruits of one's creative and intellectual talents and labors. Third, the intersection between copyright, technology, and social institutions are interdependent, dynamic, and function as locales of contention and reconfiguration in sociotechnical systems.

During the colonial period, U.S. copyright law mirrored English law in several key respects: copyright was recognized as common law, the grant of exclusive rights to authors as an incentive to create was the accepted justification and, due to considerations of practical application, it was agree that the law must be federal. By the time the U.S. Constitution (1776) and the Copyright Act (1790) were drafted, however, the perspective on copyright had changed in a few important ways. For example, in England, copyright protection emerged out of a

[61] Leval, P. N. (1990). Toward a fair use standard. *Harvard Law Review*, *103*(5), 1105-1136, p. 1107.

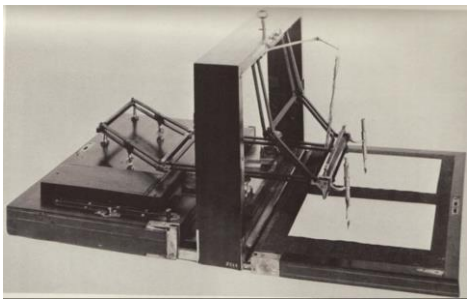
Lockean natural law approach; authors are entitled to state protection of the fruits of their labors as a matter of moral right. By contrast, property rights in the U.S., including copyright rights, are a creature of statute; authors are not morally entitled to reap the benefits of their ingenuity and labor but rather benefit only to the extent the state deems appropriate and enforces. In addition, unlike England, which empowered government officials to adjust down exorbitant pricing, copyright scholar William Patry notes that the U.S. Copyright Act relied solely on the marketplace to determine fair prices.[62]

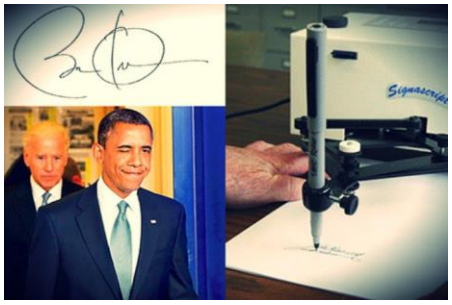
The Copyright Act underwent two major revisions, first in 1909 and then second in 1976. The revisions were spurred, in large part, by sociotechnical changes. New technologies for creating (or fixing), reproducing, and distributing cultural and intellectual works prompted specific revisions to the Act. In addition, the industrialization of the means of production created powerful lobbies which persuaded Congress to enact favorable reforms such as copyright term extensions, changes in formalities requirements, dissolutions in the distinctions between personal and public uses, and industry-specific protections including provisions targeting broadcast and cable networks, and libraries and archives. In the next subsection, relevant legal principles and doctrines will be discussed, but before moving on I will spend a moment tracing some of the key transformations in technical and social aspects of reproduction and distribution of copyrighted works, drawing attention to some of the key moments between the invention of the movable type printing press and library digitization at scale.

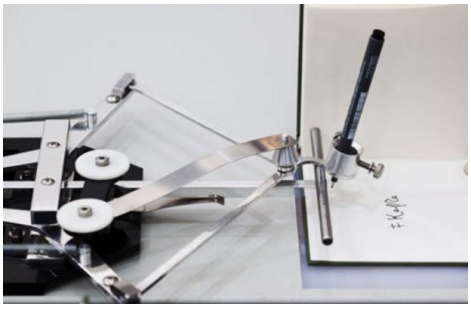
Notwithstanding the advent of the printing press, the vast majority of reproductions continued to be accomplished on a small, as-needed, scale. Copyists, scribes, and scriveners were employed to copy works, in part or in their entirely, by hand. Early reproduction technologies continued to be oriented around handwritten rather than typed works. For example, in the early nineteenth century, polygraph machines enabled instantaneous duplication of a handwritten document by attaching a second pen to the primary one held by the writer (Fig. 3). By the end of the century, Edison's pen enabled non-contemporaneous duplication of handwritten documents through what was essentially a stencil-creating pen (Fig. 4). The Robot Pen or Autopen, created in the 1930s, recorded handwriting, such as a signature, on a storage device for later reproduction (Fig. 5). Although these handwriting reproduction devices have largely been usurped by other reproduction technologies, autopens continue to be used, albeit

[62] Patry, 2000: 30.

somewhat controversially, by government officials for state business; in 2005, the U.S. Justice Department issued an advisory opinion upholding the right of the President to sign bills by autopen[63] and, in 2011, President Obama used an autopen to sign legislation extending the Patriot Act[64] (fig. 6).







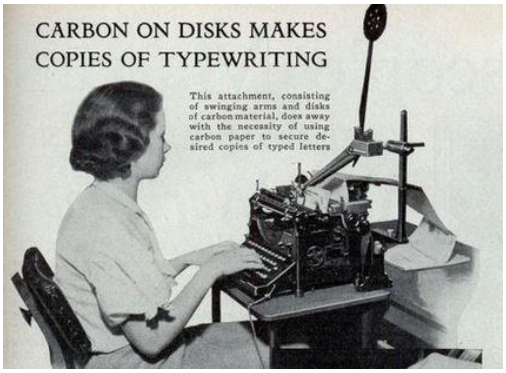
Clockwise from top left: Figure 3. Polygraph Machine; Figure 4. Edison's Electric Pen; Figure 5. Autopen; Figure 6. President Obama signs bill into law using Autopen

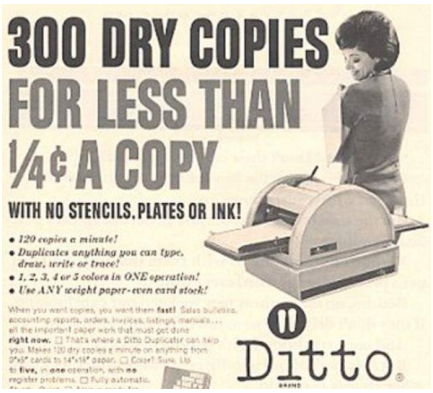
A distinct set of issues of arguably of greater relevance to the topic of this thesis are raised by duplication technologies designed to copy not unique, one-of-a-kind signatures and seals, but widely published materials such as books, pamphlets, and articles. The earliest copiers, called letter copying presses, where invented in the late eighteenth and underwent refinements through the late nineteenth century and essentially consisted of a moistened tissue paper being inserted between an existing printed pages and a new page that, when pressed, would transfer or imprint the text to the new sheet (Fig. 7). Other technologies such as carbon paper (Fig. 8), mimeographs (Fig. 9), and ditto machines (Fig. 10) proliferated during the twentieth century. Through the work of companies such as Xerox and others, photocopiers grew

[63] Nielson, Howard C., Jr., (2005). "Whether The President May Sign a Bill by Directing That His Signature Be Affixed To It," Report of the United States Department of Justice, available at http://search.justice.gov/search?query=whether+the+president+may+sign+a+bill+directing&op=Search&affiliate=justice .

[64] Shear, Michael D., (2011). "Making Legislative History, With Nod from Obama and Stroke of an Autopen," The New York Times, May 28, 2011, available at http://www.nytimes.com/2011/05/28/us/politics/28sign.html?_r=0

to dominate much of the twentieth century and continue to be a staple in most offices, notwithstanding the rise in digital document creation, scanning, and delivery (Fig. 11).





Clockwise from top left: Figure 7. Letter Copying Press;
Figure 8. Carbon Paper; Figure 9. Mimeograph; Figure 10. Ditto Machine

Figure 11. Xerox Machine

The first commercially available scanner, marketed in 1978, was initially embraced by the business and government sectors and, in particular, organizations that were paper-laden and had fairly predictable workflows.[65] Many of these early efforts, particularly those in the governmental sector, were primarily driven by a desire to simplify and streamline *print distribution* channels. For example, in one of the earliest large-scale digitization efforts, the Library of Congress ("LOC") began scanning *Newsweek*, *Time*, and other periodicals onto 14-inch optical platters. The scans were used to generate additional print copies that were then distributed to congressional staffers for research and information purposes. Digitization was not much more than an intermediary step in the process of improved print distribution.

It was not long before early digitizers began contemplating other uses of converted materials. For example, in the early 1980s the National Library of Medicine ("NLM"), which had been digitizing medical journals and journal articles in an effort to streamline distribution, began envisioning the potential of digitization to enable *digital document delivery.* Removing the step of printing the digital scan would certainly seem to remove friction in the interlibrary loan process but, at this early stage, digitization was still very much conceptualized and understood in analog terms, as a more convenient, efficient, perhaps less expensive version of print material.

[65] *Digital Pioneers* interview with Paul Conway, dated May 25th, 2010. Available at http://digitalpioneers.library.du.edu/paulConway/transcript.html.

A few years later, around 1985, the National Archives provided $2 million in funding for the Optical Digital Imaging Storage System ("ODISS") project. This was one of the first concerted forays into exploring technical standards for digitization, suggesting that perspectives had begun to shift from seeing digitization as a transitional step in an essentially analog process to an end point with independent value. ODISS was evidence that digitization could offer a new, viable way to storage and archive content. Using equipment that was state of the art at the time — Honeywell scanners capable of producing a 200 dpi bi-tonal image — ODISS was characterized as by one of the pioneers of digitization as "an experiment to see how bad digital images could be and still be acceptable to users."[66]

By the late 1980s, interest in the potential of digitization to improve access and facilitate preservation of print collections was widespread among national and academic libraries and archives. These early efforts were characteristically project-specific, goal-oriented, and relatively short-lived. Digitization was incredibly costly, both in terms of human resources and technological resources. Compounding things further, digitization projects seemed to be waging a constant (often losing) war against obsolescence. The rate of progress and development in digital technologies fast outpaced the human processes of acquiring funding resources, curating, collecting and transporting the content to be digitized, scanning and turning pages, and so forth. Unfortunately, many of these early digitization efforts failed to reach their goals and, even worse, many of the artefacts from their early efforts were disposed of, leaving few physical traces and diminishing intangible traces in the recollections of participants. [67]

For example, toward the end of the ODISS project, 14-inch optical platters were no longer standard. The technology had shifted to CD-ROMs and CD-writable disks. One of the central figures in the ODISS project reflected:

> "We no longer had equipment to make use of the large optical scans but then I figured out that the Bush Presidential library in Texas had a bridging technology that we could use to get the images off of our disks using their system. I figured out a way to do that and the whole thing was going to cost about $125,000 more, but in the end, what do you have but a bunch of 200 dpi scanned images? The conclusion was that it wasn't worth it and so all of the scans and indexing from that $2 million project were thrown away."[68]

[66] Interview with Peter Hirtle, dated September 30, 2014. Transcript on file with author.
[67] Interview with Peter Hirtle, dated September 30, 2014. Transcript on file with author.
[68] Interview with Peter Hirtle, dated September 30, 2014. Transcript on file with author.

Another failed project was the scan-to-print-to-microfilm project (Paul Conway). Those scans eventually became worthless due to technical incompatibilities that developed.[69] Early digitization efforts were prone to fail and, unfortunately for researchers studying digitization, much of the historical record of these early forays disappeared, was thrown away, or became unreachable as the technology obsolesced. As one of my participants remarked, "I've been involved with more failed digitization projects than probably anyone in history," which is why it is even more important to build a tangible record of this history that, otherwise, lives in the recollections of early digitization pioneers.

One of the ways organizations and institutions worked to combat or stem the seemingly inevitable slip into obsolescence was through the formation of partnerships. Cooperative action was a way to both spread out the immense expense of digitization and also generate greater accountability for projects' maintenance and success. In 1987, for example, the National Agriculture Library held a conference on the applications of optical scanning in libraries and participants there discussed a joint project between Syracuse and the Kellogg School to digitize continuing education materials for adults.

Most of the early pioneers in retrospective digitization point to the Making of America project, begun in 1995, as one of the first significant and successful collaborative large scale digitization efforts. Funded originally by the Mellon Foundation, Making of America was a joint effort of Cornell University and the University of Michigan and sought to fulfil three basic goals. First, the project sought to preserve and make accessible through digital technology a significant body of primary sources related to the development of the United States' infrastructure, focusing on documenting American social history from the antebellum period through reconstruction. At Cornell University, 109 monographs (267 volumes) and 22 journals (955 volumes) were digitized.[70] The University of Michigan digitized approximately 1,600 books and 10 journals.[71] By 2007, the year the website was last updated, the collection totalled approximately 10,000 books and 50,000 journal articles.

In addition to making content digitally available, Making of America was founded upon a preservation goal. Many of the scanned materials were brittle and thus digitizing this content

[69] Interview with Peter Hirtle, dated September 30, 2014. Transcript on file with author.
[70] Cornell University Library, *Making of America*, website available at http://digital.library.cornell.edu/m/moa/about.html
[71] University of Michigan Library, *Making of America*, website available at http://quod.lib.umich.edu/m/moagrp/about.html

enabled libraries to print them onto acid-free paper, bind them, and put them on the shelf for patrons to use.

The third goal of the Making of America project was to obtain and engage the broader research and national institutional community to "develop common protocols and consensus for the selection, conversion, storage, retrieval, and use of digitized materials on a large distributed scale."[72] There was a tension in these early projects between manufacturers of the technical devices and the user community, which consisted primary of government and academic libraries and archives. The technological developers were motivated to develop newer and better technological innovations and did not necessarily see a rationale for providing continued technical support or interoperability with previous digitization technologies. The institutions, on the other hand, seemed to be throwing money away on digitization projects that would invariably grow obsolete and often become wholly inaccessible on a very quick timeline. The ODISS project, discussed below, provides one such example.

Some of the pioneers involved in these early projects viewed the establishment of norms, technical standards and protocols, and techniques around digitization as a much-needed source of stability in the rapidly changing digital environment. Cornell University, in particular, was one of the early leaders on this, benefitting from the experience and expertise of individuals that had worked on the ODISS project. Through its numerous early digitization efforts and experiments, Cornell University became a leader in developing standards that were subsequently incorporated into the federal digitization guidelines standards for preservation set forth by the National Archives and Records Administration ("NARA") which, in turn, became the basis for the federal digitization guideline standards.[73] But, the tension between diversity and standardization continued to play out, despite NARA's ratification.

By the turn of the millennium, momentum was building around large-scale digitization. In addition to the Making of America project, Raj Reddy and colleagues at Carnegie Mellon University embarked on the "Universal Library Project," a project aimed at creating a free-to-read, searchable collection of one million books, primarily in the English language, available to

[72] Cornell University Library, *Making of America*, "About the Project," website available at http://ebooks.library.cornell.edu/m/moa/about.html.

[73] Interview with Peter Hirtle, dated September 30, 2014. Transcript on file with author.

everyone in the world over the Internet.[74] Brewster Kahle, founder of the Internet Archive, was also immensely influential as an early leader in digital archives and universal access efforts, followed a few years later by an active and expansive digitization program, with scanning centers and Internet Archive "nodes" dispersed across geographies and supported through diverse institutional partnerships.

When Google embarked on its digitization project, what would ultimately come to be recognized as the world's first *mass* digitization project, it added several layers of influence (or disruption) into the digitization universe. For example, it chose not to follow the preservation standards, because doing so would have required more time and more storage capacity and would have ultimately slowed the project down. As one of my participants who had been active in the standards setting process described: "The Google Books project comes along and says 'We're going to ignore all of this.' That's fine. It would have been nicer if they had done everything to our preservation standards but something is better than nothing, isn't it?"

Google's decision to make satisficing scans may have been motivated by efficiency and expediency, but it also has a number of second- and third-order effects. Standards can emerge through a deliberative process, as illustrated by Cornell University's work with NARA, but they can also emerge as a by-product of design choices made by first-movers who gain rapid dominance over a new market, industry, or technology. One of my study participants analogized this process to the impact of a dominating feature of the natural environment: "In Seattle, people say that Mount Rainier makes its own weather and I think the same can be said of Google's mass digitization project; it changes the existing knowledge environment and enables new standards to emerge."

In addition, the Google project raised significant copyright concerns in comparison to its predecessors (which primarily dealt in public domain materials). As one noteworthy copyright scholar noted at the time, Google's mass digitization project …

> "… strikes at the very heart of the copyright system and reveals that we tend to rely on the rickety structure of fair use to support too many essential public values. Google's Library Project threatens to unravel everything that is good and stable about the copyright system. It injects more uncertainty and panic into a system that is already in disequilibrium."[75]

---

[74] Reddy, R. and StClair, G. (2001). "The Million Book Digital Library Project," Dec. 1, 2001, available at http://www.rr.cs.cmu.edu/mbdl.htm

[75] Vaidhyanathan, S. (2006). Googlization of Everything and the Future of Copyright, The. *UC DAViS l. reV.*, *40*, 1207.

While the potential copyright implications in existence contemporaneously with the start of Google's mass digitization project will be discussed in the next subsection, it is worth signalling the non-obvious, but potentially significant, linkages between those early decisions about scan quality standards and available legal rationale for the digitization. For example, if Google intended to defend its actions based upon an archival preservation argument, its decision to scan at a quality level significantly lower than NARA's standards could have had a bearing on the success of its defense. Decisions about technical standards have potentially important ramifications for what the intended uses of the scans might be and, as a copyright matter, what potential legal rationale might be successfully advanced for doing the copying in the first place.

Doctrinal Analysis of Mass Digitization

This section walks through the various elements of copyright law implicated by the mass digitization of in-copyright works to illustrate both the legal complexity and the potential for confusion surrounding the practice. The focus is on the law as it existed at the time the University of Michigan and Google entered into an agreement to digitize UM Library's print collection in the fall of 2004.[76] The reasons for adopting this bright-line are twofold. First, because this research is interested in sensemaking and decision-making in the lead up to, and at the outset of, the mass digitization project, the contemporaneous legal context is crucial to understanding. Second, because this research is also interested in how sensemaking and decision-making evolved over time, as part of a larger and more complex process of sociotechnical transformation sparked, in part, by the mass digitization project, it is helpful to have an established starting point from which we can begin to trace and disentangle the copyright, technology, and social practice "strands." Pausing the doctrinal analysis in 2004 allows the findings relating to sensemaking, decision-making, and processes of transformation around mass digitization and HathiTrust to emerge through the empirical study. In addition, because retrospective and historical accounts may be prone to normalization and other forms of hindsight bias, the following detailed accounting of copyright law as it existed in 2004 also

[76] Serious discussions for the project began earlier, in late 2002. But as the discussion demonstrates, there were no key relevant copyright decisions during the contemplation period (late 2002 – late 2004) and therefore this section uses late 2004 as the critical date for purposes of determining the copyright law as it existed when the project was embarked upon.

provides a way to anchor participants' recollections about the risk and uncertainty of the undertaking against a relatively more objective measure.

The following review begins with a brief discussion of the goals of copyright, the "exclusive rights" in §106, and remedies for infringement including damages and profits in §504 and state sovereign immunity in §511. The focus then shifts to potentially relevant limitations on the exclusive rights including: fair use in §107; library and archives exemptions in §108; computer programs exemption in §117; exemption for blind and disabled persons in §121; and relevant provisions of the Digital Millennium Copyright Act, §§1201-1205. Because the statute provides a snapshot of the law rather than a living, breathing interpretation of it, case law (up through 2004) will also be cited, particularly with respect to fair and/or transformative uses and the sovereign immunity of universities.

Goals and Purposes of Copyright Law

While the previous subsection explored some of the historical and philosophical foundations of copyright law, particularly with respect to technological change, this section focuses more narrowly on its representations in positive law. Article I, Section 8, clause 8 of the U.S. Constitution gives Congress the power "To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries."[77] This section also provides the basic means-end formula Congress ought to abide by when crafting the Copyright Act, aiming to should strike a balance between the Miltonesque desire for broad social benefit (through free expression and the ready exchange of cultural and intellectual works) and the Hobbesian desire for private benefit (in the form of personal monopolies and enforceable property rights in the fruits of one's labor). The result is a regime that, at its essence, uses state-created property rights simultaneously as an economic trade regulation and as a fount of cultural and intellectual activity for the benefit of society as a whole.

[77] The framers of the Constitution used the term "science" to refer to knowledge and learning, and used the term "useful Arts" to refer to industry. Although it may be somewhat peculiar to our present-day lay usage of the terms, copyright law (which we typically associate with creative expression) is intended promote the progress of science while patent law (which we typically associate with technological innovation) is intended to promote the progress of the useful arts. See Eldred v. Ashcroft, 537 U.S. 186, 242-243 (U.S. 2003) (Breyer, J. dissenting) (citing Walterscheid, E. C. (2002). *The nature of the intellectual property clause: A study in historical perspective*. William S. Hein & Co., Inc..

While this means-end formulation may seem inherently fraught given the strident debates and never-ending battles between rightsholders and public interest advocates, scholar William Patry explains that, when this clause was written into the Constitution, no such tension existed:

> "[T]he public interest *fully coincides* with the interest of authors having exclusive rights in their works. The source of this harmony between public and private interests is not difficult to discern once we strip away our present-day, consumer-oriented perspective: in place of government control, the Founding Fathers believed private property, including intellectual property, was the best way to ensure the triumph of democracy over the tyranny of the aristocracy. As former Register of Copyrights Abraham Kaminstein observed:
>
> *The basic purpose of copyright is the public interest, to make sure that the wellsprings of creation do not dry up through lack of incentive, and to provide an alternative to the evils of an authorship dependent upon private or public patronage. As the founders of this country were wise enough to see, the most important elements of any civilization include its independent creators – its authors, composers and artists – who create as a matter of personal initiative and spontaneous expression rather than as a result of patronage or subsidy. A strong, practical copyright is the only assurance we have that this creative activity will continue."*[78]

By giving creators (referred to collectively as "authors" in copyright parlance) a limited monopoly over their creations (referred to as "works"), copyright permits them to exercise their creativity independent of governmental control or systems of patronage, and control and receive compensation for many aspects of the communication and use of their works. As the Supreme Court explained:

> "The economic philosophy behind the clause … is the conviction that encouragement of individual effort by personal gain is the best way to advance the public welfare through the talents of authors …."

Therefore, the goal or purpose of copyright is a fundamentally public one as it seeks to promote the progress of society through knowledge and learning, and the means set out to accomplish this goal are property rights-based — copyright incentivizes the creation and

[78] Patry, 2000: 24, citing the testimony of Register of Copyrights Abraham Kaminstein, *Copyright Law Revision: Hearings Before the Subcomm. on Patents, Trademarks, and Copyrights of the Comm. on the Judiciary on S. 1006*, 89th Cong., 1st Sess.65 (1965). *See also* Register Kaminstein's further remarks in Copyright Law Revision Part 6: Supplementary Report on the Register of Copyrights on the General Revision of the U.S. Copyright Law: 1965 Revision Bill, 89th Cong., 1st Sess.xiv-xv (House Comm. Print 1965).

communication of the fruits of authors' creative and intellectual labor by establishing a marketable right to the use of one's own expression.[79]

> "The copyright law celebrates the profit motive, recognizing that the incentive to profit from exploitation of copyrights will redound to the public benefit by resulting in the proliferation of knowledge."[80]

Despite the founders' beliefs that the "public good fully coincides … with the claims of individuals,"[81] as the previous section described, tensions have emerged as to whether or not (and the extent to which) copyright laws fairly balance the public's interest and rightsholder interests. This has become particularly true as revisions to the Act have repeatedly expanded the scope of the exclusive rights and extended the duration of copyright protection (which now stands as seventy years beyond the life of the author for published works and ninety-five years from publication for works-made-for-hire). Furthermore, broadening the protections for authors comes at a time when significant changes in information and communication technologies have fundamentally altered the ways in which works can be created, preserved, disseminated, used, modified, remixed, and so forth. So while the copyright regime seems increasingly focused on the expansion and enforcement of rightsholder protections, technologies continue to emerge that enable new forms of participation with protected works in furtherance of copyright's essential goals.

With the purpose of copyright laid out, we can now turn to relevant sections of the Copyright Act.

Exclusive Rights - 17 U.S.C. §106

If intellectual property laws add the "fuel of interest to the fire of genius,"[82] as Abraham Lincoln declared before the Springfield Library Association in 1860, the exclusive rights may be the primary source of combustion. Copyright protection in the U.S. subsists in original works of authorship that are fixed in a tangible medium of expression from which they can be perceived,

---

[79] Harper & Row Publishers, Inc. v. National Enters., 471 U.S. 539, 558 (1985).
[80] American Geophysical Union v. Texaco, Inc. 882 F. Supp. 1, 27 (S.D.N.Y. 1992) , *aff'd,* 60 F.3d 913 (2d Cir. 1994).
[81] Madison, J. (1788). THE FEDERALIST NO. 43.
[82] Lincoln, A., & Miller, M. M. (1908). *The wisdom of Abraham Lincoln*. New York: A. Wessels company. "Lecture before the Springfield, IL. Library Association: Discoveries, Inventions and Improvements," (Feb. 22, 1860), at pg. 104. Lincoln's statement was made in reference to the Patent Law system but is also applicable to the Copyright Law system.

reproduced, or otherwise communicated.[83] These works can be literary, musical (compositions and sound recordings), dramatic, choreographic, pictorial, graphic or sculptural, audiovisual, or architectural. In §102(b), ideas, procedures, processes, and systems, methods of operation, concepts, principles, and discoveries are expressly excluded from Copyright protection.[84]

Authors of copyrighted works are entitled to certain "exclusive rights" specified in the Act. The relevant rights for purposes of this research include the right to do and authorize any of the following:

(1) to reproduce the copyrighted work in copies;

(2) to prepare derivative works based upon the copyrighted work;

(3) to distribute copies of the copyrighted work to the public by sale or other transfer of ownership, or by rental, lease, or lending;

(4) to perform the copyrighted literary work publicly; and

(5) to display the copyrighted literary work publicly.[85]

Copyright rights vest automatically the moment a work is created, regardless of whether or not it is made public. Formalities such as affixing a "©" to the work, registering the work with the Copyright Office, and depositing copies of the work with the Library of Congress are not a prerequisite to protection, although these measures may afford rightsholders additional benefits or remedies, such as statutory damages, in some cases. The lack of a comprehensive recording process makes tracking ownership rights a challenge, if not an outright impossibility. In addition, as property lawyers and scholars are keen to point out, property rights really consist of a "bundle of rights" meaning that each of the enumerated rights may be subdivided and partitioned indefinitely, and each piece of the exclusive right may be owned and enforced

[83] 17 USC §102(a).
[84] 17 USC §102(b).
[85] Right to distribution, performance, and display are conditioned on "public" communication. In §101, the Act tells us "publicly" means a display or performance that is "at a place open to the public or at any place where a substantial number of persons outside of a normal circle of family and its social acquaintances is gathered" or that is transmitted or otherwise communicated to such a place or "to the public, by means of any device or process, whether the members of the public capable of receiving the performance or display receive it in the same place or in separate places and at the same time or at different times." The accompanying House Report further explains that "one of the principle purposes of the definition was to make clear that performances (and/or displays) in semipublic places such as … schools are public performances (and/or displays) subject to copyright control" and that the transmission guidelines are "broad enough to include all conceivable forms and combinations of wired or wireless communications media" and apply with equal force to co-located and geographically dispersed transmissions as well as transmissions that are received contemporaneously and/or at staggered times. House Report no. 94-1476 (1990, 1995, 1999, 2002).

separately.[86] The result is that, over the life of a copyright, the bundle of rights may become dispersed and difficult to trace making permission-seeking for uses implicating the exclusive rights untenable.

Compounding these concerns further is the fact that copyright infringement is a strict liability tort, meaning that the subjective intent of the actor is irrelevant to the question of culpability.[87] There is no such thing as "innocent," "excusable," or "justifiable" infringement. If a person interferes with any of the rightsholder's exclusive rights, and their activity is not subject to one of the enumerated limitations or exceptions discussed below, they may be found liable for copyright infringement and subject to the penalties outlined in Chapter V of the Act.

Statutory Remedies & Questions of Liability

This section discusses remedies for infringement and potential relevant limitations on liability including §504 — Damages and Profits, and §511 — State Sovereign Immunity.

DAMAGES AND PROFITS - §504

Remedies for copyright infringement can include monetary damages,[88] temporary or permanent injunction,[89] impounding and destruction of infringing copies,[90] and costs and attorney's fees.[91] Instead of recouping actual damages and profits, prevailing plaintiffs who timely registered their work may elect statutory damages that range from $750 - $30,000 per instance of infringement, and up to $150,000 per instance of willful infringement.[92] Willful infringement is not defined by the Act, but courts have typically treated the willfulness determination as consisting of two prongs: 1) the defendant engaged in the (infringing) acts, and 2) the defendant knew or should have known that the acts were infringing.[93] The burden is on the plaintiff to prove willfulness by a preponderance of the evidence. In addition, although they

[86] House Report no. 94-1476 (1990, 1995, 1999, 2002).
[87] But see Goold, P. R. (2015). Fair Use: Why Copyright Infringement Is Not a Strict Liability Tort. *Berkeley Technology Law Journal*, arguing that the fair use defense suggests that culpability for copyright infringement actually based on a negligence standard.
[88] 17 USC §504.
[89] 17 USC §502.
[90] 17 USC §503.
[91] 17 USC §505. A prevailing plaintiff may only recover costs and fees if the copyrighted work was timely registered.
[92] 17 USC §504. Statutory damages range from $750 - $30,000 per infringement. Damages for "willful" infringements can run up to $150,000 per infringement. (17 USC §504(c)).
[93] Halo Electronics, Inc. v. Pulse Electronics, Inc., 769 F.3d 1371 (Fed. Cir. 2014), at 1381-1382.

are rarely imposed, criminal penalties including imprisonment may be ordered in some cases of copyright infringement.[94]

Plugging these potential remedies into the mass digitization context we can speculate that the "worst case scenario" in terms of possible copyright infringement liability is extreme. Assume, for example, that the University of Michigan's library contains five million in-copyright works and the digitization project meets the threshold for willfulness, the potential liability for making a single copy is seven hundred and fifty billion dollars. Of course, Google would need its own copy (double that figure) and Michigan would need at least one back-up copy (double that figure again). And then if Michigan or Google wanted to actually do anything with the copies, that might trigger additional instances of infringement. This is all to say that, in terms of the worst case scenario, the risks were extreme.

That said, however, §504(c) of the Act includes a potentially significant added source of protection for libraries and archives that may be relevant to the mass digitization project:

> "The court shall remit statutory damages in any case where an infringer *believed and had reasonable grounds for believing* that his or her use of the copyrighted work was a *fair use* under section 107, if the infringer was an employee or agent of a *nonprofit educational institution, library, or archive* acting within the scope of his or her employment who, or such institution, library, or archives itself, which infringed by reproducing the work in copies …." [95]

Thus, if the infringing activity was undertaken by a nonprofit educational institution, library, or archive under the belief — subjectively and reasonably under an objective standard — that the infringing activity qualified as a fair use, it would be insulated from an award of statutory damages. It is worth noting that a belief that its activities were non-infringing under other exemptions (such as §108, §117, and §121 discussed infra) would not be sufficient to trigger §504(c)'s protections. Furthermore, libraries associated with for-profit institutions or universities would not be eligible to avail themselves of this added protection.

STATE SOVEREIGN IMMUNITY - §511

Another potential limit on public institutions' liability for copyright infringement is the doctrine of sovereign immunity. Most scholars agree that the doctrine stems from the U.S. Constitution although there appears to be some disagreement about its specific source within that

[94] 17 USC §506.
[95] 17 USC §504 (c)(2)(i).

document. The dominant understanding is that state sovereign immunity derives from the Eleventh Amendment which provides, in pertinent part:

> "The Judicial power of the United States shall not be construed to extend to any suit in law or equity, commenced or prosecuted against one of the United States by Citizens of another State, or by Citizens or Subjects of any Foreign State."

The express text of the amendment does not mention suits brought against a state by its own citizens. However, over time, the principle has evolved beyond a strict textual reading to encompass suits filed by citizens of the defendant state.[96]

Although state sovereign immunity has been interpreted fairly broadly in recent years, the principle is still undergoing active exploration and thus is not well-settled. As one scholar notes, "the Eleventh Amendment has emerged from relative obscurity to become a major focus of constitutional controversy."[97] While it was relatively ignored through the first half of the twentieth century, court opinions have mentioned it liberally in recent years.[98] In the last two decades the Supreme Court has made several important Eleventh Amendment rulings that reflect a deeply divided court. Majority opinions, as will be discussed momentarily, are frequently coupled with vehement dissents. Many of the decisions are 5-4 splits and the differences of opinion between the majority and dissent often revolve around crucial and foundational aspects of Eleventh Amendment interpretation.

Before distilling some of key aspects of sovereign immunity in the context of public universities' potential liability for copyright infringement, I will take a moment to offer a general observation on the highly contentious and unsettled nature of state sovereign immunity jurisprudence drawing upon the Supreme Court's 1999 decision in *Alden v. Maine*.

In *Alden v. Maine*, the majority held that Congress may not use its Article I powers to abrogate state sovereign immunity to subject unconsenting states to suit in state court, writing:

> "We have … sometimes referred to the States' immunity from suit as 'Eleventh Amendment immunity.' The phrase is convenient shorthand but something of a misnomer, for the sovereign immunity of the States neither derives from, nor is limited by, the terms of the Eleventh Amendment. Rather, as the Constitution's structure, its history, and the authoritative interpretations of this Court make clear, the States' immunity from suit is a fundamental aspect of the sovereignty which

[96] *Hans v. Louisiana*, 134 U.S. 1, 10 S. Ct. 504, 33 L. Ed. 842 (1890).
[97] Meltzer, D. J. (1996). The Seminole Decision and State Sovereign Immunity. *The Supreme Court Review*, *1996*, 1-65:1.
[98] Meltzer, (1996:): "The Amendment was cited in only ten Warren Court decisions (over sixteen terms) [1953-1969], but has been mentioned in 125 decisions in the twenty-seven Terms since."

> the States enjoyed before the ratification of the Constitution, and which they retain today (either literally or by virtue of their admission into the Union upon an equal footing with the other States) except as altered by … certain constitutional Amendments."

Note that under the majority's perspective, States' immunity from suit is so essential to the fundamental structure of the United States that the principle not only predates the Constitution but it does not even require express articulation and ratification in that crucial founding document; it is just *that* obvious.

Writing for the dissent, Justice Souter argued that the Eleventh Amendment and the principle of state sovereign immunity should be interpreted narrowly to limit only the diversity jurisdiction of the federal courts. In Souter's view (which is joined by three other Justices), even if States had broad sovereign immunity at some point they necessarily surrendered it when they ratified the Constitution.[99]

One might reasonably wonder whether the majority and the dissent could be any farther apart in terms of their respective constructions of the role of the Eleventh Amendment and interpretations of the doctrine of state sovereign immunity. *Alden v. Maine* demonstrates that questions of state sovereign immunity are extremely contentious and open to debate. Keeping in mind the stark difference of opinion and interpretation regarding the doctrine of state sovereign immunity, consider the following discussion.

At its essence, the doctrine of state sovereign immunity precludes *unconsenting* states from being sued for monetary damages or equitable relief in federal[100] and state[101] courts. As immunity turns on questions of consent, states can waive their sovereign immunity, in whole or in part,[102] either expressly or constructively through the operation of certain rules of civil procedure.[103] In addition, based on still evolving case law, it appears that the protections of state sovereign immunity extend to some arms of the state including libraries, archives, and universities run by states or their instrumentalities.[104] Sovereign immunity does not, however,

---

[99] *Alden v. Maine, 527 U.S. 706, 119 S. Ct. 2240, 144 L. Ed. 2d 636 (1999).*
[100] *Seminole Tribe of Fla. v. Florida*, 517 U.S. 44, 116 S. Ct. 1114, 134 L. Ed. 2d 252 (1996).
[101] *Alden v. Maine, 527 U.S. 706, 119 S. Ct. 2240, 144 L. Ed. 2d 636 (1999).*
[102] Most states, for example, have waived sovereign immunity for liability based negligence in tort actions but few, if any, have waived immunity for liability from intentional wrongs.
[103] *Lapides v. Board of Regents of Univ. System of Ga.* 535 U.S. 613 (2002).
[104] *Seminole Tribe v. Florida*, 517 U.S. 44, (1996);
Chavez v. Arte Publico Press, 204 F.3d 601 (5th Cir. 2000).

apply to state municipalities and counties[105] although, as a practical matter, if municipal or county officials are sued in federal court in their official capacity and any relief granted would have a significant effect on the state treasury, courts will consider it a "suit against the state" and sovereign immunity would kick in.[106]

Aside from states waiving the protection, another important potential limitation derives from Congressional abrogation of states' immunity. The seminal decision was the Supreme Court's 1996 ruling in *Seminole Tribe v. Florida.* Prior to *Seminole,* state universities and libraries were understood to be subject to damages for copyright infringement.[107] In this case, however, the Court held that Congress cannot use its Article I powers to abrogate state sovereign immunity to subject unconsenting states to suit in federal court. A year later, however, the Court ruled in *City of Boerne v. Flores* that Congress may abrogate state sovereign immunity pursuant to its powers under §5 of the Fourteenth Amendment provided that it gives (1) an express statement of intent and (2) the abrogation is a constitutionally valid exercise of power."[108] These two decisions may have important implications on questions of the potential liability of universities for copyright infringement stemming from mass digitization.

First, with respect to the first factor cited in *City of Boerne,* it is uncontroverted that Congress has attempted to abrogate state sovereign immunity in the copyright context. In §511 of the Copyright Act, Congress provided an express statement of intent to make states and their instrumentalities liable for copyright infringement:

> "Any State, any instrumentality of a State, and any officer or employee of a State or instrumentality of a State acting in his or her official capacity, shall not be immune, under the Eleventh Amendment of the Constitution of the United States or under any other doctrine of sovereign immunity, from suit in federal court … for a violation of any of the exclusive rights of a copyright owner provided by sections 106 through 122 … or for any other violation under this title."

This section further provides that states and their instrumentalities may be subject to any of the standard remedies for infringement including impounding and disposition of infringing copies, actual damages and profits, statutory damages, and costs and fees (§511(2)).

---

[105] *Lincoln County v. Luning,* 133 U.S. 529 (1890); *Ford Motor Co., v. Dept. of Treasury,* 323 U.S. 459 (1945); *Edelman v. Jordan*, 415 U.S. 651 (1974).
[106] *Pennhurst State School & Hospital v. Halderman,* 465 U.S. 89 (1984).
[107] Section 108 Study Group Report (2008). An Independent Report sponsored by The United States Copyright Office and the National Digital Information Infrastructure and Preservation Program of the Library of Congress, March 2008. Available at http://www.section108.gov/docs/Sec108StudyGroupReport.pdf.
[108] *Seminole Tribe of Fla. v. Florida*, 517 U.S. 44 (1996); *City of Boerne v. Flores,* 521 U.S. 507 (1997).

Under a plain reading of the statute, §511 satisfies the first prong of the *Boerne* test because it offers an express statement of intent to abrogate state sovereign immunity with respect to copyright suits. Therefore the inquiry shifts to the second prong: whether or not §511 reflects a constitutionality valid exercise of Congress' power. Perhaps unsurprisingly, this question turns out to be far more contentious.

As previously discussed, Congressional authority to pass copyright laws stems from the powers granted it under Article I of the Constitution. Under this clause, Congress may pass laws:

> "To promote the Progress of Science and useful Arts, by securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries."[109]

While Congress clearly has constitutional authority to pass laws *granting rights* under Article I, *City of Boerne* tells us the key inquiry with respect to sovereign immunity turns on whether or not Article I gives Congress the power to *enforce rights*, particularly with respect to states, as it attempts with to §511 of Act.

In *Seminole,* the Court seems to answer this question with a resounding "no." In that decision, the majority plainly states that Congress may not abrogate state sovereign immunity under its Article I powers. *Seminole* would therefore suggest that, despite its express intent, §511 of the Copyright Act does not actually abrogate states' immunity from copyright suits absent some other valid source of constitutional authority.

In *City of Boerne,* the majority held that Congress may use its enforcement powers under §5 of the Fourteenth Amendment to limit state sovereign immunity provided that certain conditions are satisfied. The relevant provisions of the Fourteenth Amendment provide in pertinent part:

> "Section 1. … No State shall … deprive any person of life, liberty, or property, without due process of law ….
>
> Section 5. The Congress shall have power to enforce, by appropriate legislation, the provisions of this article."

Applying the Fourteenth Amendment to the copyright context, we must ask whether or not copyright infringement by states or their instrumentalities could reasonably constitute the deprivation of property without due process of law. Copyright rights are unquestionably a form

[109] U.S. Const., Art. I, §8, cl. 8.

of (albeit intangible) property, however not all deprivations are legally cognizable. Section 1 tells us that only deprivations that occur "without due process of law" trigger Congress's enforcement powers against the states. In other words, states' sovereign immunity from liability for copyright infringement could only be abrogated if state courts are unwilling or unable to adequately remedy the harm.

A nearly identical question was brought before the Court in *Florida Prepaid Postsecondary Educational Expense Board v. College Savings Bank* in the patent law context.[110] (The Patent Act contains a provision, "PRCA," that is functionally identical to §511 of the Copyright Act.) The majority in *Florida Prepaid* stated the applicable rule for purposes of the Fourteenth Amendment as follows:

> "[O]nly where the State provides no remedy, or only inadequate remedies, to injured patent owners for its infringement of their patents could a deprivation of property without due process result."[111]

Given that both patent and copyright law falls under the exclusive jurisdiction of the federal courts, and state courts are therefore powerless to rule on patent and copyright infringement actions, a reasonable inference might be that state courts would be unable to provide adequate remedies in infringement suits. This was the view held by Justice Stevens in his dissenting opinion (joined by Justices Souter, Ginsburg, and Breyer) in *Florida Prepaid.* On the due process question, Justice Stevens explains:

> "Given the absence of effective state remedies for patent infringement by States and the statutory pre-emption of such state remedies, the [PRCA] was an appropriate exercise of Congress' power under §5 of the Fourteenth Amendment to prevent state deprivations of property without due process of law."

The majority did not align with this conclusion because, in its view, Congress "barely considered the availability of state remedies for patent infringement" based on laws properly within the jurisdiction of state courts such as contract law, tort law, and so forth. In other words, the legislative record did not present persuasive evidence suggesting that Congress adequately considered the possibility of re-tooling patent and copyright infringement cases based on state

---

[110] *Florida Prepaid Postsecondary Ed. Expense Bd. v. College Savings Bank*, 527 U.S. 627, 119 S. Ct. 2199, 144 L. Ed. 2d 575 (1999).

[111] *Florida Prepaid Postsecondary Ed. Expense Bd. v. College Savings Bank*, 527 U.S. 627, 119 S. Ct. 2199, 144 L. Ed. 2d 575 (1999) at 643; citing *Parratt v. Taylor,* 451 U.S. 527, 539-541 (1981); *Hudson v. Palmer,* 468 U.S. 517, 532-533 (1984); *id.*, at 539 (O'connor, J. concurring) ("[i]n challenging a property deprivation, the claimant must either avail himself of the remedies guaranteed by state law or prove that the available remedies are inadequate …. When adequate remedies are provided and followed, no … deprivation of property without due process can result").

law principles. While lack of support in the legislative record is not determinative, in this instance the Court concludes the record was insufficient to warrant abrogating state sovereign immunity:

> Congress must "identify conduct transgressing the Fourteenth Amendment's substantive provisions, and must tailor its legislative scheme to remedying or preventing such conduct" and "[t]here must be a congruence and proportionality between the injury to be prevented or remedied and the means adopted to that end." [112]

In addition, the majority did not find in the legislative record substantial evidence that Congress passed the PCRA in response to "a history of 'widespread and persisting deprivation of constitutional rights' of the sort Congress has faced in enacting proper prophylactic §5 legislation."[113] While the record contained testimony suggesting a fear that infringement of patents and copyrights by states might be on the rise, there was little evidence in the record to demonstrate that these fears were substantiated. In the majority's view, "the record at best offers scant support for Congress' conclusion that States were depriving patent owners of property without due process of law by pleading sovereign immunity in federal-court patent actions."[114] Therefore, it concluded that the provision was neither congruent nor proportional in "in light of the evil presented" and thus was not a constitutionally valid exercise of Congress's §5 powers.

The dissent countered by arguing:

> "It is true that, when considering the [provision], Congress did not review the remedies available in each State for patent infringements and surmise what kind of recovery a plaintiff might obtain in a tort suit in all 50 jurisdictions. But, it is particularly ironic that the Court should view this fact as support for its holding. Given that Congress had long ago pre-empted state jurisdiction over patent infringement cases, it was surely reasonable for Congress to assume that such remedies simply did not exist."

Moreover, the dissent argues that the Court's holding is unsupported by *City of Boerne* and actually conflicts with the Courts reasoning in that case:

> "[T]his Court has never mandated that Congress must find 'widespread and persisting deprivation of constitutional rights' in order to employ its §5 authority. It is not surprising, therefore, that congress did not compile an extensive legislative record analyzing the due process (or lack thereof) that each State might afford for a patent infringement suit retooled as an action in tort. In 1992,

[112] *Florida Prepaid* at 639.
[113] *Florida Prepaid* at 645; quoting *City of Boerne* at 526.
[114] *Florida Prepaid* at 646.

> Congress had no reason to believe it needed to do such a thing; indeed, it should not have to do so today."

The central difference between the majority and dissent in this case might boil down to a difference of opinion around the how Congress's Article I authority under the Copyright Clause to pass laws *granting rights* intersects with its power to pass laws *enforcing rights* against the states. The majority concluded that Article I, in granting Congress the power to pass copyright laws, did not also give Congress jurisdiction over a State that does not consent to be sued for violating those rights. While the contested provision ensured uniformity of the patent system and closed a potential loophole in the uniform federal scheme, which are valid Article I purposes, the provision did not satisfy the purposes of the Due Process clause as required by constitutional law.

In contrast, the dissent viewed the Article I powers as necessarily encompassing both the power to pass copyright laws *and* the power to ensure their enforcement:

> "Article I … calculus is directly relevant to this case because it establishes the constitutionality of the congressional decision to vest exclusive jurisdiction over … infringement cases in the federal courts. That basic decision was unquestionably appropriate. It was equally appropriate for Congress to abrogate state sovereign immunity in … infringement cases in order to close a potential loophole in the uniform federal scheme, which, in undermined, would necessarily decrease the efficacy of the process afforded to [rights] holders."

In addition, Justice Stevens argued that the PRCA, the legislation that expressed Congress's intent to abrogate state sovereign immunity in patent infringement cases, did not alter any substantive rule of state law but "merely effectuates settled federal policy to confine patent infringement litigation to federal judges."[115] Recognizing the inherent injustice of sovereign immunity in this context, the United States waived its immunity from patent infringement suits via a 1910 Act of Congress.[116] And, in the 1973 *Goldstein v. California* case, the Supreme Court said: "When Congress grants an exclusive right or monopoly, its effects are pervasive; no citizen *or State* may escape its reach." (emphasis added) The dissent therefore remains sharply critical of the majority's rationale. After all, why then should states receive an added cloak of protection against liability?

[115] *Florida Prepaid* at 662.
[116] "An Act to provide additional protection for owners of patents of the United States." Ch. 423, 36 Stat. 851.

While the Supreme Court has not ruled precisely on the validity of §511, it is reasonable to assume the holding in *Florida Prepaid* would transfer to the copyright context. Indeed, lower courts in the First, Second, Fourth, Fifth, Sixth, Ninth, and Eleventh Circuits have applied *Florida Prepaid* to the copyright context, concluding that, notwithstanding the express intent of §511 to make states and their instrumentalities amenable to suit in copyright cases, the principle of state sovereign immunity applies.[117] To date, no court has enforced §511 against a state. Therefore, at the time the mass digitization project was in contemplation, it would appear that state sovereign immunity might provide a significant layer of protection for public universities against liability for copyright infringement (although universities could still be enjoined from using the scans, an outcome that would render the digitized copies effectively useless).

Statutory Limits on Exclusive Rights - 17 U.S.C. §§107-122

Aside from the potential complete bar to liability offered by the state sovereign immunity doctrine, copyright holders' rights are limited in several important ways.[118] Rightsholders cannot, for example, interfere with subsequent distribution of copies of their work after the initial sale.[119] Libraries are permitted to make preservation copies under certain conditions specified in the Act.[120] Owners of computer programs are permitted to make a backup copy and/or a copy made as an incidental and necessary step in using the program (i.e. a RAM copy).[121]

This section describes four of the primary potential limitations on authors' exclusive rights that may be applicable to the mass digitization context: fair use (§107), library and archive exception (§108), the "dark archive" exception for computer software (§117), and the exception for uses made in the provision of services to blind and disabled persons (§121).

FAIR USE - §107

For purposes of this research, the most significant potential limitation on copyright holders' exclusive rights is the fair use limitation codified in §107 of the Act.[122] The doctrine of fair use limits an author's exclusive rights by allowing "the public to draw upon copyrighted

[117] *Chavez v. Arte Publico Press*, 204 F.3d 601 (5th Circuit, 2000) (Under *Seminole* and *City of Boerne* §511 is an invalid exercise of Article I legislative power and upholding it as a valid exercise of legislative power pursuant to section 5 of the Fourteenth Amendment could be an impermissible end-run around *Seminole*.) at 604.
[118] 17 USC §107-122.
[119] 17 USC §109.
[120] 17 USC §108.
[121] 17 USC §117.
[122] 17 USC §107.

materials without the permission of the copyright holder in certain circumstances."[123] As the Supreme Court notes, fair use is essential to fulfilling the Act's overriding goal:

> "From the infancy of copyright protection, some opportunity for fair use of copyrighted materials has been thought necessary to fulfill copyright's very purpose, '[t]o promote the Progress of Science and useful Arts …'"[124]

Fair use is an "equitable rule of reason" that asks courts to undertake a sensitive balancing of interests, taking into account relevant facts, legal precedent, and equitable considerations to determine whether or not a secondary use of a protected work is infringing or non-infringing. The precise meaning of fair use has always been someone difficult to pin down in practice. Legislative and administrative guidance has motioned, at various times, toward clarifying the doctrine. For example, in contemplation of its inclusion in the Copyright Act (prior to the last major revision to the Act in 1976), the Register of Copyrights produced a report that provided a non-exhaustive list of core fair use examples:[125]

- Quotation of excerpts in a review or criticism for purposes of illustration or comment;
- Quotation of short passages in a scholarly or technical work, for illustration or clarification of the author's observations;
- Use in a parody of some of the content of the work parodied;
- Summary of an address or article, with brief quotations, in a news report;
- Reproduction by a library of a portion of their works to replace part of a damaged copy;
- Reproduction by a teacher or student of a small part of works to illustrate a lesson;
- Reproduction of a work in legislative or judicial proceedings or reports;
- Incidental and fortuitous reproduction, in a newsreel or broadcast, where the work is located in the scene of an event be reported.

What the Register's list communicates (perhaps implicitly) is that fair use is a fundamentally fact-based, case-by-case, determination. This gives courts a tremendous amount of discretion in terms of the doctrine's application, but also infuses the copyright system with a tremendous amount of ambiguity and uncertainty.

Fair use was given an express statutory recognition for the first time in §107 of the 1976 Act. The specific fair use language adopted in the Act is the result of a process of accretion

---

[123] *Authors Guild, Inc. v. HathiTrust,* 755 F.3d at 95 (2nd Cir. 2014).
[124] *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. at 574 (1994).
[125] Page 24 of Register's 1961 report.

resulting from repeated collisions in the courts between rightsholders and the defendants asserting their practices were in support of the overriding public policy goals of copyright. The Act requires courts to weigh together four nonexclusive factors in assessing whether a particular use is fair:

> (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, including whether it is primarily creative or instructive (which copyright tends to value and seek to foster) or primarily factual (in which the law of fair use recognizes a greater need to disseminate);
>
> (3) the amount and substantiality of the portion used in relation to the copyrighted work as a whole, including whether the secondary use employed no more than was necessary to effectuate any valid purpose under the first factor; and
>
> (4) the effect of the use upon the market for or value of the copyrighted work. [126]

Notwithstanding this guidance provided by Act and substantial case law, fair use still retains a seemingly unshakeable reputation for being a murky, ill-defined concept:

> "Although the courts have considered and ruled upon the fair use doctrine over and over again, no real definition of the concept has ever emerged. Indeed, since the doctrine is an equitable rule of reason, no generally applicable definition is possible, and each case raising the question must be decided on its own facts."[127]

Examples of fair use arising from pre-2004 litigation include:

- Book reviewers and biographers quoting from an original work in order to illustrate a point and substantiate criticism;[128]
- Rap artists using copyrighted music in a commercial parody of the original;[129]
- Internet search engines displaying low-resolution thumbnails of copyrighted images in order to direct users to the website hosting the original;[130]
- Viewers recording a television broadcast for later viewing;[131]

---

[126] 17 U.S.C. §107.
[127] House Report no. 94-1476.
[128] *Folsom v. Marsh*, 9 F Cas. 342, 344 (C.C.D. Mass 1841); *Wright v. Warner Books, Inc.*, 953 F. 2d 731 (2nd Cir. 1991).
[129] *Campbell v. Acuff-Rose Music*, 510 U.S. 569 (1994). (Quoting Justice Souter, "While I shall think myself bound to secure every man in the enjoyment of his copyright, one must not put manacles upon science.")
[130] *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 818-822 (9th Cir. 2002).
[131] *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 447-450 (1984).

- Competitors copying protected software for purposes of reverse engineering;[132]

The perceived lack of a unifying theory has prompted lengthy discussions and debates amongst scholars, judges, policymakers, and commentators on what the doctrine means and how it should be interpreted and applied, particularly with respect to new technologies and the emerging social practices they enable.

In his seminal work on the subject, Leon Seltzer, an expert on the intersections of copyright law and scholarly publishing in the years leading up to and immediately following the 1976 Act, argued that fair uses are those that are "productive." He argued that the emphasis on productivity could assist courts in characterizing the essence of the fair use trade-off: "reasonable portions of the work of a prior author" may be used to create a new work which, in turn, "adds to the fount of public knowledge."[133]

Seltzer's viewpoint was adopted by the Ninth Circuit in its seminal *Sony v. Universal* decision. *Sony* asked the Court to consider whether or not the manufacturer of the Betamax should be found contributorily liable for the infringing uses of its consumers when the home video tape recorder was capable of substantial non-infringing uses. The Ninth Circuit determined that, as a matter of law, use of a home video tape recorder was not a fair use because it was not a "productive use."[134] On appeal, however, the Supreme Court declined to ascribe to a rule requiring fair uses to be productive ones. Still, it is fair to say that productivity played an important, although not determinative, role in the Court's decision.

One of the central (ultimately determined to be non-infringing) uses the Court focused on in *Sony* was time-shifting — recording a television broadcast off the air for later viewing. Time-shifting, the court noted, "enlarges the television viewing audience" for rightsholders' content which they broadcast for free for at-home viewing. Although time-shifting involves users making verbatim copies of entire copyrighted works, and no associated new expressive or creative secondary work is produced, the Court implied that the copying may nevertheless be

---

[132] *Sony Comp. Entertainment, Inc. v. Connectix Corp.*, 203 F.3d 596, 599-601 (2000).

[133] Seltzer, L. E. "Exemptions and Fair Use in Copyright: The Exclusive Rights Tensions in the New Copyright Act." *Bull. Copyright Soc'y USA* 24 (1976): 215, discussed by Patry, W. (2005). "Productive Use, Transformative Use, Complementary Use: Who's Right?," The Patry Copyright Blog, Oct. 28, 2005, available at http://williampatry.blogspot.com/2005/10/productive-use-transformative-use.html

[134] "Without a `productive use,' *i. e.* when copyrighted material is reproduced for its intrinsic use, the mass copying of the sort involved in this case precludes an application of fair use." *Sony*, 659 F. 2d, at 971-972.

"productive" because it "expands public access to freely broadcast television programs" and, in so doing, "yields societal benefits."[135]

This reasoning drew upon an earlier case that cited "First Amendment policy of providing the fullest possible access to information through the public airwaves;"[136] applying that rationale to time-shifting, the Court noted that:

> "… access is not just a matter of convenience …. Access has been limited not simply by inconvenience but by the basic need to work. Access to the better program has also been limited by the competitive practice of counterprogramming."[137]

Thus, while the Supreme Court disagreed with the Ninth Circuit that "productive use" was an absolute rule, it nevertheless referenced the link between fair use and productive use as "helpful in calibrating the balance" of interests on review, and for constructing meanings around issues of access.

Perhaps more important to the Court's fair use analysis, however, were its concerns around the potential economic harms (or lack thereof) stemming from the Betamax. While the advent of the Betamax may "require some adjustments in marketing strategy" for rightsholders, the Court concluded that the practice of time-shifting did not impair the commercial value of the broadcast content nor create any likelihood of future economic harm.[138] Potentially foreclosing the manufacture and sale of an article of commerce useful for numerous non-infringing uses because of some potential speculative future harm to certain copyright holders' interests would result in a miscarriage of justice. In cases involving emerging technological innovations that have the potential to majorly alter the market for copyrighted materials, courts should be "circumspect in construing the scope of rights created by a legislative enactment which never contemplated such a calculus of interests."[139] Quoting from Justice Stewart's 1975 opinion:

> "The limited scope of the copyright holder's statutory monopoly, like the limited copyright duration required by the Constitution, reflects a balance of competing claims upon the public interest. Creative work is to be encouraged and rewarded, but private motivation must ultimately serve the cause of promoting broad public availability of literature, music, and the other arts. The immediate effect of our copyright law is to secure a fair return for an 'author's' creative labor. But the

---

[135] *Sony* at 454.
[136] *Columbia Broadcasting System, Inc. v. Democratic National Committee*, 412 U. S. 94, 102. *Sony.* at 454.
[137] *Sony,* 464 U.S. at fn. 40 (1984).
[138] *Sony,* 464 U.S. at 454 (1984); see also *Sony,* 464 U.S. at fn. 40 (1984).
[139] *Sony,* 464 U.S. at 431.

> ultimate aim is, by this incentive, to stimulate artistic creativity for the general public good. 'The sole interest of the United States and the primary object in conferring the monopoly,' this Court has said, 'lie in the general benefits derived by the public from the labors of authors.' When technological change has rendered its literal terms ambiguous, the Copyright Act must be construed in light of this basic purpose."[140]

Justice Blackmun, writing for the dissent, disputed the Court's interpretation and application of fair use doctrine, writing:

> "It may be tempting as, in my view, the Court today is tempted, to stretch the doctrine of fair use so as to permit unfettered use of this new technology in order to increase access …. But such an extension risks eroding the very basis of copyright law, by depriving authors of control over their works and consequently of their incentive to create."

Of particular concern to Blackmun were potential market harms caused by "unproductive 'ordinary' uses" like (in his view) time-shifting that, though appearing harmless when viewed in isolation could become "a major inroad on copyright that must be prevented" when taken in the aggregate.[141]

Following on the heels of *Sony,* the Supreme Court made another major fair use ruling in the *Harper & Row v. Nation Enterprises* case.[142] The subject matter of that case involved a memoir written by former President Gerald Ford and, in particular, portions that described his decision to pardon Richard Nixon. Ford had licensed his publications rights in the memoir to Harper & Row. Prior to its official publication, however, *The Nation* magazine "scooped" its competitors by publishing portions of the memoir without permission. The content published by *The Nation* consisted of direct quotes comprising approximately 300-400 words of the approximately 500-page memoir.

Harper & Row sued *The Nation* for copyright infringement. In its defense, *The Nation* asserted that Ford was a public figure, his reasons for pardoning Nixon were a matter of public concern, and therefore its appropriation should qualify as a fair use. Noting that the right of first publication is a particularly strong right, the Supreme Court held that there was no 'public figure' exception to copyright. In applying the four fair use factors, the Court ruled in favor of Harper & Row, finding: (1) that *The Nation*'s use ("scooping" a competitor) was not a good faith use of the

[140] *Twentieth Century Music Corp. v. Aiken*, 422 U. S. 151, 156 (1975) citing *Fox Film Corp.* v. *Doyal*, 286 U. S. 123, 127. See *Kendall v. Winsor*, 21 How. 322, 327-328; *Grant v. Raymond*, 6 Pet. 218, 241-242.
[141] *Sony* 482.
[142] 471 U.S. 539 (1985).

fair use principle, (2) that the nature of the copyrighted work was informative, (3) that, although the amount copied was small in relation to the whole, it comprised the "heart of the work," and (4) there was actual harm to the market for the original (*Time* magazine cancelled a contract with Harper & Row on the basis of *The Nation's* publication). Beyond the specific facts of the case, this decision is often referenced for the principle that "the single most important element of fair use" is the fourth fair use factor: the effect of the use on the potential market.[143]

Driven by a concern that fair use cases were too often "adjudicated upon ad hoc perceptions of justice without a permanent framework," Judge Pierre Leval modified the concept of productive use and proposed "transformation" as a cogent governing principle for fair use determinations in his seminal article, *Toward a Fair Use Standard.*[144] In describing this new approach, Leval argued that the secondary use must be productive, i.e. not merely repackage or republish the original.[145] In determining whether a use is productive, the concern is whether the secondary use:

> "… adds value to the original – the quoted matter is used as raw material, transformed in the creation of new information, new aesthetics, new insights and understandings – this is the very type of activity that the fair use doctrine intends to protect for the enrichment of society."[146]

In the landmark case, *Campbell v. Acuff-Rose Music, Inc.*,[147] the Supreme Court revisited fair use and applied, for the first time, a transformation analysis to a fair use determination. Since that decision, courts have increasingly focused on whether the purpose and character of a secondary use is transformative:

> "A use is transformative if it does something more than repackage or republish the original copyrighted work. The inquiry is whether the work 'adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message …'"[148]

Despite the inclusion of the transformation analysis, subsequent fair use cases illustrate the doctrine's persistent lack of clarity and signal a potentially important disconnect between the core transformative use cases (that deal with creative or expressive uses that "add, modify, or

---

[143] *Harper & Row Publishers, Inc. v. National Enters.*, 471 U.S. 539, 558 (1985).
[144] Leval, P. N. (1990). Toward a fair use standard. *Harvard Law Review*, *103*(5), 1105-1136.
[145]Leval, P. N. (1990). Toward a fair use standard. *Harvard Law Review*, *103*(5), 1105-1136.
[146] Leval, P. N. (1990). Toward a fair use standard. *Harvard Law Review*, *103*(5), 1105-1136.
[147] *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. at 574 (1994).
[148] *Authors Guild, Inc. v. HathiTrust* (2014), p. 96, quoting *Campbell v. Acuff-Rose*, 1994, p. 579, citing Leval, 1990, p. 1111.

alter" the original with new expression) and an emerging sub-set of transformative use cases (that deal with the technical manipulation of existing works, not to create new/derivative expressive works but rather to enable the subsequent discoverability of new facts and information about the originals).[149]

Yet another interpretation of fair use was offered by Judge Posner in a case before the Seventh Circuit. Drawing upon economic theory to inform legal theory, Posner did not dismiss outright the transformative use language adopted by the Supreme Court (of course, bound by precedent, he could not) but rather conceptualized fair uses as those that are "complementary" to the original as opposed to "substitutional."[150] Thus, Posner regarded fair uses as those that make reference to a pre-existing work as a way to ground the new, complementary one, adding to the market value of the original rather than reducing it by offering a substantially similar substitution.

The focus on market harm as the central factor of fair use is consistent with *Harper & Row,* but, somewhat paradoxically, Posner's approach would shift the transformation analysis from the first factor (purpose and character of the use) to the fourth factor. These tensions between the first and fourth fair use factors raise a unique and complicated tension. As already described, the first factor focuses on the questions of transformation and, almost by definition, courts are reluctant to find market harms following from such uses. Transformative uses by their very nature do not devalue the market for the original *because they are transformative.* By contrast, where a use is not likely to be found to be transformative, the fourth factor rises to dominate the court's analysis and the existence of the secondary work can, in and of itself, be suggestive of a potential market harm to the original (e.g. through lost licensing opportunities, for example). Therefore, the relationship between market harms and transformative becomes even more significant in fair use analyses.

On the heels of the *Campbell* decision, the Second Circuit applied the Court's transformative use analysis in *American Geophysical Union v. Texaco,* a case involving a pattern and practice of photocopying scientific journals in association with research-related activities. Specifically, the record showed that Texaco subscribed to numerous scientific journals for the

[149] This issue is discussed in greater depth in
Centivany, A. (2015). Innovative Deviance: A Theoretical Framework Emerging at the Intersection of Copyright Law and Technological Change, pg. 6.
[150] *Ty, Inc. v. Publications Int'l Ltd.,* 292 F. 3d 512, 2002.

use of its research scientists, that the scientists could place their names on a list to ensure that particular journals would be routed to their office or lab, that the scientists would make photocopies of entire articles of interest and/or relevance to their research, and, the copying was done, in part, in avoidance of paying for additional subscriptions or license fees. Texaco, which at the time of the litigation employed 400-500 research scientists, stipulated that it presumed all or most of its scientists photocopied scientific journal articles in support of their research.

Several key facts distinguish the *Texaco* case from the mass digitization project at issue in this research. There were technological differences: photocopying as opposed to digital scanning. The status of the parties differed: Texaco is a private for-profit company whereas at least some of the universities and institutions involved in the mass digitization project were nonprofit state instrumentalities, and all except Google enjoyed the special/privileged status of "library."[151] The organizational arrangement facilitating or "doing" the copying differed: Texaco supported and/or encouraged the selective photocopying of its researchers whereas, in the mass digitization context, the institutions themselves were active participants in the mass digitization project. The intent or purpose of the copying, and its outcomes and effect, may have differed in significant ways although, at this stage, I will focus only on Texaco's activity and reserve discussion of the intents, purposes, and outcomes of the mass digitization for subsequent chapters of this work.

Notwithstanding all of these potentially important factual differences, however, *Texaco* likely provides the most direct precedential analog to the sort of systematic, institutional, verbatim copying of print materials at issue in this thesis. In particular, the case provides guidance on how a court might apply fair use doctrine to the systematic, institutional, mechanical reproduction of scholarly works for research purposes. In addition, the majority viewed as significant the fact that there was a temporal and causal disconnect between the copying and the purported research uses. The copying done by Texaco's scientists was found not to be spontaneous reproduction, prompted by a specific, active, contemporaneous research purpose, but was rather pre-emptive or "archival" — the scientists made the copies and filed them away

[151] For purposes of this research, the special status of libraries is most acutely made manifest in section 108 of the Copyright Act, discussed *supra*, however, given that fair use is an "equitable rule of reason," it is reasonable to assume that libraries privileged position in society might afford them additional protections under fair use as well, although this is likely to be a door that swings both ways: libraries may be given broader latitude but are held to a higher ethical/legal standard because of their special status.

for possible future reference and use. Indeed, the court estimated that the majority of copies made were filed away and never used at all.[152]

At the outset of its analysis, the court draws attention to the copyright risks associated with innovations in mechanical reproducing technologies. It notes, "the invention and widespread availability of photocopying technology threatens to disrupt the delicate balances established by the Copyright Act." Quoting Blackmun's dissent in *Sony*, the Court recognizes that 'the advent of inexpensive and readily available copying machines … has changed the dimensions' of the legal issues concerning the practice of making personal copies of copyrighted materials."[153] While the focus here was on individual behaviors (personal copying) rather than organizational behaviors per se, the concern for striking the "appropriate balance between authors' interest in preserving the integrity of copyright, and the public's right to enjoy the benefits that photocopying technology offers" may be extended, arguably with more strength, to systematic institutional copying.

In one of its strongest statements, the Court casts doubt upon the very applicability of fair use to mechanical verbatim reproduction:

> "Indeed, if the issue were open, we would seriously question whether the fair use analysis that has developed with respect to works of authorship alleged to use portions of copyrighted material is precisely applicable to copies produced by mechanical means. The traditional fair use analysis … developed in an effort to adjust the competing interests of authors – the author of the original copyrighted work and the author of the secondary work that 'copies' a portion of the original work in the course of producing what is claimed to be a new work. Mechanical 'copying' of an entire document, made readily feasible and economical by the advent of xerography is obviously an activity entirely different from creating a work of authorship. Whatever social utility copying of this sort achieves, it is not concerned with creative authorship."[154]

The Court then, reluctantly, acknowledges that it is bound, under *Sony,* to apply a fair use analysis to the photocopying practices of Texaco's scientists.

---

[152] *American Geophysical Union v. Texaco,* 915-916. It is worth noting that neither he parties nor the Court attempted to discover the photocopying practices of all 400-500 of Texaco's scientists. Instead, "in order to spare the enormous expense of exploring the photocopying practices of each of them" the parties stipulated that "one scientist would be chosen at random as the representative of the entire group." This scientist, Dr. Donald H. Chickering II, was discovered to have copied eight articles from the scientific journal *Catalysis*, three of which he ended up using in his research while the remaining five copies were not used.

[153] *American Geophysical Union v. Texaco* at 917 quoting *Sony* at 801-802.

[154] *American Geophysical Union v. Texaco* at 917 (internal citations omitted)

Given its relevance to this thesis, I will now briefly address the Court's analysis with respect to each of the four fair use factors in turn. With respect to the first factor — *the purpose and character of the use* — the court is concerned with whether or not the copying was transformational or purely substitutive. It noted that Dr. Chickering's primary purpose was substitutional. He made the photocopies,

> "at least initially, for the same basic purpose that one would normally seek to obtain the original – to have it available on his shelf for ready reference if and when he needed to look at it …. Making copies enabled all researchers … to have the article readily available in their own offices. In Chickering's own words, the copies of the articles were made for 'my personal convenience,' since it is 'far more convenient to have access in my office to a photocopy of an article than to have to go to the library each time I wanted to refer to it."

Furthermore, the court characterized this photocopying practice as primarily "archival," a term it used to describe copying done for the primary purpose of providing numerous scientists access to each article without having to purchase additional copies of the original journal.[155]

The Court does, however recognize that the photocopies served other, potentially transformative, purposes as well. It postulates, for example, that, by disembodying a single article from a larger bound journal volume, the photocopying might facilitate lab research that would otherwise be hampered by excessive or bulky papers. Photocopying might also enable Texaco to preserve the original journals against the risks of deterioration or damage posed by chemicals used in the lab and ensure the originals remain free from markings and other marginalia made by scientists.

While the Court recognized that conversion of articles through mechanical reproduction technologies might enable transformative uses in some cases, in this case Texaco's photocopying "merely transformed *the material object* embodying the intangible article that is the copyrighted original work."[156] While the copies were made as part of the scientists' research process, the Court stresses that fair use is primarily concerned with "the work of authorship alleged to be a fair use, not to the activity in which the alleged infringed is engaged." Whatever added value might derive from converting an original copy into a more useable format, the Court's holding seems to suggest that transformative fair use requires a more concrete and express generativity, typically manifesting in a secondary work of authorship.

---

[155] *American Geophysical Union v. Texaco* at 919-920.
[156] *American Geophysical Union v. Texaco* at 923.

The Court determined that the dominant purpose of making the copies was to simply multiply copies without having to pay for additional subscriptions. While the link between Texaco's commercial gain and its copying was admittedly somewhat attenuated, the Court nevertheless noted that the copying did not seem to occur "in good faith to benefit the public," or to "produce a value that benefits the broader public interest"[157] the sort of broad policy considerations motivating fair use. Making copies as an end-run around purchasing additional journal subscriptions "merely supersede[s] the objects' of the original creation" and tilts the first fair use factor against Texaco.

The second and third fair use factors — *nature of the copyrighted work* and *amount and substantiality of portion* used, respectively — were relatively less important to the Court's analysis than either the first or fourth factors. The Court concluded that the scientific journal articles were primarily informative, a finding that weighed in Texaco's favor, while the fact that articles were copied in their entirely weighed against Texaco. Interestingly, the court notes the circularity or reinforcing aspects of the first and third factors noting that the amount and substantiality used (third factor) is assessed in light of whether it was reasonable in relation to the purpose of the copying (first factor). Furthermore, the Court notes that the extent of the copying can "provide insight into the primary purpose of copying," in this case concluding that copying articles in their entirety weakens Texaco's assertion that the overriding purpose and character of the use was to enable the immediate use of the articles in the lab and conversely strengthens the Court's view that the "predominant purpose and character of the use was to establish a personal library of pertinent articles" for its individual scientists.[158]

The fourth fair use factor requires the Court to consider *the effect of the use upon the potential market or value of the copyrighted work.* As previously discussed, in *Harper & Row*, the Supreme Court had characterized this factor as the "single most important element of fair use."[159] In the *Campbell* case that followed, however, that language is conspicuously absent. Subsequent courts have interpreted this omission as the Court signaling its apparent abandoning of the notion that the fourth factor enjoys primacy in fair use determinations: "*Campbell* instructs

[157] *American Geophysical Union v. Texaco* at 922.
[158] *American Geophysical Union v. Texaco* at 925-926.
[159] *Harper & Row,* 471 U.S. at 566.

that '[a]ll [four factors] are to be explored, and the results weighed together, in light of the purposes of copyright."[160]

Under this factor, Courts are only interested in "traditional, reasonable, or likely to be developed markets."[161] Transformative uses, by their very nature, do not supplant any part of the normal market for a copyrighted work. Criticisms and parodies, for example, fill a market niche that copyright holders simply have no interest in occupying and/or have no power to take advantage of.[162] Thus, there is a negative reciprocity between the first and fourth factors; if a court finds a use transformative, that sharply undercuts the possibly of the fourth factors weighing in the rightsholders' favor. As previously noted, the Court concluded that the dominant purpose of the photocopying was archival and non-transformative and thus, the Court was tasked with evaluating the potential effect photocopying had on the market for or value of the original scientific articles.

After noting that the marketing pattern for scientific articles consists of their inclusion in composite journals sold through subscription or as back issues, and there was not an existing market for single stand-alone articles, the Court acknowledged that this complicates the application of the fourth fair use factor. In particular, the Court says:

> "Quite significantly … in the unique world of academic and scientific articles, the effect on the marketability of the composite work in which individual articles appear is not obviously related to the effect on the market for or value of the individual articles."

That said, however, the court noted that Texaco's use would suggest a potential market for licensing revenues derived from individual articles. Through the establishment of the Copyright Clearance Center ("CCC"), publishers have "created … a workable market for institutional users to obtain licenses for the right to produce their own copies of individual articles via photocopying."[163] It is therefore proper, in the Court's view, to consider the existence of a ready market means for paying for the uses.

In addition, the Court tells us that "Congress has impliedly suggested that the law should recognize licensing fees for photocopying as part of the "potential market for or value of" journal

---

[160] *American Geophysical Union v. Texaco* at 926.
[161] *American Geophysical Union v. Texaco* at 930 citing *Campbell* 114 S.Ct. at 1178.
[162] *American Geophysical Union v. Texaco* at 930, citing *Harper & Row* at 568, *Twin Peaks,* 996 F.2d at 1377, *Pacific and Southern Co. v. Duncan*, 744 F.2d 1490, 1496 (11th Cir. 1984).
[163] *American Geophysical Union v. Texaco* at 930.

articles in two ways. The first justification can be derived from the library and archive exceptions in §108 (discussed *supra*) which "narrowly circumscribes the conditions under which libraries are permitted to make copies of copyrighted works" and "implicitly suggests that Congress views journal publishers as possessing the right to restrict photocopying, or at least demand a licensing royalty" from entities that are not eligible for §108 protections.[164] The second justification offered by the Court is that Congress itself "prompted the development of CCC by suggesting that an efficient mechanism be established to license photocopying" and thus it would be illogical to conclude that Congress "did not believe that fees for photocopying should be legally recognized as part of the potential market for journal articles."[165] Due to the potential market for licensing copies of the article, paired with the existence of a reasonable payment mechanism (the CCC) the Court concluded that the fourth factor weighed against a finding of fair use.

Weighed together, the Court concludes that "the institutional, systematic, archival multiplication of copies" was not a fair use. In particular, because the court did not find the copying to be transformative, and a licensing regime existed whereby publishers could extract payments for copies of disembodied single articles, the fourth factor played a major role in its determination. These conclusions, however, raised a sharp dissent that will now be briefly discussed.

First, the dissent found the copying to be "integral to transformative and productive ends of scientific research."[166] The key distinction between the majority and the dissent turns on their differing understandings of what constitutes "research." The majority adopted a more narrow approach to research:

> "Though Texaco claims that its copying is for "research" as that term is used in the preamble of section 107, this characterization might somewhat overstate the matter. Chickering has not used portions of articles from *Catalysis* in his own published piece of research, nor has he had to duplicate some portion of copyrighted material directly in the course of conducting an experiment or investigation. Rather, entire articles were copied as an intermediate step that might abet Chickering's research."

[164] *American Geophysical Union v. Texaco* at 931.
[165] *American Geophysical Union v. Texaco* at 931 citing S. Rep. No. 983, 93d Cong., 2d Sess. 122 (1974); S.Rep. No. 473, 94th cong., 1st Sess. 70-71 (1975); H.R.Rep. No. 83, 90th Cong., 1 Sess. 33 (1968).
[166] *American Geophysical Union v. Texaco* at 932.

Citing sociological and anthropological research on how research practices unfold in scientific labs, the dissent notes the fundamental role scientific journals play in disseminating and communicating information and the long-standing traditions of note-taking, i.e. interacting with, emphasizing, and copying portions of protected expression. Under the *Williams & Wilkins* case, a decision affirmed by an equally divided Supreme Court:

> "It is almost unanimously accepted that a scholar can make a handwritten copy of an entire copyrighted article for his own use, and in the era before photoduplication it was not uncommon (and not seriously questioned) that he could have his secretary make a typed copy for his personal use and files. These customary facts of copyright-life are among our givens."[167]

In the dissent's view, the copying at issue in *Texaco* was part of this note-taking process, rather than, as the majority concluded, archival multiplication of copies to avoid paying for additional subscriptions or license fees. Photocopying is a "technologically assisted form of note-taking" in line with these long-standing customs and practices among researchers.[168] Photocopying saves researchers "the toil and time of recording notes on index cards or in notebooks, and improves the accuracy and range of the data, charts, and formulas he can extract from the passing stream of information."[169] The photocopying of journal articles, and the use of them, "is customary and integral to the creative process of science."[170]

Under *Harper & Row*, "the fair use doctrine is predicated on the author's implied consent to 'reasonable and customary' use"[171] and, under *Williams & Wilkins*, the dissent argues that a reasonable and customary use does not become unfair when the copyright holder develops a way to exact an additional price for the same product.[172] Since copying was a reasonable and customary part of research practices, particularly the note-taking practice, the dissent argues that the first fair use factor should weigh in Texaco's favor.

Going even further, the dissent argues that the photocopying of entire articles for note-taking purposes is transformative. It is not probative, as the majority contended, to consider

---

[167] *Williams & Wilkins*, 487 F.2d at 1350 (1973), *aff'd by an equally divided court,* 420 U.S. 376 (1976).
[168] *American Geophysical Union v. Texaco* at 934.
[169] *American Geophysical Union v. Texaco* at 934.
[170] *American Geophysical Union v. Texaco* at 93-935, citing Shapin, S. (1981). Bruno Latour and Steve Woolgar, Laboratory life. The social construction of scientific facts, Beverly Hills, Calif., and London, Sage Publications, 1979. 8vo, pp. 272, illus.,£ 1 1.25 (£ 5.50 paperback). *Medical history*, *25*(03), 341-342.
[171] *Harper & Row* at 550.
[172] *American Geophysical Union v. Texaco* at 934.

whether or not the material object embodying the copyrighted work is physically transformed by the secondary use:

> "Good notes, being as precise and copious as time allows, do not aspire to transform the original text, but are useful in research only to the extent that they faithfully record the original. Such notes, however, are important raw material in the synthesis of new ideas."

Thus, under the dissent's view, the photocopying of entire articles was transformative as means of note-taking integral to the creative research process.[173]

Second, the dissent found the effect of the use upon the potential market for or value of the work to be illusory:[174]

> "There is a circularity to the problem: the market will not crystallize unless courts reject the fair use argument that Texaco presents; but, under the statutory test, we cannot declare a use to be an infringement unless (assuming other factors also weigh in favor of the secondary user) there is a market to be harmed."

The dissent further adds that only a small proportion of publishers have sought to exact these fees, through CCC or using some other mechanism, which implies that "there is no normal market in photocopy licenses, and no real consensus among publishers that there ought to be one." And, even if extracting such fees should become "administratively tolerable" this fact alone should not determine whether or not photocopying articles is unfair in the first place.[175]

Emphasizing that fair use is an "equitable rule of reason" the dissent questions how the interests of the authors of photocopied journal articles would be promoted by a ruling against Texaco:

> "The single fact that evidences the fair use expectation of the people whose creativity Congress seeks to stimulate, is that they *give away their copyright* in order to promote their work, their ideas and their reputations. … The authors of scientific articles publish in order to gain distinction, appointment, resources, tenure. But they seek and derive absolutely no direct cash benefit from publication. It seems to me that this fact is of great importance: it means that, so long as the copyright system assures sufficient revenue to print and distribute

[173] Relatedly, the dissent argues that the majority's adoption of the term "archival" was a misnomer: "an archive is ordinarily a bulk of documents accumulated by a bureaucratic process and servicing as a resource for public or institutional reference." at 933. And, in addition, doubts that the photocopying was systematic or institutional because the copying was initiated by individual researchers and involved a process of selection. Both of these points are mentioned in passing because, with respect to questions of archival systematic or institutional copying, the arguments of the majority would only be strengthened in the mass digitization context. Therefore, the dissent has little bearing on that analysis.

[174] *American Geophysical Union v. Texaco* at 932.

[175] *American Geophysical Union v. Texaco* at 938.

> scientific journals, the level of copyright revenue is not among the incentives that drive authors to the creative acts that copyright laws are intended to foster." [176]

While the CCC's licensing fees would benefit the copyright holders, the link between that marketing strategy and the sort of creative activity copyright law was designed to foster is tenuous at best. As the dissent argues, courts should consider the incentives for authors "chiefly from the perspective of the authors and scientists … [and] [f]rom their point of view … what is truly important is the wide dissemination of their works to their colleagues" through the cooperative practice of scholarly communication.[177] For those reasons, the dissent finds that the fourth factor should also weigh in Texaco's favor.

Aside from the cases already discussed, one additional fair use case decision is worth bringing in. *Kelly v. Arriba Soft*[178] was a Ninth Circuit decision from 2003 and dealt with a very different sort of activity than what was at issue in *Texaco* but is arguably still potentially relevant to sensemaking and decision-making around the mass digitization project.

The central issue before the Court in *Kelly* was whether or not it was fair use for an Internet search engine to make copies of protected images and display "thumbnails" of the images as part of its search functionality. Kelly was a professional photographer who posted some of his protected works on his website and other licensed websites. Arriba Soft was an Internet search engine that developed software to crawl the web, copying full-sized copies of images from other websites that were then converted into lower resolution thumbnail images, indexed, and used to generate image-based search results to users. The Ninth Circuit concluded that Arriba Soft's creation, use, and display of "thumbnails" of Kelly's works in the search engine was a fair use.

In reaching its conclusion, the Court found Arriba's use of Kelly's images for its thumbnails was transformative.[179] Importantly, although exact replications of Kelly's images were made:

> "the thumbnails were much smaller, lower-resolution images that served an entirely different function than Kelly's original images. Kelly's images are artistic works intended to inform and to engage the viewer in an aesthetic manner. Arriba's use of Kelly's images in the thumbnails is unrelated to any aesthetic purposes. Arriba's search engine functions as a tool to help index and improve

[176] *American Geophysical Union v. Texaco* at 939-940.
[177] *American Geophysical Union v. Texaco* at 940-941.
[178] *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, (9th Circuit) 2003.
[179] *Kelly* at 818.

> access to images on the internet and their related web sites. In fact, users are unlikely to enlarge the thumbnails and use them for artistic purposes because the thumbnails are of much lower-resolution than the originals; an enlargement results in a significant loss of clarity of the image, making them inappropriate as display material."[180]

Acknowledging that courts "have been reluctant to find fair use when an original work is merely retransmitted in a different medium," the court found those cases inapposite because the secondary uses were the same as the original uses. [181] In this case, the Court observes, Arriba's use of the images was for improving access to information on the Internet rather than Kelly's use — artistic expression. Furthermore, because the images are low-resolution, the court concluded it was unlikely that users of the search engine would find the thumbnails substitutional with respect to the originals. Therefore, the Court concluded, Arriba's use was transformative because it does not supersede Kelly's use but rather creates a different purpose for the images.

The Court based its decision, in part, on its earlier decision in *Worldwide Church of God v. Philadelphia Church of God.* That case involved the copying of an entire book to create additional copies for distribution to a different audience. The court concluded that the copying in that case was not transformative because the secondary use was for the same intrinsic purpose as the original – to serve religious practice and education.[182] Quoting from Justice Story in the landmark 1841 fair use case, *Folsom v. Marsh*:

> "There must be real, substantial condensation of the materials, and intellectual labor and judgment bestowed thereon; and not merely the facile use of the scissors; or extracts of the essential parts, constituting the chief value of the original work."[183]

Verbatim copying of an entire book to facilitate its reading and study by a new audience was not transformative because the dominant uses of both communities were the same. Such was not the case in *Kelly* where the thumbnails were a tool for discovering and accessing information rather than observing the aesthetic value that might only be readily available in a higher-resolution version.

Taken together, the statutory construction of fair use and its interpretation by courts suggests that, while the exception is unquestionably crucial to the overall health and function of

[180] *Kelly* at 818.
[181] *Kelly* at 819.
[182] *Worldwide Church of God v. Philadelphia Church of God*, 227 F.3d 1110 (9th Cir.2000).
[183] *Folsom v. Marsh*, 9 F. Cas. 342 at 345 (Supreme Court, 1841).

the copyright regime, it remains a relatively murky predictive tool, particularly where new activities emerging from technological innovations may be concerned. While the purpose and character of the use, including whether or not it is transformative, is central to the fair use inquiry, these questions are subject to complicated and diverging interpretations and opinions. Even concepts like "research" can be prone to disagreement and debates can ensue about what uses may "dominate" or take "primacy" over others.

Several fair uses cases involved questions of verbatim copying of entire print works, and in each instance the use was not deemed to be transformative or fair. But in each of those instances, the copies were made and distributed in print form, not electronic form. While simple conversion of an original work to a new medium does not, in and of itself, result in transformation for purposes of copyright law, new, potentially transformative affordances may be possible by virtue of (or may be embedded in) the digital format. In addition, *Kelly* suggests that improving access to information may qualify as a transformative use so long as steps are taken to guard against the copy superseding the original, i.e. through restrictions on quality or other metrics.

A more implicit thread running through fair use jurisprudence with respect to technological change is the notion that fair use functions as a sort of pressure valve within the system of copyright law. By permitting certain reasonable exercises of discretion, guided by equitable considerations, and in light of changes in technology and social practice, fair use enables otherwise ridged legal doctrine and procedures to accommodate the emerging and evolving contours of the "copyright world." In practice, this has led some scholars to opine that "judges do not apply the four-factor fair use test to *discover* the outcome of the case but rather decide upon an outcome and use the factors to reason backward toward an explanatory rationale"[184] but, as subsequent sections of this work describe, this may be a somewhat inescapable feature of all sensemaking, whether in a legal context or other setting.

The accuracy or common sense appeal of this understanding of fair use jurisprudence does not, however, assuage concerns around the predicative value (or lack thereof) of fair use jurisprudence. No matter how thorough or deep one's understanding is of fair use, the process of identifying and evaluating the risks and opportunities associated with a new and innovative

[184] Interview with legal scholar, transcript on file with author. (AP)

course of action will always be rife with uncertainty. As the sociologist and economic Thorstein Veblen noted:

> "It is not easy in any given case – indeed it is at times impossible until the courts have spoken – to say whether it is an instance of praiseworthy salesmanship or penitentiary offense.'"[185]

Notice, in the above quote, Veblen uses the catch-all "it" to describe some non-specific category of emerging behavior. Even more challenging, however, are emerging behaviors that we cannot anticipate, understand, or even recognize until they have already occurred. The meaning and significance of digitizing an entire print collection of a major research library is likely to yield some the obvious outcomes — we don't have to go to the library to read this book — as well as emerging, unpredictable, possibly still utterly unfathomable outcomes. While fair use doctrine provides clues and breadcrumbs, how and whether it will accommodate and respond to changes on this scale — sociotechnical transformations — remains an open and important question.

Library and Archives Exemptions, §108

Libraries and archives have enjoyed special privileges under copyright law since 1935, when the National Association of Book Publishers and the Joint Committee on Materials for Research of the American Council of Learned Societies entered into a voluntary gentleman's agreement that set out the standard for acceptable reproduction practices undertaken by libraries and archives.[186] The agreement permitted libraries, archives, or similarly situated institutions to make a single photographic reproduction of a copyrighted work for a scholar provided that the scholar represented in writing that he or she would use it for research purposes.[187] Additional restrictions included the prohibition on copying for profit, the requirement that the scholar-recipient was given notice that the use of the reproduction could result in copyright infringement, and precluded substitutional copies (i.e. copies that would substitute for the purchase of a book because of the extensiveness of copying or were deemed unfair "for any other reason.")[188]

The gentleman's agreement and its progeny governed library and archive practices around reproduction and distribution of copyrighted works until the 1976 Copyright Act was

[185] Merton (1957:195).
[186] Section 108 Study Group Report (2008: fn 39).
[187] Section 108 Study Group Report (2008).
[188] Section 108 Study Group Report (2008).

passed and included a codified and fairly complex series of targeted provisions pertaining to libraries and archives, in section 108. Prior to embarking on a description of §108 and its potential implications for library mass digitization, it is important to note that fair use and the library and archives exemptions operate in concert with each other; neither should be read to limit or proscribe the other. In addition, under §108(f), the Act makes explicit the fact that the library and archives exemptions discussed below would not pre-empt or trump contrary contractual obligations[189] and thus the role of contracts and licensing in the mass digitization project and, more generally, in the publishing environment, require continued attention even where one might find §108 protections. Finally, with respect to questions of liability (discussed in greater detail in the remedies section of this chapter) absent a determination to the contrary in an applicable administrative rulemaking proceeding, libraries and archives are not permitted to circumvent technological protection measures that effectively control access to a work for the purposes of exercising the section 108 exceptions.[190] The intent of the drafters of §108 was that the section should be revisited every five years in order to remain current and responsive to changing technologies and practices. Despite a study group report drafted in 2008, it appears that no modifications have been made to this provision.

General Library Exceptions and Requirements for Eligibility:

§108 provides libraries and archives with two general categories of exception: (1) copying and distribution for library purposes such as replacement and preservation, and (2) copying and distribution at the request of patrons for private study, scholarship, and research. These two broad categories will be discussed in greater detail shortly, but at the outset, it is important to keep in mind that §108's exceptions, and all of the various and often confusing subsections, must ultimately relate to one or both of these two broad categories of use.

Keeping that caveat in mind, §108(a) lays out the general library and archives exception. In particular, the first part of this section reads in pertinent part:

> "it is not an infringement of copyright for a library or archives … to reproduce no more than one copy … of a work, except as provided in subsections (b) and (c), or to distribute such copy …, under the conditions specified by this section…."

[189] The section 108 study group reviewed section 108(f) and agreed that the terms of any negotiated, enforceable contract should continue to apply notwithstanding the section 108 exceptions however the group disagreed as to whether or not section 108, especially the preservation and replacement exceptions, should preempt contrary terms in nonnegotiable agreements. To date, no action has been taken to revise section 108(f).
[190] Section 108 Study Group Report (2008).

This section has generated some confusion right out of the gate. Based on a simple reading of the text, §108(a) would seem to suggest that eligible libraries and archives enjoy a stand-alone exception permitting them to make and distribute a single copy of any work, for any purpose, irrespective of the subsequent sections. Under this interpretation, §108(a) provides a minimum allowance that may be expanded by subsequent subsections such as (b) and (c). This interpretation, however, is not controlling. The Copyright Office drafted a comprehensive report indicating that the legislative history of the 1976 Act makes clear that, instead of providing a minimum stand-alone exception, "108(a) instead serves as a chapeau for the specific exceptions set forth in the subsequent provisions."[191] As evidence, this report notes that, after laying out the requirements of eligibility under §108, the House Report accompanying the section then states that "the rights of reproduction and distribution under section 108 apply in the following circumstances:" and goes on to discuss the remainder of section 108's subsections.[192] §108(a) is therefore more accurately understood as enabling the more-specific library and archives-related uses described in the subsequent subsections.

The primary function of §108(a) would therefore seem to be that it states the three basic conditions for eligibility. As the Copyright Act does not provide a definition of "library" or "archives," §108(a) provides a series of threshold requirements for determining which libraries and archives, and which of their activities, are eligible for the 108 exemptions. To qualify under §108, the library or archive must be:

- Open to the public, or at least to researchers in a specialized field;
- The reproduction and distribution activities may not be for direct or indirect commercial advantage; and
- The library and/or archives must include a copyright notice on all copies provided or, if no notice appears on the original copy, it must provide a legend indicating that the work may be protected by copyright.

[191] Section 108 Study Group, "Overview of the Libraries and Archives Exception in the Copyright Act: Background, History, and Meaning," April 14, 2005. Available at http://www.section108.gov/docs/108BACKGROUNDPAPER%28final%29.pdf

[192] Section 108 Study Group, "Overview of the Libraries and Archives Exception in the Copyright Act: Background, History, and Meaning," April 14, 2005. Available at http://www.section108.gov/docs/108BACKGROUNDPAPER%28final%29.pdf

While the eligibility requirements may appear relatively straightforward, like most sections of the Copyright Act, they too have generated some discussion and debate. In particular, as the technological terrain continues to morph conceptions of what constitutes libraries and archives, tensions have emerged surrounding questions of whether and/or how §108 eligibility can or should extend to evolving forms of knowledge and memory institutions such as, for example, the Internet Archive. Relatedly, an active debate has been circulating around whether or not a virtual-only library or archive can or should be eligible for the protections of §108. There is a compelling argument that subsequent subsections of §108 implicitly require a brick-and-mortar establishment; for example, in §108(b)(2) libraries and archives are restricted from distributing digital copies "outside the premises of the library or archives" and this sort of language is inherently difficult to square with a virtual-only context.

Furthermore, in passing the Digital Millennium Copyright Act (discussed in subsequent sections), Congress indicated that it did not intend to broaden the scope of §108 eligibility to include purely virtual institutions:

> "[J]ust as when section 108 of the Copyright Act was first enacted, the term "libraries" and "archives" as used and described in this provision still refer to such institutions only in the conventional sense of entities that are established as, and conduct their operations through, physical premises in which collections of information may be used by researchers and other members of the public. Although online interactive digital networks have since given birth to online digital "libraries" and "archives" that exist only in the virtual (rather than physical) sense on websites, bulletin boards and homepages across the Internet, it is not intended that section 108 as revised apply to such collections of information."[193]

While the changing technological environment and evolving conceptions around the role, function, and "appearance" of libraries and archives leaves some open questions, for the time being it appears that eligibility under §108 is limited to a more traditional and conventional image of libraries and archives.[194]

If §108(a) does not, in and of itself, permit library and archives reproduction and distribution of copyrighted works but rather provides the eligibility requirements, one might

[193] S. Rep. No. 105-190, at 62 (1998).
[194] The 2008 §108 Study Group Committee contemplated this question and ultimately concluded that it would not issue a recommendation that virtual-only libraries and archives should not be eligible for §108 protections.

reasonably wonder why it bothers including the confusing language regarding "making one copy." A compelling explanation may be gleaned from another subsection: §108(g).

Prohibition against Related, Concerted, and Systematic Copying - §108(g):

Under §108(g), libraries' and archives' rights under 108 only extend to "isolated and unrelated reproduction and distribution of a single copy … of the same material on separate occasions." In addition, in §108(g)(1) "related or concerted" copying and/or distribution of "multiple copies of the same material, whether made on one occasion or over a period of time, and whether intended for aggregate use … or for separate use" is expressly prohibited. Furthermore, in §108(g)(2) "systematic reproduction or distribution" of copies made at the request of patrons for private study, scholarship, or research is expressly prohibited.

Tying the prohibitions against related, concerted, and systematic copying back to §108(a) gives new insight into why Congress included the "make one copy" language where it did not intend to grant any stand-alone rights for libraries and archives. The language contained in §108(a) may be tied to concerns around libraries and archives using the exception to avoid paying for additional copies, the assumption being that there is less risk of financial harm to copyright owners where only a single copy is presumed.

The link between general library and archive copying and distribution, and potential financial harms to copyright owners, is made more explicit when read in conjunction with the accompanying House Report:

> "There is a direct interrelationship between the 'indirect commercial advantage' requirement and the prohibitions against 'multiple' and 'systematic' photocopying in section 108(g)(1) and (2). Under section 108, a library and profit-making organization would not be authorized to: (a) use a single subscription or copy to supply its employees with multiple copies of material relevant to their work; or (b) use a single subscription or copy to supply its employees, on request, with single copies of material relevant to their work, where the arrangement is 'systematic' in the sense of deliberately substituting photocopying for subscription or purchase; or (c) use 'interlibrary loan' arrangements for obtaining photocopies in such aggregate quantities to substitute for subscriptions or purchase of material needed by employees and their work.
>
> Isolated, spontaneous making of single photocopies by a library and a for-profit organization, without any systematic effort to substitute photocopying for subscription or purchase, would be covered by section 108, even though copies are furnished to the employees of the organization for use in their work. Similarly, for-profit libraries could participate in interlibrary arrangements for

> exchange of photocopies, as long as the reproduction or distribution was not systematic. These activities by themselves would not ordinarily be considered for direct or indirect commercial advantage since the advantage referred to in this clause must attach to the ultimate profit-making motivation behind the reproduction or distribution itself, rather than to the ultimate profit-making motivation behind the enterprise in which the library is located. On the other hand, section 108 would not excuse reproduction or distribution if there was a commercial motive behind the actual making or distributing of the copies, if multiple copies were made or distributed, or if the photocopying activities were systematic in the sense that their aim was to substitute for subscriptions or purchases."

§108(g)'s prohibitions against related, concerted, and systematic copying raise potentially serious implications for library mass digitization. As subsequent sections of this work describe, the mass digitization project that enabled and/or prompted the emergence of HathiTrust was, at its core, *about* the related, concerted, and systematic copying of protected works. In addition, it was formed around a partnership between a for-profit organization (Google) and libraries (for-profit and nonprofit). A central point of contention with respect to 108(g) as discussed in the House Report relates to the "direct or indirect commercial advantage" prong — whether or not the intent of Google and the libraries in engaging in the mass digitization project was to avoid paying subscription and purchase costs associated with obtaining copies. The House Report's language seems to point toward the subjective intent of the library or archive doing the copying.

This does not seem to be an interpretation shared by the Copyright Office. According to the Copyright Office's 1983 report, cited by its draft 2005 report, "whether or not reproduction is 'systematic' is an objective test; if the reproduction is done via a common plan, regular interaction, organized or established procedure, then it is infringing."[195] The 1975 Senate Report issued in advance of the 1976 Copyright Act revisions noted that, while a definition of "systematic copying" is impossible, a few clear-cut examples might include:

> (1) A library with a collection of journals in biology informs other libraries with similar collections that it will maintain and build its own collection and will make copies of articles from these journals available to them and their patrons on request. Accordingly, the other libraries discontinue or refrain from purchasing subscriptions to these journals and fulfill their patrons' requests for articles by obtaining photocopies from the source library.

[195] The Register of Copyrights, LIBRARY REPRODUCTION OF COPYRIGHTED WORKS (17 U.S.C. 108) 14 (1983), at 139; Section 108 Study Group (2005), "Overview of the Libraries and Archives Exception in the Copyright Act: Background, History, and Meaning," April 14, 2005. Available at http://www.section108.gov/docs/108BACKGROUNDPAPER%28final%29.pdf at 29.

(2) A research center employing a number of scientists and technicians subscribes to one or two copies of needed periodicals. By reproducing photocopies of articles the center is able to make the material in these periodicals available to its staff in the same manner which otherwise would have required multiple subscriptions.
(3) Several branches of a library system agree that one branch will subscribe to particular journals in lieu of each branch purchasing its own subscriptions, and the one subscribing branch will reproduce copies of articles from the publication for users of the other branches.[196]

A proviso to the "systematic copying" clause clarifies that it is not intended to prevent interlibrary loan activities so long as their purpose or effect is not to provide a receiving library with such aggregate quantities of material so as to substitute for purchase. The House intended understandings of this provision to be clarified by the Commission on New Technological Uses of Copyrighted Works ("CONTU")[197] and guidelines were published in the Conference Report for the 1976 Act, which although not binding with the force of law, was agreed to be a "reasonable interpretation of the proviso of section 108(g)(2) in the most common situations to which they apply."[198] Admittedly, the mass digitization project central to this thesis inhabits a vastly different technological world than when CONTU issued its "reasonable interpretation" of systematic copying nearly 30 years prior. The thorny issue of systematic library copying is highlighted here, and will be returned to in the empirical sections of this thesis.

Finally, as mentioned at the outset of the discussion of §108, libraries would not only need to overcome the concerns of §108(g) to qualify for the protections of §108, but their copying and distribution must also fall under one of the two broad categories of use: copying for library purposes such as preservation and replacement, or copying and distribution initiated by library patrons.[199] Doctrine relating to each category will be addressed in turn.

---

[196] S. REP. NO. 94-473, at 70. In addition, the U.S. Court of Appeals for the Second Circuit has analyzed the meaning of "systematic" copying in the context of actions by a library in a for-profit corporation. This analysis, however, was within the fair use context (discussed *infra*), and did not directly address 108(g). *See* American Geophysical Union v. Texaco, Inc., 60 F.3d 913, 916, 919-20, 924-25 (2d Cir. 1994).
[197] *See* H.R.Rep. No. 94-1476, at 78.
[198] Conf. Rep. No. 94-1733, at 71-74 (1976).
[199] With respect to terminology in the Act, although §108 applies to both libraries and archives, in an effort to reduce excess verbiage I will only use the term "library" instead of "library and archives" in the discussion of §108 following this note.

Copying for Library Purposes — §108(b) and (c):

Protections afforded to copying and distribution for library purposes differ depending on whether or not the copyrighted work was published or unpublished.

Copying and Distribution of Unpublished Works

Under §108(b), a library may *make and distribute* up to three copies of an *unpublished* work for *preservation or security* purposes, or for deposit for research use in another eligible library, if the copy is currently in the first library's collection. In addition, if the preservation or security copy of the unpublished work is in a digital format, the copy may *not be distributed digitally outside the premises* of the library. "Premises" is generally understood to mean the physical buildings in which the library or archives is housed, not the wider campus or community in which those buildings are situated.[200]

A potentially important aspect of this subsection is the fact that the copies may only be made for preservation or security purposes and thus may not be lent or shared with patrons. Since the works under this provision are unpublished, a draft report of the Copyright Office in 2005 noted that such lending would "infringe the copyright owner's right of first publication."[201] As previously discussed, however, case law suggests that fair use might provide some coverage where the unpublished works pertain to public figures and/or matters of public concern.

When §108(b) was initially passed, the exemption did not extend to reproductions in "machine-readable language for storage in an information system."[202] There may have been some implicit assumption that such copies could not satisfy the preservation and security requirements of the subsection. Given the advances in digital technologies and the rapid rise of electronic publications and subscriptions, this limitation was lifted in 1998 when the DMCA

[200] Section 108 Study Group Report (2008). An Independent Report sponsored by The United States Copyright Office and the National Digital Information Infrastructure and Preservation Program of the Library of Congress, March 2008.
[201] The Register of Copyrights, LIBRARY REPRODUCTION OF COPYRIGHTED WORKS (17 U.S.C. 108) 14 (1983), pg. 105-106.
[202] Section 108 Study Group Report (2008). An Independent Report sponsored by The United States Copyright Office and the National Digital Information Infrastructure and Preservation Program of the Library of Congress, March 2008.

broadened §108(b) to include the making and distributing of digital reproductions for preservation purposes.[203]

Copying of Published Works

Under §108(c), a library may *make* up to three copies of a *published* work for *replacement* purposes if the copy is *damaged, deteriorating, lost, stolen,* or *if the existing format in which the work is stored has become obsolete*[204] if the library has, through reasonable effort, determined that an *unused replacement cannot be obtained at a fair price* and, if it is a digital copy, the copy is *not distributed digitally outside the premises* of the library.

Worth noting, this subsection only refers to making copies for replacement, not distributing the copies. This is somewhat of a peculiarity since, presumably if a library or archive were making a replacement copy of a work in its collection, it would presumably intend to put the replacement copy into circulation. The Copyright Office report (2005) notes this peculiarity: "this provision … was designed to make sure that items in the library collections are preserved in usable form despite factors — like time, change, and technology — beyond the library's control."[205] That report goes on to say that despite the subsection omitting language related to distribution, "it is nevertheless implied that the library will retain the same rights of distribution to the copy as it did to the original version of the work (under the first sale doctrine), since the purpose of the provision is to permit continued access to the work."[206] In addition, given that the rights under this subsection only kick in when a work is already damaged, deteriorating, lost, stolen, or in an obsolete format, it is also implied that interlibrary arrangements for replacement copying and distribution are permitted.[207]

The language referencing the risks associated with obsolescence was added in 1998 via the DMCA and notes: "a format shall be considered obsolete if the machine or device necessary

[203] Section 108 Study Group Report (2008). An Independent Report sponsored by The United States Copyright Office and the National Digital Information Infrastructure and Preservation Program of the Library of Congress, March 2008.
[204] The Act notes: "a format shall be considered obsolete if the machine or device necessary to render perceptible a work stored in that format is no longer manufactured or is no longer reasonably available in the commercial marketplace."
[205] Section 108 Study Group (2005), "Overview of the Libraries and Archives Exception in the Copyright Act: Background, History, and Meaning," April 14, 2005, pg. 27. See S. Rep. No. 105-190, at 62 (1998).
[206] Section 108 Study Group (2005), "Overview of the Libraries and Archives Exception in the Copyright Act: Background, History, and Meaning," April 14, 2005, pg. 27. See S. Rep. No. 105-190, at 62 (1998).
[207] Section 108 Study Group (2005), "Overview of the Libraries and Archives Exception in the Copyright Act: Background, History, and Meaning," April 14, 2005. pg. 27. The Register of Copyrights, LIBRARY REPRODUCTION OF COPYRIGHTED WORKS (17 U.S.C. 108) 14 (1983), pg. 114.

to render perceptible a work stored in that format is no longer manufactured or is no longer reasonably available in the commercial marketplace."

Copying and Distribution at Patrons' Requests — §§108(d) and (e):

In addition to certain types of copying and distribution for library purposes, the Act also permits libraries to make and distribute copies for the private study, scholarship, and research of patrons under certain circumstances.

Copying and Distributing a Single Article or Part of a Larger Work

Pursuant to §108(d), libraries may make *one copy* of a *single article* contained within a larger collection, or a *small part* of a larger work, at the *request of a patron* or other library under the following four conditions:

(1) The work must be in the collection of the library where the patron makes the request, or from that of another library (§108(d));
(2) The copy must become the property of the requesting patron, and cannot be added to the library's collections (§108(d)(1));
(3) The library must have no notice that the copy will be used for anything other than private study, scholarship, or research (§108(d)(1)); and
(4) The library must display a copyright warning where copy orders are made and attach the same warning to copy order forms (§108(d)(2)).

Copying and Distributing an Entire or Substantial Part of a Work

Pursuant to §108(e), libraries may make *one copy* of an *entire work* or substantial parts thereof, at the *request of a patron* under the following five conditions:

(1) The library must first consult the copyright holder or customary trade sources to determine that a *used or unused copy* is not available *on the market* at a *fair price* (§108(e));
(2) The work must be in the collection of the library where the patron makes their request, or of another library (§108(e));
(3) The copy must become the property of the requesting patron, and cannot be added to the library's collections (§108(e)(1));
(4) The library must have no notice that the copy will be used for anything other than private study, scholarship, or research purposes (§108(e)(1));

(5) The library must both display a copyright warning where copy orders are made, and attach the same warning to copy order forms (§108(e)(2)).

It is worth noting that, because of subsection (1) above, this provision is intended to apply to out-of-print works. In addition, although the express text of this subsection only accounts for patron requests, not requests made by other libraries, the accompanying House Report notes that libraries may invoke this right through interlibrary loan arrangements.[208]

Other General Exemptions — §108(f), (h), and (i):

The Act contains a number of other general exemptions, the relevant portions of which are noted below:

- *Vicarious Liability*: Under §108(f)(1), libraries are exempted from liability for contributory or vicarious infringement stemming from the unsupervised use of reproducing equipment located on its premises so long as it provides notice to the person(s) using the equipment that they may be engaging in infringing activity.
- *Effect of Contrary Contracts:* Under §108(f)(4), if there is an express contractual prohibition against library reproduction and distribution for any purposes, §108 shall not be construed as justifying a violation of the contract. However, fair use may still apply.
- *Orphan Works:* Reflecting the Copyright Term Extension Act ("CTEA") instituted in 1998 which extended the duration of copyright by an additional twenty years, §108(h) was added to permit libraries to copy and distribute published orphan works for preservation, scholarship, or research purposes in the last twenty years of its copyright term.[209] Orphan work status may be inferred where the library has determined after reasonable investigation that:
    - The work is not currently subject to normal commercial exploitation (§108(h)(2)(A));
    - A new or used copy of a work in not available at a reasonable price (§108(h)(2)(B)); or

[208] Section 108 Study Group Report (2008). An Independent Report sponsored by The United States Copyright Office and the National Digital Information Infrastructure and Preservation Program of the Library of Congress, March 2008.

[209] Note that the exception only applies to libraries and archives, not to their patrons or downstream uses of the work.

- The rightsholder has not notified the Copyright Office that the work is either subject to normal commercial exploitation or is available at a reasonable price.[210]

DARK ARCHIVE EXEMPTION FOR COMPUTER PROGRAM, §117

Pursuant to §117 of the Act, owners of copies of computer programs are entitled to make a single additional copy provided that the new copy is created as an essential step in the utilization of the computer program and it is used in conjunction with the computer or machine (i.e. a RAM copy), or that the copy is created for archival purposes only. A further requirement is that the archival copy must be destroyed if and when ownership of the original copy ceases being lawful. In addition, copies made in accordance with this section may be leased, sold, or otherwise transferred only as part of the lease, sale, or transfer of the original copy.

While this section may not pertain in any obvious way to mass digitization, particularly when compared to fair use or the library and archives provisions, there has been some suggestion that this provision may provide libraries with some additional protections with respect to digital content. This topic will be addressed more fully in the empirical sections of this thesis.

EXEMPTION FOR THE PROVISION OF COPIES TO PERSONS WITH DISABILITIES, §121

Pursuant to §121 of the Act, some reproductions and distributions of published works to persons with disabilities may be exempt from copyright infringement if certain conditions are met. The relevant conditions include:

- The work is a nondramatic literary work that is neither a standardized, secure, or norm-referenced test or testing material, nor a computer program (except for the portions that are in conventional human language and displayed to users in the ordinary course of using the computer programs) (§121(a) and (b)(2));
- The work is reproduced and distributed in specialized formats (i.e. braille, audio, or digital text) exclusively for use by blind or other persons with disabilities (§121(a) and (b)(1)(A));
- The copy bears notice that further reproduction and distribution in non-specialized formats constitutes infringement (§121(b)(1)(B));

[210] According to Section 108 Study Group (2005), "Overview of the Libraries and Archives Exception in the Copyright Act: Background, History, and Meaning," April 14, 2005, no rightsholder has ever filed a notice under this provision. Pg. 31.

- The copy bears notice identifying the copyright holder and the date of original publication (§121(b)(1)(C));
- The copying and distribution is done by an "authorized entity" meaning a nonprofit organization or a governmental agency that has a primary mission to provide specialized services relating to education, adaptive reading, or information access needs of blind or disabled persons under 17 U.S.C. 135a (1931) ("An Act to provide books for the adult blind") (§121(d)).

Copyright as Social Relationship

Since its earliest beginnings, copyright laws and law-like privileges have been deeply entangled with technological innovation and social practices. With the Statute of Anne, in England in the early eighteenth century, copyright law's basic means-end formulation, attempting to strike a balance between public and private benefit, was established. This basic formulation continues today. Significant changes in information and communication technologies and reforms of the legal mechanisms (such as scope and duration of protection) although technological changes have enhanced tensions between the public policy underpinning copyright and the legal mechanisms for achieving the law's goals.

Mass digitization highlights the tensions and uncertainties at the intersection of copyright law, technological change, and emerging social practice. This chapter provides an in-depth account of copyright doctrine as it existed when the mass digitization project was being contemplated and demonstrates that, from a copyright law perspective, the potential legality or potential illegality of the project was uncertain. Copyright law (and law in general) mediates our decisions about opportunities and risks when we consider whether or not to engage in a particular course of action, such as the digitization of books. That said, copyright law is inherently retrospective; particularly when new and innovative uses are at issue, it is often difficult to predict in advance how a court might interpret the use for purposes of copyright law and fair use. While copyright considerations played a role in sensemaking and decision-making around the mass digitization project, it is both limiting and misleading to view emerging social practices involving cultural and intellectual works as an algorithmic or mechanized response to (often ambiguous) existing copyright law and doctrine.

While it is true that copyright law favors rules, elements, factor tests, precedent, and other sorts of formulas for human behavior, some would argue that copyright at its essence is less about *property* than it is about *social relationships.* Under this view, copyright is not a "thing" that is created, traded, and protected, but rather copyright is about the creation of value through social collaboration and interaction.[211] This perspective prioritizes the social aspects of copyright in a couple of key ways.

First, in terms of the production, this perspective recognizes that cultural and intellectual works are not created *de novo*, out of whole cloth, but rather emerge as a result of active collaboration between members of society.[212] One of the most notable proponents of this view is Henry Jenkins, founder of participatory culture theory, who famously analogized the creation of cultural works to the philosophical ponderings of the Skin Horse in *The Velveteen Rabbit*:

> "The value of a new toy lies not in its material qualities (not 'having things that buzz inside you and a stick-out- handle'), the Skin Horse explains, but rather in how the toy is used, how it is integrated into the child's imaginative experience: 'Real isn't how you are made. It's a thing that happens to you.'[213] …The Rabbit is fearful of this process, recognizing that consumer goods do not become 'real' without being actively reworked: 'Does it hurt?...Does it happen all at once, like being wound up or bit by bit?'[214] Reassuring him, the Skin Horse emphasizes not the deterioration of the original but rather the new meanings that get attached to it and the relationship into which it is inserted: 'It doesn't happen all at once. You become. It takes a long time.' [215]

Similarly, the value of copyrighted works lies not so much in its *making* as in the process by which the works *become real*, a process that is essentially and fundamentally social, collaborative, and participatory.

Support for this perspective can also be found in research on collective action, or user-focused, models of innovation. In contrast to the 'private investment' model of innovation, which assumes that returns to the innovator result from private goods and efficient regimes of intellectual property protection, this literature recognizes a growing phenomenon of user innovation characterized by users innovating to solve their own as well as shared problems, and

---

[211] Patry, 2009:97.
[212] Litman, J. (2007). Lawful personal use. *Texas law review*, *85*, 1871.
[213] Bianco, M. W. (1926). *The velveteen rabbit*. Pioneer Drama Service, Inc. at pg. 4.
[214] Bianco, 1983: 4-5.
[215] Bianco, 1926:5; Jenkins, H. (2012). *Textual poachers: Television fans and participatory culture*. Routledge, at pp. 51-52.

then freely revealing their innovations to others, thereby transforming the would-be proprietary innovation into a public good.[216]

Beyond the tendency of users to freely reveal their innovations, researchers have observed that user innovation is distinct from traditional sources and practices of innovation in a number of key respects. For example, in many industries users are in a better position than manufacturers to innovate.[217] Eric von Hippel developed a theory of "lead users" to explain this phenomenon. Lead users "are users whose present strong needs will become general in a marketplace months or years in the future."[218] These are users who are so linked in to an existing product or service that they are able to anticipate, imagine, and develop improvements, enhancements, and modifications of those products and services and therefore "can serve as a need-forecasting laboratory for marketing research … (and) … can provide new product concept and design data as well."[219] Furthermore, because sticky information — information which is costly to acquire, transfer, and use in a new location — can be difficult to come by through traditional models of innovation but is readily developed by users, von Hippel suggests that organizations should find it in their own self-interest to support, and possibly harness, the creative and intellectual potential of users by offering toolkits and other customizable options for their products and services. When viewed in this light, the problem of copyright and technological change is not a problem of unauthorized uses but rather a problem of industries being slow to learn "to exploit the value of lead user innovations for commercial advantage."[220]

Granting users a wide berth to innovate on existing products and services may be economically pragmatic for the reasons articulated by von Hippel and others, but structural elements of the existing intellectual property regime tend, in practice, to lead rightsholders away from that approach. While some scholars contend that intellectual property laws' impacts on innovation depend entirely on "the ease with which innovators can enter into agreements for

---

[216] Hippel, E. V., & Krogh, G. V. (2003). Open source software and the "private-collective" innovation model: Issues for organization science. *Organization science*, *14*(2), 209-223.
[217] Hippel, E. V. (1982). Appropriability of innovation benefit as a predictor of the source of innovation. *Research policy*, *11*(2), 95-115.
95.
[218] Hippel, E. V. (1986). Lead users: a source of novel product concepts. *Management science*, *32*(7), 791-805.

[219] Hippel, E. V. (1986). Lead users: a source of novel product concepts. *Management science*, *32*(7), 791-805, pp. 791.
[220] Franke, N., Von Hippel, E., & Schreier, M. (2006). Finding Commercially Attractive User Innovations: A Test of Lead-User Theory*. *Journal of product innovation management*, *23*(4), 301-315.

rearranging and exercising those rights,"[221] von Hippel notes that "it is becoming increasingly clear that … rearranging and exercising intellectual property rights is often difficult rather than easy."[222] Furthermore, it has been fairly widely observed that rightsholders strategically deploy their rights to "achieve private advantage at the expense of general innovative progress."[223] This lends support to the contention that policymakers and theorists are simply out of touch with the ways in which characteristics of intellectual property regimes are actually experienced by innovators and creators and prompted some researchers to conclude "that intellectual property rights are bad for innovation (and competition) in many cases."[224]

Researchers studying sociotechnical systems from a variety of disciplinary perspectives have observed the tendency of industries, particularly those characterized by a single corporate monopoly, to routinize innovation and ultimately shift toward technological conservatism as a technique of self-preservation.[225] This has led some scholars to caution about the "tragedy of the anticommons"[226] which is said to occur when a resource, such as innovation-related information, is underused because "multiple owners each have a right to exclude others and no one has an effective privilege of use."[227] The result in both cases is that the system of intellectual property law, the purpose of which is to stimulate artistic and intellectual creativity, can have the opposite effect, stifling innovation, particularly with respect to its non-traditional sources.

It is worth noting that the research just discussed deals with innovation in general and thus captures behaviors that might implicate patent law as well as copyright. Moreover, while scholars conducting research in this area acknowledge that owners of large collections of copyrighted works in the movie, publishing, and software fields can employ strategies that concentrate information production and retard innovation more generally, [228] there is a tendency

[221] Gallini, N., & Scotchmer, S. (2002). Intellectual property: when is it the best incentive system?. In *Innovation Policy and the Economy, Volume 2* (pp. 51-78). MIT Press.
[222] Von Hippel, E. (2009). Democratizing innovation: the evolving phenomenon of user innovation. *International Journal of Innovation Science*, *1*(1), 29-40.
[223] Foray, D. (2004). *Economics of knowledge*. MIT press; Hippel, E. V. 2009. See also Centivany, A., & Jackson, S. (2012, February). Concentration, incumbency, and interconnection: broadband development and the lessons of history. In *Proceedings of the 2012 iConference* (pp. 317-324). ACM.
[224] Hippel, E. V. (2009:112-113).
[225] See e.g. Centivany & Jackson (2011); Wu, T. (2010). The master switch: The rise and fall of digital empires.
[226] Heller, M. A. (1998). The tragedy of the anticommons: property in the transition from Marx to markets. *Harvard law review*, 621-688; Heller, M. A., & Eisenberg, R. S. (1998). Can patents deter innovation? The anticommons in biomedical research. *Science*, *280*(5364), 698-701.
[227] Von Hippel, E. V. (2009).
[228] von Hippel, E.V. (2009).Benkler, Y. (2002). Coase's Penguin, or, Linux and" The Nature of the Firm". *Yale Law Journal*, 369-446.

amongst scholars to focus their efforts primarily on technological innovation and the invention of useful articles (the purview of patent) rather than primarily expressive creative works (the purview of copyright). This may be due to a perception that restrictions on innovation are less problematic in the copyright context because expressive innovators have more access to alternatives, both in terms of the expressive possibility space and legal alternatives such as creative commons licensing.[229]

Economic pragmatists put a lot of credence into the potential of alternative licensing regimes, such as creative commons, to temper the property-focused maximalist tendencies which currently dominate copyright discourse. Von Hippel posits: "If the commons then grows to contain reasonable substitutes for much of the proprietary intellectual property relevant to the field, the relative advantage accruing to large holders of this information will diminish and perhaps even disappear."[230] This perspective fails to recognize, however, that the nature of copyrighted works, being primarily oriented around original creative and intellectual expression, may make them ill-suited to a pure market substitution analysis. For example, format-shifting aside, what might a reasonable substitute be for the contents of a major academic research library?

Viewing copyright as a social relationship recognizes that the meaning and the value of cultural works emerges from a dynamic and evolving relationship between authors, objects, and their audience. Furthermore, the fact that boundary lines between each of these categories is blurred, not well-delineated, and not static raises implicit concerns about the propriety of legal structures that, by their very design, draw sharp distinctions between authors, users, and so forth and are triggered to an instant in time (when a work is "fixed") and a particular agent (the "author").

The perspective of copyright as a social relationship may seem somewhat far-fetched in light of the dominant discourse and debates in copyright, but it is clearly visible in copyright's overriding goal. The purpose of copyright is not to promote the interests of authors but rather to promote the *progress of science for the whole of society*. As one legal scholar commented:

> "[W]e have focused so hard on the idea that copyright is an incentive for authors and publishers that it is almost as if we thought that we could achieve the

[229] Hippel, (2009:114).
[230] Hippel, (2009).

> "Progress of Science" just by filling up some stockroom somewhere with lots of works of authorship."[231]

The fundamental reason we have copyright laws is because we care about learning which, as any educator or student will tell you, is fundamentally a process of social collaboration. The increasing tendency of copyright law and surrounding discourse to conflate copyright's means (which focus on authors, rightsholders, and property) with copyright's end (societal progress) suggests a troubling shift away from seeing copyright as being by, of, and for society and, instead, being by, of, and for rightsholders.

Finally, the view of copyright as social relations rejects the construction that copyright law, technological change, and emerging social practices are oppositional. They may develop in tension with each other, and may problematize norms, expectations, and traditional ways of doing things, but these three strands are mutually constitutive and, in tangled combination, are crucial to societies' success in accomplishing the goal of copyright.

In *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*, Jenkins and his co-authors described a participatory culture as one with: (1) relatively low barriers to artistic expression and civic engagement; (2) strong support for creating and sharing one's creations with others; (3) some type of informal mentorship whereby what is known by the most experienced is passed along to novices; (4) members who believe that their contributions matter; and (5) members who feel some degree of social connection with one another (at the least they care what other people think about what they have created).[232] Jenkins' approach recognizes that copyright is one part of a larger sociotechnical system and our concerns should focus, more broadly, on accomplishing the larger objectives of society. Thus, while interaction *with* cultural works is an important part of the social-relational aspects of copyright, interacting *through* cultural works as part of the functional, dynamic, and complex interplay between law, emerging technology, and institutions/organizations should be a primary focus.

This chapter discussed the primary goals and purposes underpinning copyright law. While copyright-like protections may have arisen as a reaction to the invention of the movable type printing press, taken more broadly, the law emerged as a way of negotiating the intersections of technological change and social practices with respect to information and

[231] Litman, (2007: 176-177).
[232] Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. NYU press, at pg.7.

communication. From its early beginnings, we see that copyright law is, at its core, driven by Enlightenment principles around freedom of thought and expression and the communication of ideas and expression as a core value. In addition, this chapter described relevant legal doctrine as it existed in 2004 enabling us to recognize the uncertainty and ambiguity of copyright law with respect to the mass digitization of in-copyright works. This could lead us to speculate that a narrow, "conformist," view of the law might lead decision-makers to adopt a "wait and see" approach. Similarly, a risk-averse organization or institution might refrain from digitizing in-copyright works because of the potential copyright risks. On the other hand, one might argue that leadership is about taking risks, and the big risk here was the open question of whether the mass digitization of in-copyright works, and their subsequent uses, would be deemed to be within the law after the fact. The law is *inherently* limited when it comes to preparing us for the future. The next chapter shifts away from a legal analysis to a discussion of relevant social science literature, and considers non-legal theories and approaches to sensemaking and decision-making under conditions of uncertainty, ambiguity, and disequilibrium.

Chapter III: Related Social Science Literature

In this chapter I review social science perspectives on the intersections among law, technological change, and processes of social transformation. As the last chapter described, the tensions between copyright law's overriding public policy of promoting the progress of society through the creation and dissemination of cultural and intellectual works operates in tension with the existing mechanisms for effectuating that policy (namely copyright doctrine). In the context of mass digitization, the tension is further complicated by doctrinal uncertainty: in 2004, when the mass digitization project was embarked upon, it was not clear whether a court would view the activity as a legitimate exercise of fair use (or some other exception or limitation) or whether it would generate a cognizable claim of mass copyright infringement.

How do individuals, organizations, and institutions make sense of this uncertainty? How do they identify and evaluate the potential promise and peril of an activity like mass digitization? How do we as a society, through our judicial system or otherwise, make sense of and make decisions regarding about emerging activities at the fulcrum of the public-private intersection of copyright? How do we navigate the intersections of copyright, technological change, and emerging social practice?

This chapter explores perspectives on these questions through relevant social science literature. The chapter is organized around two main sections. The first section discusses relevant literatures that adopt what might be called a "top down" theoretical approach. These are perspectives that focus on behavior as a natural outflow of systemic, institutional, structural social arrangements. The centerpiece of this section is Robert K. Merton's structural strain theory, but it also discusses social disorganization theory, relevant work of economists including Hirschman and Schumpeter, and Hughes' concept of reverse salient. This literature suggests that some of the answers to the above questions, and some of the insights into processes of transformation, are likely tied to structural and institutional aspects of sociotechnical systems.

But transformation of the sort exemplified by the mass digitization project is not purely a matter of systemic, structural, or institutional arrangements. Transformation is also shaped by more dispersed, distributed, relatively more modest instances of human agency that coalesce and percolate up to generate change. Therefore while the first section of this chapter provides a top-down orientation, the second part provides a bottom-up approach to the processes of sociotechnical transformation. The centerpiece of this section is Karl Weick's research on sensemaking and decision-making but it also brings in relevant organization science theories about change and learning. This section highlights how, in addition to being impacted by structural and institutional forces, processes of transformation are also activated through instances of human agency, interaction, and cooperation.

The goal of this chapter is to lay out and begin weaving together these different but related social science perspectives, synthesizing an analytic framework that contributes to the study and understanding of the interplay among copyright law, emerging technologies, and processes of social transformation. A framework that blends social science and law-based approaches can best take into account fairly complex copyright-related concerns while remaining attuned and responsive to other (non-legal) social, organizational, institutional, and structural pieces of the HathiTrust story.

Structural aspects of transformation: A top-down approach

This section begins with the work of social disorganization theorists. The social disorganization perspective provides a long chain of analytical thinking about the relationships among change and social practices beginning with the work of the Chicago School of sociologists in the 1920s and continuing through recent times.

At the heart of social disorganization theory is the belief that rapid change — particularly in the areas of technology, institutions, and demography — damages society's web of normative controls and causes social disorganization. The theory emerged in the period after World War I when the United States' economy, modes of industrial and agricultural production, and communications infrastructures were undergoing fundamental changes. One of the responses to the intense growth and innovation stimulated by the war was that social institutions, organizations, and interactions were becoming increasingly mechanized. Social disorganization theorists adopted an approach characterized by professional detachment and technical efficiency,

viewing human relationships and functioning as a complex system that could be tweaked and enhanced (later explored most pointedly through the work of cybernetics theorists) much as post-War business organizations became increasingly mechanized, algorithmic, and structured.[233] To the extent that technological change generated normative dissension and produced social deviance, these ill-effects were technically correctable by tweaking and modifying the social systems of interaction.[234]

The near-term effect of significant technological change is that it creates a sort of normative disequilibrium that results in increases in deviance. Over the longer-term, however, social disorganization theorists suggest that the social disequilibrium is a form of normative competition. In other words, deviance is a marker or evidence of the process of sociotechnical transformation whereby a society disrupted will, eventually, move toward reorganization around a new series of norms, expectations, and shared practices.

An example of this process was explored by Thomas and Znaniecki in *The Polish Peasant in Europe and America,* one of the earliest social disorganization studies (1919). This study highlighted the problems endemic to immigrants who are unable to successfully transplant the "ways of the old world to the new world" and who also have "tremendous difficulty assimilating the norms and standards of their new social environments." [235] Social disorganization, the authors note, is the "process by which the authority and influence of an earlier culture and system of social control is undermined and eventually destroyed." Significant change epitomized, in this case, by their relocation to the United States, unmoors shared social norms amongst Polish immigrants and facilitates their drift into deviance.

In terms of methodological approaches, the hallmark model for studying social disorganization was developed by Park and Burgess in their work *Human Communities*.[236] Park and Burgess viewed the normative order of a well-organized society as existing in a state of symbiosis amongst its members. Using the metaphor of invasion, Park and Burgess developed an ecological model of social disorganization whereby: 1) change introduces an invasive/competing species into a community; 2) there is conflict/competition for dominance between species; 3) the weaker species accommodate the demonstrably dominant one(s); and, 4)

[233] See e.g. Yates, J. (1993). *Control through communication: The rise of system in American management* (Vol. 6). JHU Press, for a discussion of the rise of systematic management.
[234] Pfohl, S. J. (1994). *Images of deviance and social control: A sociological history* (p. 2). New York: McGraw-Hill.
[235] Pfohl, (1994: 186).
[236] Park and Burgess, *Human Communities,* 1952.

all members of the community assimilate a new order of symbiosis based upon the accommodative outcomes of the previous three stages. Under this ecological model of social disorganization, deviance remains widespread until society becomes symbiotically reorganized around a new dominant form of normative order through the processes of accommodation and assimilation.[237] Although the ecological model may be somewhat unnecessary limiting in a modern context, it echoes the sorts of mutual readjustment and social reconfigurations that STS scholars recognize as fundamental to processes of sociotechnical transformation.

Social disorganization theories were subsequently adopted and modified by social control theorists who applied a social-psychological perspective to the concerns originally articulated by the Chicago School. Control theorists agreed that normative disequilibrium results in disrupted socialization but were primarily interested in how this disruption was internalized. In other words, recognizing that social disorganization is a causative factor in, rather than a determinant of, deviance, control theorists focused on identifying factors which might be helpful in differentiating which individuals or groups might be more or less likely to engage in deviant behavior. As one scholar notes, control theorists are particularly interested in making explicit something most disorganization theorists leave implicit: the link between social disequilibrium and the likelihood of acting outside norms, or what came to be called deviance.[238]

For example, Reckless, one of Park's students, generated a model to help explain why, when faced with acute social disorganization, some individuals and communities hold steadfastly to traditional norms while others fall into forms of deviance. A central feature of Reckless's model is the notion that social-structural factors impacting socialization, what he calls "outer containment factors," are mediated by social-psychological factors, what he called "inner containment factors." During periods of high social disorganization, when normative constraints deteriorate due to rapid change, outer containment factors, conceptualized as structural buffers in a person's immediate social world, would fail to adequately protect the person from slipping into deviance unless there is high strength of inner containment.[239] (Fig. 12)

[237] Pfohl, (1994: 188).
[238] Pfohl, (1994:205).
[239] Pfohl, (1994: 205-206).

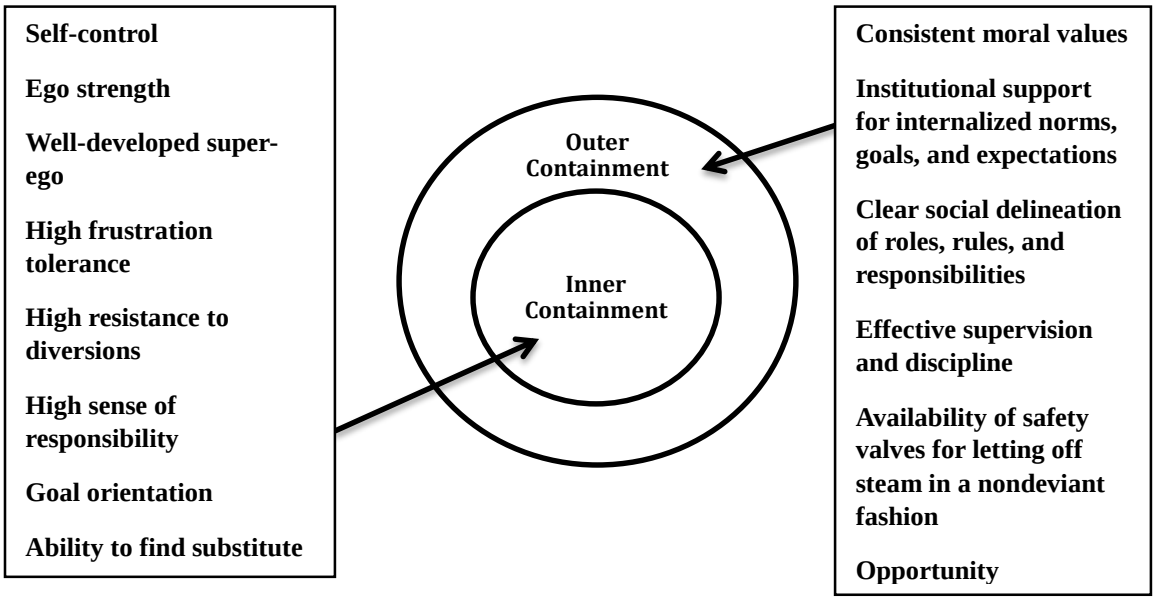


Figure 12. Reckless' containment factors impacting socialization

Similarly, in *Causes of Delinquency,* Hirschi focused on the role of social bonds in influencing the likelihood of deviance or conformity in the face of normative disequilibrium.[240] In Hirschi's view, social disorganization primes deviant behavior but whether or not a particular individual or community engages in deviant behavior ultimately depends on the condition of their bond to society. If the social bonds are weak or broken, deviant behavior is likely. Social bonding, in Hirschi's view, has inner and outer dimensions. (Fig. 13) Significant technological changes may result in deviance by disrupting the social bonding process through weakening or suspending the power of internalized beliefs (norms) and/or external attachments, commitments, and involvements.[241]

Attachment: strength of a person's ties to others, particularly those who conform to society's standards

Commitment: degree to which a person is tied to conventional ways of behaving by virtue of the social rewards obtained by acting in accordance with prevailing norms

Involvement: proportion of a person's time engaged in the pursuit of conventionality

[240] Hirschi, T. (2002). *Causes of delinquency*. Transaction publishers.
[241] Pfohl, (1994: 206-207).

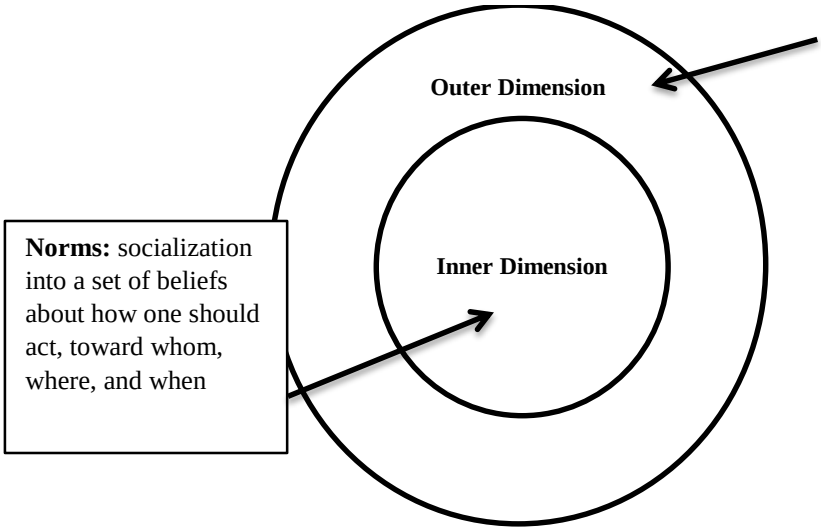


Figure 13. Hirschi's dimensions of social bonds impacting socialization

We can take a number of useful lessons from social disorganization and social control theory. The first may be the recognition that as significant innovations and changes in technology permeate society, enabling new forms of social practice and participation and disrupting old ones, the result may be a disruption or questioning of previously shared norms and expectations within a given organization or institutional context, or within society more broadly. Technological change can result in social disequilibrium where normative consensus is replaced by normative dissension. Social disorganization theorists suggest that, while this normative dissension may be characterized as deviance within its immediate context, over the longer term this behavior tends to reflect normative competition or what cyberneticists might call a process of homeostasis. Significant, rapid change is a trigger for mutual readjustment of laws, technologies, and social practices and, in combination, the system is transformed through dynamic evolutions between balance and disequilibria.

In addition, drawing from the social control literature, we can begin to distinguish and refine the relationship between change and deviance. Rather than assuming the relationship is linear and determinative, control theorists suggest that change *primes* rather that directs deviance. Also important are the social-psychological factors which recognize the important role of human agency in determining outcomes. For example, social structures such as the degree of

institutional support for internalized norms, goals, and expectations, how committed, attached, and involved a person or organization might be in the broader social context, the availability and effectiveness of safety valves for engaging in innovative and unorthodox behavior without penalty, and the opportunity to accomplish important personal and social goals through conforming behavioral channels can play key roles in terms of how disorganization and disequilibrium is responded to. Also important are what Reckless calls the "inner containment" factors which is roughly analogous to what Hirschi called the "inner dimensions of social bonds" which can include considerations such as toleration for frustration, sense of responsibility, goal orientation, the ability to find substitute behaviors, and so forth. In the context of the mass digitization project, these may provide touchstones by which we can begin to distinguish and understand the various organizational and institutional responses to the risks and opportunities of the project.

While social disorganization and control theories provide some potentially useful ways of approaching and analyzing the mass digitization project — in digitizing millions of books, the project rapidly and fundamentally altered the ecology of traditional academic research libraries — these theories suffer from a number of potential weaknesses or limitations. For example, critics of these approaches contend that the theories fail to fully operationalize the concept of disorganization and control. Perhaps what might be characterized as disorganization could more accurately be seen as differential organization. In addition, these approaches have faced criticism for neglecting socially productive forms of deviance, such as innovation. Finally, social disorganization theorists' perspective that deviance results from one's spatial location in the natural ecology of a rapidly changing society fails to account for the causative potential of structured differences in social power and social class, so-called structural inequalities. These critiques are answered, in part, by bringing in additional theoretical perspectives discussed below beginning with Robert K. Merton's structural strain theory of deviance.

Focusing also on the generation, function, and effects of social instability, Robert K. Merton posited in his 1938 book, *Social Structure and Anomie*, that anomie did not derive from normlessness (as Émile Durkheim had argued) but was rather a consequence of structural inequalities in society. In particular, Merton was concerned with structural inequalities that reflected a mismatch between cultural goals and socially available, i.e. legitimate, means to achieve such goals. Deviance, in Merton's view, is the normal product of an unequal society.

The imbalance or disequilibrium between socially accepted goals and the availability of legitimate institutionalized means to accomplish the goals produces a strain that pushes people toward deviance.

Merton further argues that the deviance will follow one of four possible paths. (See Fig. 2 reproduced below as Fig. 14)

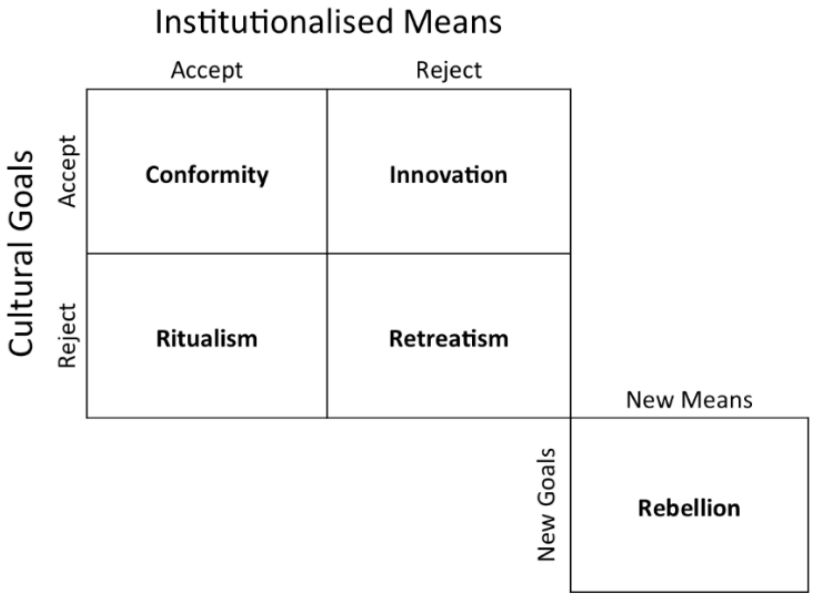


Figure 14. Robert K. Merton's Typology of Deviance

Under this typology Merton distinguished between behavior on the basis of whether or not it: 1) accepts or rejects a given cultural goal, and 2) accepts or rejects institutionalized means for accomplishing the cultural goal (although he acknowledges that, in some cases, institutionalized means may not exist). Working counter-clockwise from *Conformity* in Figure 14, *Ritualism* is behavior that is deviant because it "follows the rules" but is not motivated by the culturally accepted goal. An example of ritualistic deviance might be a person who attends college because they feel like "it's what people do" rather than because they want to develop knowledge, be successful, have more employment opportunities, and so forth. *Retreatism* is behavior that is deviant because it rejects both the culturally accepted goal and the institutionalized means of accomplishing the goal. An example of retreatist deviance might be

the iconic loner or hermit who shuns the basic mainstays of life, preferring instead to live off the grid, separate from the rest of society.

Rebellion describes deviant behavior that does not fit neatly within Merton's 2x2 typology. Rebellious deviance is behavior that not only rejects both the culturally accepted goal and the institutionalized means to accomplish it, but seeks to replace either or both with a new order. The residents of Copenhagen's Christiania (also called "Freetown Christiania") may be an example of rebellious deviance as they took over a mostly abandoned military base in the city to create a purportedly autonomous new society with its own goals, rules, and forms of governance.

Finally, *innovation* describes behavior that is deviant because, while it accepts the culturally accepted goal, it rejects the institutionalized means for accomplishing the goal or such means are unavailable. Merton's classic example of innovative deviance is Al Capone, a person who whole-heartedly accepted the success goal but rejected the legitimate (i.e. legal) means of accomplishing success. It is worth noting that, of the four paths of deviance Merton identifies, innovative deviance is the only path which promotes the culturally accepted goal. In this work, I will focus on this path of deviance and, in subsequent sections, will argue that participation in the mass digitization of in-copyright works was an example of innovative deviance not because the participants rejected the institutionalized means of accomplishing the digitization but because legitimate means to do so were unavailable for reasons I will later describe.

Merton defines innovative deviance as the "creative use of illegitimate means to obtain valued legitimate ends" and "the rejection of institutional practices but the retention of cultural goals."[242] "Innovative deviance is a normal outgrowth of having accepted cultural goals without having been provided with the opportunity to legitimately achieve those goals."[243]

In his later works, Merton focuses on innovative deviance as tool in economic or business dealings. Explaining the cause of deviant innovation, Merton writes:

> "Great cultural emphasis upon the success-goal invites this mode of adaptation through the use of institutionally proscribed but often effective means of attaining at least the simulacrum of success – wealth and power. Innovative deviance occurs when the individual has assimilated the cultural emphasis upon the goal without equally internalizing the institutional norms governing the way and means for its attainment." [244]

[242] Merton (1938: 230).
[243] Merton (1938).
[244] Merton (1957: 195).

Furthermore, Merton suggests that we tend to look upon upper class, or white-collar, innovative deviance with a degree of tactic approval. At the top economic levels, he observes:

> "[T]he pressure toward innovation not infrequently erases the distinction between the business-like strivings this side of the mores and sharp practices beyond the mores. As Veblen observed: 'It is not easy in any given case – indeed it is at times impossible until the courts have spoken – to say whether it is an instance of praiseworthy salesmanship or penitentiary offense.'"[245]

The focus on white-collar crime and organized crime reflects one possible version of innovative deviance, namely behaviors that *reject* institutionalized means for accomplishing cultural goals. But innovative deviance can also result from circumstances where legitimate institutionalized means for the accomplishing a goal simply do not exist. This is often the case, according to Merton and many others, due to structured inequalities in society based on race, gender, class, wealth, and so forth. For those in society's lower strata, where the pressure toward deviance is greatest and, not coincidentally, deviance generates the strongest social control response innovative deviance have very little to do with a rejection of institutionalized means.[246]

In the context of mass digitization, the structured inequalities are based on an imbalance between the public's interest and the private interests of rightsholders, an imbalance exacerbated by significant changes in information and communication technologies and recent compounding reforms to copyright law which favor rightsholders at the expenses of the public. The relevant mismatch or imbalance is between the culturally accepted goal — promoting the progress of society through the creation, dissemination, and use of cultural and intellectual works — and the institutionalized means of accomplishing the goal — the Copyright Act and associated case law. The transition from an analog world to a digital one through the mass digitization of libraries' entire print collection would advance the goals of copyright, yet there were no clearly legitimate means for accomplishing the goal based on copyright law as it existed at the time. Given the uncertainty and ambiguity in the law, mass digitization could be interpreted as a form of socially productive potential infringement. As the empirical sections of this work will describe, participants and observers held different opinions on the potential legality (or illegality) of the mass digitization project.

[245] Merton (1957: 195).
[246] Merton (1957: 198).

Furthermore, to the extent that different libraries and institutions responded differently to the mass digitization project, some taking on roles as willing partners, some limiting participation to public-domain works, others refusing to participate, suggests that internalized norms discussed earlier and mentioned again by Merton also played a key role in the outcome.

Merton also acknowledges the pragmatic advantages to researching innovative deviance as compared to the other paths of deviance he identified. In particular, because "the law provides formal criteria of this form of deviance, it is relatively visible and readily becomes a focus for study." In contrast, "other forms of behavior which are sociologically though not legally departures from accepted norms … are less visible and receive little attention."[247] That said, the research attention Merton cites is in the domain of criminal deviance such as theft, violent crimes, and so forth, rather than intellectual property infringement. Extending Merton's theories to the IP context could provide some useful insights not already addressed by research in this area.

Several sociologists have carried forward Merton's theories on anomie and the paths of deviance. Richard Cloward, for example, contributed to Merton's theory by arguing that structured inequalities in society may actually block access to both legitimate and illegitimate means of goal attainment. Cloward calls this "double failure" stating: "If illegitimate means are unavailable, if efforts at innovation fail, then retreatist adaptations may be the consequence, and the 'escape' mechanism chosen by the defeated individual may perhaps be all the more deviant because of his 'double failure.'"[248]

In addition, in work with Ohlin, Cloward observed that social adjustments to the strain of blocked opportunity can be collective, calling these collectives "subcultures,"[249] and moreover, by adopting elements of Edward Sutherland's learning perspective of deviance, concluded that deviance must be, at least partially, learned in interaction with others. This notion, which has come to be called the cultural transmission theory, focuses on the relational dynamics by which one form of adjustment to frustration is selected instead of others. Subsequent research on cultural transmission suggests that it is a communicative process dependent on innovation in the

[247] Merton (1957: 230).
[248] Cloward, R. A. (1959). Illegitimate means, anomie, and deviant behavior. *American sociological review*, 164-176: 175.
[249] Cloward, R. A., & Ohlin, L. E. (2013). *Delinquency and opportunity: A study of delinquent gangs* (Vol. 6). Routledge.

creation of new, problem-solving responses to changes in the environment and the ability to communicate this behavior and/or imitate the behaviors of others.[250]

Structural strain theories have received their fair share of criticism. For example, Albert Cohen argued that Merton's theory was too atomistic in the sense that it places undue emphasis on the individual and the discontinuity of the deviant act. Cohen argues, instead, that the way a person experiences frustration and selects a particular mode of deviant adaptation is highly dependent upon his or her interpersonal associations with his or her social reference group. This criticism is well-founded and, to some extent, was resolved by Cloward and Ohlin's cultural transmission theory and their studies of collective deviance. In addition, Merton himself also ultimately incorporated Cohen's concerns as his own.

Cohen also took issue with Merton's focus on utilitarian deviance. Cohen acknowledges that, while Merton's theory of deviant adaptation is a "sociologically sophisticated and highly plausible … explanation" of utilitarian, or instrumentally-oriented, forms of deviance, it failed to account for non-utilitarian or expressive forms of deviance."[251] Subsequent researchers have picked upon on the distinction between utilitarian and non-utilitarian forms of deviance. For example, Gould (1987) contrasts deviant innovation with deviant aggression. Citing Merton's theories, Gould states that deviant innovation "implies a rational, utilitarian innovation to obtain the facilities necessary to achieve a desired goal. The innovation will be termed deviant when and if it results in acts that violate the institutionalized normative order…."[252] By contrast, deviant aggression, according to Gould, is non-utilitarian deviance. Ultimately, for Gould, the distinction comes down to the subjective intent of the deviant actor. Providing the example of a person who breaks the windshield of a car, Gould reasons that the individual would be characterized as a deviant innovator if they were a "glazier who does so for profit" — because they would be paid to repair the damage – versus a deviant aggressor who broke the windshield as nothing more than an act of vandalism.[253]

---

[250] Lehmann, L., Feldman, M. W., & Kaeuffer, R. (2010). Cumulative cultural dynamics and the coevolution of cultural innovation and transmission: an ESS model for panmictic and structured populations. *Journal of evolutionary biology*, *23*(11), 2356-2369.
[251] Cohen, A. K. (1955). DELINQUENT BOYS; THE CULTURE OF THE GANG.
[252] Gould, M. (1987). *Revolution in the Development of Capitalism: the Coming of the English Revolution*. Univ of California Press, at pg. 75.
[253] Gould, M. (1987). *Revolution in the Development of Capitalism: the Coming of the English Revolution*. Univ of California Press, at pg. 366.

It is worth noting that Gould does not appear to be applying the philosophical doctrine of utilitarianism stemming from the work of Jeremy Bentham and others which defines utilitarian as "the greatest good for the greatest number," since it would be impossible to know whether the glazier's profit was a greater good than the loss of property to the car owner. Rather, it appears that Gould is employing a definition that more closely aligns utilitarian with utility – the glazier's breaking of the windshield is utilitarian because it is "of or relating to utility" or, drawing upon Cohen's work, is "instrumentally-oriented" because it motivated by an economic self-interest rather than (merely) expressive aggression.

Furthermore, Gould's critique may fail to fully incorporate the goal orientation of Merton's concept of innovative deviance. While it is true that innovative deviance does contain some inherent reference to utilitarianism via its premise that the behavior in question be motivated by the pursuit of a culturally accepted goal, it is worth noting that none of the other paths of deviance articulated by Merton include such a condition. Therefore, it may be that non-utilitarian or expressive forms of deviance may be captured by ritualist, retreatist, and rebellious forms of deviance.

Until this point, I have focused on sociological theories which have framed processes of change and transformation in terms of interactions between systemic, structural, and institutional forces and shared norms and practices. I would like to break away from these normative frameworks to introduce two additional perspectives. The first perspective is that of the economist Albert Hirschman. The second is the perspective of the historian of technology Thomas Parke Hughes.

Hirschman's approach to change and transformation is oriented around notions of rationality, efficiency, and participation through political and economic action rather deviance and conformity. Furthermore, while the previously discussed theories used change — particularly significant, rapid technological change — as a jumping off point, Hirschman uses change — particularly decline and deterioration in goods and services — as his jumping off point. Notwithstanding this fairly significant difference, I believe Hirschman's approach offers a complementary perspective to those discussed earlier and offers potentially important insights into processes of sociotechnical transformation in the context of mass digitization context.

In *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations, and States* (1970), Hirschman begins with the premise that "lapses from efficient, rational, law-abiding,

virtuous, or otherwise functional behavior always happen, no matter how well a society's basic institutions are devised."[254] Moreover, "society learns to live with a certain amount of such dysfunctional or mis-behavior" because, in his view, this sort of deterioration is:

> "… the inevitable counterpart of man's increasing productivity and control over his environment. Occasional decline as well as prolonged mediocrity – in relation to achievable performance levels – must be counted among the many penalties of progress."[255]

Societal deterioration, thus, goes hand-in-hand with economic surplus and technological progress and any "homeostatic controls" with which societies are equipped to correct the elements of decline are "bound to be rough."[256] This basic framework seems to align with the previously discussed perspectives on sociotechnical reconfiguration in responses to technological change.

While the search for social arrangements that wholly eliminate deviance or deterioration may be futile, Hirschman notes that society nevertheless continues to:

> "… marshal from within itself forces which make as many of the faltering actors as possible revert to the behavior required for its proper functioning … lest the misbehavior feed on itself and lead to general decay."[257]

In fact, Hirschman argues that the progress of society is almost inversely correlated to the severity of sanctions it seeks to impose on deviance. Recognition of the unpleasant truth that some level of deviance is unavoidable "has been impeded by a recurring utopian dream: that economic progress, while increasing the surplus above subsistence, will also bring with it disciplines and sanctions of such severity as to rule out any backsliding that may be due, for example, to faulty political processes."[258] Explaining further:

> "The common assumption of these constructs is simply stated: while technical progress increases society's surplus above subsistence it also introduces a mechanism of the utmost complexity and delicacy, so that certain types of social misbehavior which previously had unfortunate but tolerable consequences would now be so clearly disastrous that they will be more securely barred than before. As a result, society is, and then again it is not, in a surplus situation: it is producing a surplus, but is not at liberty *not* to produce it or to produce less of it

[254] Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states* (Vol. 25). Harvard university press.
[255] Hirschman (1970: 6-7).
[256] Hirschman (1970: 6-7).
[257] Hirschman (1970:1).
[258] Hirschman (1970:7).

> than is possible; in effect, social behavior is as simply and as rigidly prescribed and constrained as it is in a no-surplus, bare subsistence situation."[259]

Within this context, where progress, deterioration, and deviance co-exist and are to a large extent interdependent, Hirschman contends that organizational behavior reflects a characteristic he describes as "slack." Organizational behavior is primarily about satisficing, he argues, rather than obtaining the highest possible profits.[260] Therefore, organizational development may more appropriately be understood "not so much (as) finding optimal combinations of resources and factors of production as on calling forth and enlisting for development purposes resources and abilities that are hidden, scattered, or badly utilized."[261]

How are an organization's members to respond when the organization seems "permanently and randomly subject to decline and decay" or, stated somewhat differently, "to a gradual loss of rationality, efficiency, and surplus-producing energy, no matter how well the institutional framework within which they function is designed"?[262] Hirschman's theory suggests two possible alternative answers. First, members can "exit" by making use of market competition to withdraw from the relationship and possibly select an alternative. Second, they can "voice" their complaints and hope that, through communication of the grievance, they can prompt the organization to repair the perceived lapse or defect. Exit and voice thus represent a convergence of economic and political action.

Hirschman argues that economists bias exit over voice because it is neat, impersonal, indirect, and most importantly efficient. However, exit can only signal an organization's possible decline and does not, as compared to voice, communicate much information about the reasons for the decline. Therefore, Hirschman argues that organizations should prefer voice over exit. Furthermore, one way in which organizations can sway members' cost-benefit analysis regarding whether to choose exit or voice is to foster loyalty.

Loyalty is therefore the lynchpin in Hirschman's analytic framework. Fostering member loyalty carries with it its own challenges because those members who care most about the organization "and therefore are those who would be the most active, reliable, and creative agents of voice are for that very reason also those who are apparently likely to exit first in the case of

[259] Hirschman (1970: 8-9).
[260] Hirschman (1970::11).
[261] Hirschman, A. O. (1959). The Strategy of Economic Development, p. 5.
[262] Hirschman (1970:15).

deterioration."[263] When the organization creates mechanisms through which members can make their voice heard and effectuate reform, members' loyalty grows as does their devotion to the organization's success.

Hirschman's exit, voice, and loyalty framework is relevant to inquiries into processes of sociotechnical transformation and mass digitization because it can provide useful clues about the relationship between social structures, such as economic and political institutions, whose effects filter down through society, and more individual or dispersed forms of engagement and participation, such as exit, voice, and loyalty, that percolate and coalesce to generate transformations from the bottom-up. His framework, which is oriented around concepts of rationality/irrationality, efficiency/inefficiency, exit/voice, and so forth, offers a nice analytic counterpoint to sociological theories that are primarily focused on norms, conformity, and deviance. Decisions involving the mass digitization of in-copyright works may be partly motivated by norms and shared expectations, but they may also be motivated by economic and political concerns.

Another non-normative concept that is potentially relevant to the topic of sociotechnical transformation at the intersection of copyright law, technological change, and emerging social practices is Thomas Parke Hughes' work on the social construction of technology and, in particular, his concept of reverse salient.[264] Again, this research is a bit of a departure from the sources previously discussed. Hughes was primarily interested in historical and sociological aspects of technological and infrastructural development. I draw upon it here because mass digitization may be fairly characterized in terms of a transition between analog and digital information environments. Indeed, the development of HathiTrust is, in many ways, an example of knowledge infrastructure development that draws upon and interoperates with centuries-old knowledge infrastructures while simultaneously carving out a new, different path that takes advantage of networked communication technologies, data mining and data analytics, distributed computer-supported collaborative work, and other emerging forms of knowledge technology and practices.

---

[263] Hirschman (1970:47).

[264] Hughes, T. P. (2012). The Evolution of Large Technical Systems, pp. 51-82 in Bijker, W. E., Hughes, T. P., Pinch, T., & Douglas, D. G. (2012). *The social construction of technological systems: New directions in the sociology and history of technology*. MIT press; Hughes, T. P. (1993). *Networks of power: electrification in Western society, 1880-1930*. JHU Press.

Hughes argued that the development of infrastructure is not as a grand scheme conceptualized from the outset. Instead, he stressed the local and entrepreneurial construction of systems that, over time, and were assembled into larger systems, networks, and Internetworks. We can apply these observations more generally to processes of sociotechnical transformation that shapes and is shaped by dynamic, entangled interactions of law, technological change, and emerging social practice.

Of particular note is Hughes' introduction of the concept of the "reverse salient." Borrowing from military terminology, a reserve salient is a part of a system that underperforms the rest and therefore prevents the system from achieving its full potential. A reverse salient can be technological, such as a capacitor in an electrical system, or social, such as founder's myopia that biases organizations against innovation. In this research, I would extend the concept to also include existing legal mechanisms and doctrines in copyright law that unnecessarily hamper the progress of society's useful arts and science. As a practical matter, reverse salients become focal points in processes of sociotechnical transformation. An obvious sticking point highlighted by a backward protrusion against forward momentum, reverse salients function as a convenient locale for orienting debates and reforms. Thus, in the context of mass digitization (and previous controversies involving technological change) debates around how and why copyright law must be reformed given the new technical realities can be understood as an effort to alleviate the copyright's effect as a reverse salient.

Distributed sources of transformation: A bottom-up approach

In seeking to describe, explain, and understand how individuals and organizations navigate the intersections of copyright law, technological change, and emerging social practice and influence processes of transformation, organizational science theory provides useful insights particularly with respect to sensemaking and decision-making. This section focuses primarily on the work of Karl Weick and colleagues before touching briefly on organization change literature.

As a preliminary matter, organizational science literature does not provide a specific, concrete, concise definition of "sensemaking." Somewhat like the sensemaking process itself, sensemaking theories seems to ask researchers to take on the role of bricoleur,[265] building

[265] As described by Strauss, C. L. (1962). *Savage mind*. University of Chicago at pg.11: "The 'bricoleur' has no precise equivalent in English." The term is often used to refer to a person who makes things using "whatever is at

interpretations and meanings over time through an iterative process of redrafting and synthesis subject to potentially endless revision. Preeminent sensemaking scholar Karl Weick analogizes sensemaking to cartography. Maps reflect where the cartographer looked, how they looked, what they wanted to represent, and their tools of representation; the same can be said of sensemaking. Moreover, as with sensemaking, there is no one best map of a particular terrain because, for any terrain, there will be an indefinite number of useful maps.

As a researcher engaged in my own sensemaking about the sensemaking of an organization, the irony of this endeavor does not escape me. Like the participants in my study, I too will cobble together a retrospective story that invariably privileges certain features and aspects of the story and leaves others hidden. But, as one of my committee members wisely counseled, in communicating research, the story is not going to come from *it*, it is going to come from me looking at *it* and saying *here is what I see*. In sensemaking, the "terrain keeps changing and the task is to carve out some momentary stability in this continuous flow."[266] With that said, I will attempt to sketch some of the features of sensemaking and decision-making literature that I have found most prominent, relevant, and useful in my research on sociotechnical transformation and, more specifically, to a case study of mass digitization and the emergence and evolution of HathiTrust.

Sensemaking is not a normative process about Truth and getting it right. Instead, it is "about continued redrafting of an emerging story so that it becomes more comprehensive, incorporates more of the observed data, it is more resilient in the face of criticism."[267] Sensemaking generates understanding that is "provisional, plausible, subject to revision, swift, directed toward continuation of interrupted activity, ready to hand, tentative, infused with ignorance, and sufficient for current purposes." Sensemaking starts out as "a momentary, expedient understanding. But the sense thus created often lingers and gets stored as if it were the product of a far more deliberate, intentional analysis."[268] Sensemaking by an organization is an instance of the organization giving order to flux:

---

hand" rather than tools or materials that are project specific. As Levi-Strauss stresses: "the 'bricoleur' is still someone who works with his hands and uses devious means compared to those of a craftsman."

[266] Weick, K. E. (2012). *Making sense of the organization, Volume 2: The impermanent organization* (Vol. 2). John Wiley & Sons.

[267] Weick, Sutcliffe, and Obstfeld (2005: 415).

[268] Weick, (2012:96).

> "sensemaking does not begin *de novo*, but like all organizing occurs amidst a stream of potential antecedents and consequences. … All of these activities furnish a raw flow of activity from which he may or may not extract certain cues for closer attention."[269]

Weick thus identifies sensemaking as the primary site where "meanings materialize that inform and constrain identity and action."[270]

The relationship between sensemaking and decision-making is complex and interdependent. Both are processes with no clear starting or ending point and each feeds into the other, creating swirling currents of meaning within a larger temporal stream. As a practical matter, this can lead to difficulties in terms of knowing where to jump in to an analysis or discussion. For purposes of this literature review and its subsequent application to the empirical study on mass digitization and the emergence and evolution of HathiTrust, I organize sensemaking and decision-making theories around three phases: 1) pre-decision processes involving choosing among alternatives; 2) the choice or act captured by a behavioral commitment, and 3) the post-decision process of sensemaking.

Phase One: Choosing Between Alternatives

Preceding a decision, or an "event of behavioral commitment," would-be decision-makers operate in a state of ambiguity or uncertainty. Jones and Gerard observe that this uncertainty is a motivating force in the decision-making process. In comparison to other disciplines (such as law) that tend to privilege rationality as a central tenet of good decisions, organizational science researchers tend to problematize the deliberation process by, for example, emphasizing tradeoffs between rationality and expediency in organizational decision-making.

An approach to decision-making that prioritizes rationality and a careful assessment of alternatives might result in the discovery of a better alternative. Reducing uncertainty and ambiguity enables organizations to register fine grain details of the environment and choose an action that is sensitive to potentially important aspects of that complicated environment. One potential downside to this approach is that dedicating significant resources to assessing alternatives might dissipate some of the energy that could be used to help implement that chosen action. Thus, in working to discover the best alternative, "people (lose) some of the commitments that they needed to put it into action." However, Weick notes that by more fully

[269] Weick, Sutcliffe, and Obstfeld (2005: 411).
[270] Weick, Sutcliffe, and Obstfeld (2005: 409).

understanding a problem, decision-makers are able to craft responses that consist of small actions, executed with relatively little commitment (and thus requiring less implementing energy), to produce big results.[271]

On the other hand, an organization might resolve ambiguity not through decision rationality but through vigorous action. By just jumping in rather than carefully assessing alternatives, organizations necessarily simplify the environment so that "relatively crude analyses are sufficient to keep track of the main things that are happening."[272] This approach favors expediency and simplicity over rationality. While it enables decision-makers to proceed with alacrity, the tradeoff of this approach is that it precludes decision-makers from "any chance of learning more about the situation as it originally existed." This potential drawback is mitigated by the fact that, in the view of sensemaking theorists, meaning and significance are largely socially constructed through retrospective processes and therefore getting a firm, detailed picture of the decision-making environment may be somewhat illusory.

How do organizations navigate these tradeoffs? How do organizations decide whether to engage in decision rationality or just jump out into the unknown? As a practical matter, researchers note organizations rarely (if ever) employ formal rationality in decision-making, beginning by perceiving a problem, defining the problem carefully, generating possible courses of action solely because they might solve the identified problem, and selecting an action solely because it ought to be the best way to solve the problem.[273] One explanation offered for this is that organizations simply lack the bandwidth to deliberate in this way. The problem of decision-making amongst alternatives is a problem of requisite variety because "no sensing device can control input that is more complicated than the sensing device."[274]

The implications of requisite variety suggest that a choice amongst alternatives will ultimately come down to the control of salience, deciding whether deliberation or action is more important, and deciding which factors or aspects of the environment should emphasized or foregrounded. The effects of this decision on how to approach decision-making are often irreversible. "If you choose in favor of accurate sensing, you reduce your capability to take strong action." Carried to the extreme, looking before one leaps may result in not seeing

[271] Weick (2012:50).
[272] Weick (2012: 50-51).
[273] Starbuck, W. H. (1982). Congealing oil: Inventing ideologies to justify acting ideologies out. *Journal of Management Studies,* 19, p. 16.
[274] Weick (2012:50).

anything, and therefore may result in inaction. However, if a problem is construed such that only a relatively small corrective action is required, then then detail and accuracy are crucial."[275] On the other hand, if a problem is construed on a grand scale, as many social movement-oriented problems are, then an action-oriented approach may be better suited. By jumping in and choosing a course of action before gathering and assessing all the alternatives, action is guaranteed to generate outcomes that ultimately provide the raw material for seeing something and generating momentum.[276] Through the mechanism of salience, organizational science theorists shift the problem of decision-making from one of *uncertainty* in the environment — a framework that would indicate that increased quantities of information may provide the source of a solution — to c*onfusion* in the environment — where the problem is really a criticality resulting from multiple meanings:

> "[W]hen a decision means many different conflicting things, the problem is one of too many meanings, not too few, and the problem shifts from one of uncertainty to one of criticality."[277]

Understanding how an organization navigates the tradeoffs between careful deliberation and vigorous action is therefore tied to the question of how an organization controls salience regarding the problem.

Rationality is not abandoned by the organization that chooses efficient action over decision rationality, it simply gets shifted to a subsequent phase of the process. Organizational science researchers adopt multiple views of rationality based on its various functions. Rationality can include: 1) a set of prescriptions that change as the issue changes; 2) a façade created to attract resources and legitimacy; and/or 3) a post-action process used retrospectively to invent reasons for the action."[278] When an organization adopts a decision-making approach that privileges vigorous action over careful deliberation, rationality shifts from an evaluative function to a form of rationalization, a process of retrospective justification whereby decision-makers seek to resolve multiple, conflicting definitions of what their decision means.

---

[275] Weick (2012:50).
[276] Weick (2012: 53).
[277] Weick (2012:76); Daft, R. L., & Macintosh, N. B. (1981). A tentative exploration into the amount and equivocality of information processing in organizational work units. *Administrative science quarterly*, 207-224.
[278] Weick (2012:35); Pfeffer, J. (1981). *Management as symbolic action: The creation and maintenance of organizational paradigms*. Graduate School of Business, Stanford University, pp. 194-196.

Before moving on to the retrospective justification phase, we must discuss the intermediate phase between choosing amongst alternatives and sensemaking. This is the phase where the actual act of decision-making happens, what Weick calls the "behavioral commitment" or the event which triggers the sensemaking process.

Phase Two: Behavioral Commitment

In sensemaking theory, a behavioral commitment is the trigger moment which unleashes the forces that destroy the plausibility of alternatives and remove their ability to inhibit action. As Weick notes, although rationality may play an initial role in our decision-making process, what drives, energizes, and justifies our choice is "the nonrational basis of our motivational commitment to them."[279] Thus, after the committed action has been chosen, there is little advantage to reflecting on the advantages of the rejected alternative or disadvantages of the chosen alternative. "Once a decision is made, action is more effective when probabilistic information is treated as if it were deterministic and beliefs that are only relatively true are treated as if they were absolutely true."[280] Behavioral commitment is therefore the stimulus around which people and organizations build coherent worldviews through the process of sensemaking.

The binding strength of any action depends on three conditions: 1) choice, 2) irrevocability, and 3) visibility. When a behavior involves a high degree of choice, is undeniable, and is irreversible, it stimulates a retrospective process of justification "in which behavior is rationalized by referring to features of the environment which support it."[281] Sensemaking is thus a matter of articulating a rationale, or providing a justification, for a binding behavioral commitment.

Phase Three: Sensemaking — Retrospect and Social Justification

The importance of retrospect to the sensemaking process cannot be understated. Justification and rationalization are fundamentally post-hoc devices.[282] "Societal ideologies insist that actions ought to be responses – actions taken unreflexively without specific reasons

[279] Brickman, P. (1987) *Commitment, conflict, and caring.* Englewood Cliffs, NJ: Prentice Hall, pp. 40-41.
[280] Weick (1995:25).
[281] Weick (1995:11-12).
[282] Staw, B. M. (1980). Rationality and justification in organizational life. *Research in organizational behavior*, *2*, 45-80.

are irrational and irrationality is bad."[283] Therefore, organizations point to information in their environment, such as problems, threats, success or opportunities, to justify their action.

Post-hoc rationalizations tend to underestimate the vast amount of uncertainty present during the early stages of acting, because people overestimate the predictability of past events once they know how they turned out.[284] This is essentially the issue of hindsight bias.[285] Organizational science researchers argue that hindsight bias most often implicates assumptions about reality (that what appears to happen did happen), intention (that what happened was intended to happen), and necessity (that what happened had to happen).[286] As discussed in the following chapter on Methods, hindsight bias may have some implications for how research into organizational sensemaking is conducted, but poses relatively fewer problems for constructivist research that triangulates, corroborates, and anchors the accounts of participants.

In addition to being an essentially retrospective process, sensemaking is also fundamentally social. This is true in a number of ways. First, retrospective justification is often done for external consumption, to attract resources and legitimacy, for example. The strength of the behavioral commitment to trigger subsequent justification depends in large part on the publicness of the act. When an action is irrevocable, public, and volitional, there is more at stake for the actor; the role of justification is intensified by the social context in which it occurs. The actor seeks socially acceptable justifications for the lapsed action. Shared norms and expectations affect the rationalizations developed for behavior as a process of legitimation. "People develop acceptable justifications for their behavior as a way of making such behavior meaningful and explainable." [287] As a consequence, explanations that are developed retrospectively to justify committed actions are often stronger, more tenacious, and more likely to produce selective attention and confident action.

The second way in which retrospective justifications are social is that they serve important needs internal to the organization for understanding and predicting future events.[288]

---

[283] Starbuck (1982: 94).
[284] Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human perception and performance*, *1*(3), 288.
; Fischhoff, B., & Beyth, R. (1975). I knew it would happen: Remembered probabilities of once—future things. *Organizational Behavior and Human Performance*, *13*(1), 1-16; Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual review of psychology*, *28*(1), 1-39.
[285] Weick (1995:36).
[286] Weick (1995:37).
[287] Weick (1995:11-12).
[288] Weick (1995:36).

Tenacious justifications tend to prefigure subsequent perception and action, which means they often become self-confirming, verging on self-fulfilling prophecies.[289]

The third way in which sensemaking is essentially social is that behavioral commitments are often justified by explanations that reify social structure. In particular, Weick identifies five ways in sensemaking tends to link micro-behavioral commitments to macro social consequences: 1) they begin as commitments to social relationships rather than commitments to individual behaviors, 2) the social relationships often generate their own conditions of commitment, 3) since social relationships rather than behaviors are what people become bound to, justifications tend to invoke social entities rather than individual reasons, 4) reifications that justify social commitment tend to set up expectations that operate like self-fulfilling prophecies, and 5) efforts to validate the social justifications tend to spread them to other actors.[290]

Put in more explicit terms, sensemaking is a process through which collectivity may be generated. This occurs when behavioral commitments, typically double interacts, are justified through shared committed interpretations. Commitment occurs as a result of what Weick calls a double interact represented as $A_1B_1A_2$. The double interact is also the smallest unit building block of collective structures which Weick argues form when "self-sufficiency proves problematic."[291] Thus, collective structures are built initially around interdependent means: "Neither A nor B has direct control over their outcomes, and they must entice someone else to contribute a means activity to get their own desired outcomes."[292] Furthermore, "people commit to and coordinate instrumental acts (means) before they worry about shared goals."[293] "But shared goals do emerge as people search for reasons that justify the earlier interdependent means to which they have become bound."[294]

Thus, agreements to agree introduce stability into an equivocal flow of events by means of justifications that increase social order and collectivity:

> "When people act on behalf of these committed interpretations and their reified contents, their actions become more orderly, more predictable, more organized. As a result of this tightening, their actions have more impact on others and are more likely to be imitated. Thus, both the form and substance of organization

[289] Weick (1995:13-14.
[290] Weick (1995: 15).
[291] Weick (1995:17).
[292] Weick (1995:17).
[293] Weick, K. E. (1979). Cognitive processes in organizations. *Research in organizational behavior*, *1*(1), 41-74. p. 91-95.
[294] Weick (1995:17).

> become more distinct and the world momentarily becomes slightly less chaotic. And all because some action first stuck out as more public, more irrevocable, and more attached to a set of actors than were other actions."[295]

Reification of a collectivity thus becomes a mechanism for justifying commitment. "Having become bound to interdependent action, a person might invoke macro sources of micro-constraints as in, for example, "that's the way we do things in this culture."[296]

The relationship between action, sensemaking, and reality are interdependent, circular, and often murky. Organizational science literature on sensemaking and decision-making seems to support the contention that "we need only in cold blood act as if the thing in question were real, and keep acting as if it were real, and it will infallibly and by growing and such a connection with our life but it will become real."[297] And, "if men define situations as real, they are real in their consequences."[298] When considering sensemaking around mass digitization and the subsequent emergence and evolution of HathiTrust, it makes sense to bear in mind how retrospective justification and reification might operate in combination with structural and social-psychological factors to configure sociotechnical transformation.

Moreover, while the process of sensemaking involves retrospective, socially attuned justification, "that action initially explained by reification soon generates the reality that replaces the reification with substance."[299] Once a justification begins to form, "it exerts an effect on subsequent action"[300] and can ultimately turn into preferences that "control subsequent attention and action."[301] Thus, sensemaking processes can form tighter entanglements among aspects of copyright, technological change, and social practice effecting behavioral commitments and salience around future related issues. In that way, while past action and sensemaking are not necessarily predictive, they more forge new "go to" pathways which can have a tendency to channel behavior. Whether one calls it a self-fulfilling prophecy, drawing upon the work of Weick and colleagues, or gives it a different name, identifying and investigating the relationship between innovation and a persistent, almost gravitational pull toward tradition, is a vital area of

---

[295] Weick (1995:16).
[296] Weick (1995:19).
[297] James, W. (1890). *Principles of Psychology, Vol 2. 1950 Reprint*. p. 321.
[298] Thomas, D. S. (1928). The Child in America: Behavior Problems and Programs, cited in Weick (1995:12).
[299] Weick (1995:23).
[300] Weick (1995:23); Penner, D. D., Fitch, G., & Weick, K. E. (1966). Dissonance and the revision of choice criteria. *Journal of personality and social psychology*, *3*(6), 701; Weick, K. E., & Prestholdt, P. (1968). Realignment of discrepant reinforcement value. *Journal of personality and social psychology*, *8*(2p1), 180.
[301] Penner et al (1966).

exploration in the story of HathiTrust and the context of sociotechnical transformation more broadly.

Possibly related to this, organizational science scholars identify a number of factors that can interrupt or otherwise create difficulties for sensemaking. One such factor is technological change, particularly where emerging technologies disrupt normal expectations "and hence, the efficacy of established patterns of meaning and associated behavior."[302] Weick notes that as newer information technologies become more stochastic, continuous, and abstract, more of their operation has become a problem for sensemaking with respect to how a technology works and its organizational impacts.[303] When faced with technological change that disrupts organizational order or increases flux within the organization, individuals will draw creatively on their memory — especially their personal experience — in composing a story that begins to make sense of what is happening. The transition of research libraries from a purely analog world, grounded in centuries-old practice and traditions, to a hybrid analog-digital one raises complex questions and challenges for how institutional memory intersects with technological change in the context of sensemaking and decision-making around things like mass digitization.

When technological change produces ambiguities in the environment and its "products are value laden, … commitment to goals and procedures, whatever they are, may be sufficient for proper adjustment to the environment. At the extreme, a school of thought may be created as in university life where successful organizational leaders are those who can convince others that their own commitments are the standard to be achieved."[304] This might suggest that a return to the foundational goals or missions of libraries and universities (or perhaps the institution of copyright law also) as the motivation for behavioral commitments and decisions may take primacy over past practices, processes, mechanisms, and means for carrying out those objectives which, in light of technological change, no longer hold the same relevance.

Technological change enables the introduction and development of new knowledge ecologies and alters the context in which existing institutional structures and patterns of practices operate. Sometimes, the changes reveal tremendous promise with respect to the development and introduction of new, adaptive, continuous methods for creating, debating, and spreading

[302] Coopey, J., Keegan, O., & Emler, N. (1997). Managers' Innovations as 'Sense-making'. *British Journal of Management*, *8*(4), 301-315: 312.
[303] Weick (1995:98-99).
[304] Staw (1982).

knowledge. But knowledge infrastructure development, and efforts to transition the analog world to an uncertain digital one, faces numerous challenges. The history of large-scale digitization offers relevant touchstones for exploring, understanding, and problematizing sensemaking around the mass digitization project.

Researchers note that one common source of failure or obsolescence is the failure to account for the non-technical aspects of infrastructure development. When a significant technical change is introduced it can put a tremendous strain on existing institutions (e.g. copyright law, universities, and libraries) that may be prone to inertia.[305] When new technologies fail to account for key features of the social environments in which they operate, an otherwise promising technology can end up having a disastrous and deleterious effect.[306] As Erik van der Vleuten observed, "technical infrastructure elements are increasingly intertwined with non-technical ones" such that they two systems can be self-reinforcing and similarly generate their own gravitational pull. He explains:

> "In the establishment phase(s) technical designs are adapted and coupled to an actor playing field, organization structures, marketing strategies, legal frameworks, etc.; in the expansion phases such sociotechnical intertwinement is further strengthened to the degree that technical and non-technical elements interlock and make the whole thing difficult to change."[307]

While issues at the intersection of sensemaking, technological change, and STS have been explored by Weick, van der Vleuten, Edwards and others, a related body of literature on organizational change offers additional insights. Organization change theory is concerned with how organizations construe, interpret, and respond to events that disrupt existing patterns and expectations. One important, however obvious, difference between the two approaches is that organizational change focuses on the change process, rather than the sensemaking process. Literature on organizational change is concerned with the process by which uncertainty and conflict are accepted by the organization and transformed through confirmation so that novelty and innovation, eventually, becomes routinized. [308] This research therefore has potential bearing, in particular, on HathiTrust's evolution from an emerging upstart to a more seasoned and entrenched organization.

---

[305] Edwards, P. N., Jackson, S. J., Chalmers, M. K., Bowker, G. C., Borgman, C. L., Ribes, D., ... & Calvert, S. (2013). Knowledge infrastructures: Intellectual frameworks and research challenges, pg. 5.
[306] E.g. Rob Kling, *Computerization and Controversy: Value conflicts and social choices*
[307] Vleuten, E. V. D. (2006). Patterns and principles in infrastructure development. Position paper, September 2006.
[308] Kimberly, J. R., & Quinn, R. E. (1984). *Managing organizational transitions*. McGraw-Hill/Irwin, pg.303.

Lynn Isabella's research draws upon previous research on the role of cognition and interpretation in organizational change and proposes a model for the change process.[309] She posits that the sensemaking process turns into organizational change in four key stages. The first stage is *anticipation* — when an organization's key members "assemble an interpretive portrait based on speculation and anticipation" by piecing together ill-fitting information into a "coherent and cogent frame of reference."[310] The second stage is *confirmation* — where the interpretation of the event is standardized: "Interpretations at this stage provide no new or creative insights and primarily reflect understandings that worked or are believed to have worked in the past – presumptions about what will be, based upon what has been."[311] The third stage is *culmination* — when an organization amends its view of an event:

> "Interpretations no longer represent standard or presumed use but reconstructed use, frames of reference that are being amended at the event occurs to include new information or omit information no longer of value. … A real hands-on sense of experimentation and testing and learning by doing characterize collective interpretations at this stage. Since each event brings with it the need to create new norms and execute new behaviors, old views may not be effective."[312]

The final stage is the *aftermath* during which an event is evaluated. As organizations "test and experiment with a construed reality that moves beyond the traditional boundaries of past sense making, there comes a growing, concrete realization of the permanent changes wrought and of the consequences those changes and the event itself have had for the organization and its members." In this stage, the organization evaluates its interpretation of the event in terms of its consequences and actively seeks out the strengths and weaknesses of changes brought on by the event in an attempt to (hopefully) rebuild certainty in the system.[313]

Approaching organizational change and processes of transformation with these stages in mind may help researchers guard against a tendency to follow the "new, fast-moving, and dramatic parts of the current transition." Chasing the rabbit leaves the "complex processes of mutual adjustment by which older knowledge institutions adapt to emergent ones, and vice

---

[309] Isabella, L. A. (1990). Evolving interpretations as a change unfolds: How managers construe key organizational events. *Academy of Management journal*, *33*(1), 7-41; Bartunek, J. M., & Moch, M. K. (1987). First-order, second-order, and third-order change and organization development interventions: A cognitive approach. *The Journal of Applied Behavioral Science*, *23*(4), 483-500; Van de Ven, A. H., & Ferry, D. L. (1980). *Measuring and assessing organizations*. John Wiley & Sons.
[310] Isabella (1990:17).
[311] Isabella (1990:17).
[312] Isabella (1990:23).
[313] Isabella (1990:25).

versa" under-examined.[314] In addition, it ignores the tensions that arise as boundaries between co-existing educational institutions (such as research universities) and memory institutions (such as libraries and archives) become increasingly blurred, reconfigured, and reconstituted during the process of transformation.[315] This is particularly troubling when, as is frequently the case, "[t]he institutions in which most knowledge workers live and labor have not kept pace, or have done so piecemeal, without a long-term vision or a strategy."[316] Educational institutions, libraries, publishers, intellectual property regimes, and political mechanisms have struggled to adapt to or accommodate the changing information environment."[317] Programmatic efforts to improve knowledge infrastructures frequently prioritize "investments in technical systems over research on how to effect equally crucial cultural, social, and organizational transformations."[318] These observations suggest that particular gaps exist, and continue to grow, in our understandings of the sociotechnical aspects of transformation in the context of knowledge infrastructure development in which mass digitization and HathiTrust may play important roles.

Research into processes of transformation in the context of knowledge and memory organizations and infrastructures must necessarily display a willingness to engage the complex, often messy intersections of law, technology, and social practice over time. There are, therefore, inherent challenges in studying this phenomenon but the challenges are mitigated by the enormous potentials of generating a richer, deeper understanding of the processes of transformation as they unfold:

> "This is an exhilarating possibility. Imagine what might have happened if scholars of the 15th and 16th centuries could have experimented directly with the sociotechnical reconfigurations that accompanied the advent of the printing press – as we can do today."[319]

Of course, we cannot turn back the clock and experiment directly with the transformative impact of the printing press, but we can open ourselves to the lessons that history might teach and use those lessons to generate new insights into things like mass digitization and the emergence and evolution of HathiTrust.

---

[314] Edwards et al (2013: 3).
[315] Edwards et al (2013:6).
[316] Edwards et al (2013:2).
[317] Borgman, C. L., Wallis, J. C., Mayernik, M. S., & Pepe, A. (2007, June). Drowning in data: digital library architecture to support scientific use of embedded sensor networks. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries* (pp. 269-277). ACM.
[318] Edwards et al (2013:12).
[319] Edwards et al (2013:19).

Chapter IV: Methods

Qualitative Case Study Approach

The primary goal of this research is to develop a richer understanding of processes of sociotechnical transformation arising from the intersections of copyright law, technological change, and emerging social practices. As this is a rather broad field of inquiry, I have narrowed the analytic aperture to focus on a single case: the mass digitization project and subsequent emergence and evolution of HathiTrust. Through the story of HathiTrust, this research hopes to contribute to understandings of transformation, delving into the ways copyright law, technology, and social practices co-evolve through dynamic, often murky, processes of entanglement, reconfiguration, and mutual-readjustment.

The focus of this research is necessarily interdisciplinary. As the previous chapters demonstrate, insights abound from multiple domains including law, history, sociology, economics, and organizational science (in addition to the countless other perspectives not discussed). The analytic framework for this research is based upon the theory of innovative deviance expanded through sensemaking theory and traditional legal analysis. Synthesizing approaches from various domains to describe and explain a complex case requires analytic flexibility and sensitivity to multiple, varied perspectives. The most appropriate method for studying transformation in the context of mass digitization and HathiTrust is a qualitative case study approach informed by a constructivist paradigm as this facilitates the collection of information-rich data supporting thick description and the emergence of potentially deeper insights and understandings of a complex multidimensional phenomenon.[320]

Qualitative interview studies enable researchers to develop detailed holistic descriptions, integrate multiple perspectives, learn how events are interpreted, describe process, and bridge

[320] Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications; Geertz, C. (1994). Thick description: Toward an interpretive theory of culture. *Readings in the philosophy of social science*, 213-231.

inter-subjectivities.[321] A qualitative approach is particularly useful when studying organizations and institutions as these units of analysis "lend themselves to multiple, conflicting interpretations, all of which are plausible."[322] As Weick describes, to reduce the criticality of information, researchers do not need larger quantities of information but rather richer qualitative information:

> "Information richness is defined as the ability of information to change understanding within a time interval. Communication transactions that can overcome different frames of reference or clarify ambiguous decisions to change understanding in a timely manner are considered rich. … In a sense, richness pertains to the learning capacity of a communication."[323]

In seeking to generate rich information, qualitative research privileges the perceptions of local actors 'from the inside' and the meanings they ascribe to their activities with the goal of "explicat[ing] the ways people in particular settings come to understand, account for, take action, and otherwise manage their day-to-day situations."[324]

Furthermore, a constructivist approach has been adopted by leading case study methodologists,[325] researchers studying organizational sensemaking, and sociologists who have adopted social disorganization, social control, structural strain, and labelling approaches to the study of deviance.[326] Constructivism stresses the importance of the subjective creation of meaning, encouraging participants to describe their views of reality which, in turn, enables the research to better understand the participants' conceptions and practices.[327] As sensemaking theorists explain:

> "Individuals are not seen as living in, and acting out their lives in relation to a wider reality so much as creating and sustaining images of a wider reality, in part to rationalize what they're doing."[328]

---

[321] Weiss, R. S. (1995). *Learning from strangers: The art and method of qualitative interview studies*. Simon and Schuster, 9-10.
[322] Daft and Macintosh (1981:9).
[323] Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management science*, *32*(5), 554-571, at pg. 560.
[324] Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage, pp 6-7.
[325] E.g. Stake, R. E. (1995). *The art of case study research*. Sage; Yin, R. (2003). K. (2003). Case study research: Design and methods. *Sage Publications, Inc*, *5*, 11.
[326] Pfohl (1994: 348).
[327] Robottom, I. M., & Hart, E. P. (1993). *Research in environmental education: Engaging the debate*. Deakin University.
[328] Weick (1995: 11).

Sensemaking literature further suggests three premises for constructivism: 1) the identification of any given action is subject to internal revision, 2) the anchor point for any given identification relies on a network of interdependent and continuously modifiable interpretations, and 3) any given action is subject to multiple identifications, the relative superiority of which is problematic.[329] Constructivist research therefore is not motivated by a desire to discover Truth and facts but is rather motivated by *Verstehen* — a desire to understand social phenomena, namely the processes by which meanings are created and revised, in relation to internal and external events in a given context.

Retrospect and the documentary method are essential aspects of the constructivist approach adopted in this study. As sensemaking literature reminds us, "remembering and looking back are our primary source of meaning,"[330] and thus retrospect and the documentary method are emphasized as the processes by which decisions are interpreted, justified, and understood:

> "We take seriously people's accounts of how they accomplish interpretation, mindful however, that retrospective sensemaking involve bias reconstruction of antecedents since outcomes are known at the time reconstruction occurs. This very bias in the strength of retrospect as a method of sensemaking since it edits out false starts and imposes a spurious order on an indeterminate past. But the same editing requires that investigators observe sensemaking as it unfolds if they wish to contract this bias, which often means that ethnography and use of personal experience are crucial sources of data about interpretation."[331]

Even in the case of research on institutions — including legal institutions discussed primarily in Chapter VIII's analysis of the copyright litigation involving HathiTrust and, more generally, academic institutions such as universities, libraries, and archives — retrospect and the documentary method offer approaches that are responsive to the scale, scope and rhythms of the changes being studied.[332] A sensitivity to historical context and retrospect teaches that what might otherwise appear as necessary and obvious features of a social, technological, or legal system might in fact be "historical creations which could have followed other paths."[333] By organizing the scope of this study to include the approximately twelve years from the outset of

[329] Gergen, K. J. (1982). From self to science: What is there to know. *Psychological perspectives on the self, 1*, 129-149; Weick (1995:10).
[330] Weick (1995:11).
[331] Weick (1995:12).
[332] Edwards et at (2013: 20).
[333] Edwards et al (2013: 19).

the mass digitization project to the present day, this research attempts to study the phenomenon of transformation through the lens of what Stewart Brand called "the long now."[334] Transformation and the development of things like HathiTrust take time to emerge, develop, evolve and, sometimes, obsolesce as do the meanings we generate about them.

Studies of technological change similarly highlight the importance of the long now. As Paul David's work, and in particular his notion of the "productivity paradox," describes, technological "advancements" may, at least initially, undermine productivity before enhancing it. For example, David observed that, after computers were introduced to the workplace, it took about twenty years of social, cultural, organizational, and institutional adjustments before they improved, rather than undermined, productivity.[335] Similarly, since the advent of the Internet, we are "changing our knowledge generation and expression procedures root and branch."[336] Change does not happen overnight. In the grand scheme of things, the ripple effects of mass digitization and HathiTrust are still fairly close to their epicenter. Despite the emergence of new information and communication technologies and new organizational and infrastructural forms:

> "We remain bound to the book and article format and to the classic nineteenth-century technology of files and folders. It took well over 200 years for printed books to acquire the intellectual armature we now consider intuitive (such as the index, table of contents, bibliography, footnotes, and generally agreed rules on plagiarism). Even page numbers were once an innovation."[337]

A case study approach offers distinct advantages in the context of constructivist research into complex, multi-faceted subjects because the approach emphasizes "contemporary phenomenon within its real-life context."[338] In addition, by facilitating exploration of a phenomenon within its context using a variety of data sources, this approach "ensures that the

[334] Brand, S. (1999). The Clock of the Long Now. *London: VVeidenfeld and Nicolson*
; Bowker, G. C., Edwards, P. N., Jackson, S. J., & Knobel, C. P. (2010). 1.1 The Long Now of Cyberinfrastructure. *World Wide Research: Reshaping the Sciences and Humanities*, 40.
; Ribes, D., & Finholt, T. A. (2009). The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Information Systems*, *10*(5), 375.
[335] David, P. A. (1990). The dynamo and the computer: an historical perspective on the modern productivity paradox. *The American Economic Review*, *80*(2), 355-361; Landauer, T. K. (1995). *The trouble with computers: Usefulness, usability, and productivity* (Vol. 21). Cambridge, MA: MIT press.
[336] Edwards et al (2013).
[337] Edwards et al (2013:21).
[338] Yin, R. (1994). Case study research: Design and methods. Beverly Hills, p. 13.

issue is not explored through one lens, but rather a variety of lenses which allows for multiple facets of the phenomenon to be revealed and understood."[339]

While not limited to qualitative data, case study methods generally include interviews, observations, and documentary analysis. These multiple sources of evidence allow the researcher to triangulate data, thus corroborating phenomena.[340] Case studies are particularly adept at enabling researchers to answer "how" and "why" questions when the behaviors or phenomena involved in the study are not amenable to researcher manipulation. In addition, case studies allow researchers to delve into contextual conditions that may prove relevant to the phenomenon under study or, framed a slightly different way, enable researchers to study a phenomenon when the boundaries between phenomenon and context are not clear.[341]

After noting the merits, benefits, and advantages of a constructivist case study approach, it is worth pausing to reflect on what this method is not. Conclusions drawn from case studies are not intended to be generalizable to larger populations. Instead, they aim for theoretic generalizability (Yin, 2009) and transferability. Findings could, as a result, be applied to cases with similar characteristics along the theoretical lines examined and uncovered in the study. In addition, the proposed study of HathiTrust's mass digitization project could therefore provide useful touchstones for other digitization efforts, offer probative insights for the larger copyright policy debates around innovative deviance and technological change, and provide clues about processes of sociotechnical transformation.

In addition, as already mentioned, this study is not motivated by a desire to find Truth, but rather to generate a rich description and some potential explanation for how and why HathiTrust and its progenitors navigated the legal, technological, and social aspects of mass digitizing in-copyright works. The story which unfolds from the empirical work and is discussed in the following chapters aims for *Verstehen*. I therefore refrain from interrogating or evaluating the veracity or legitimacy of the interpretations of interview participants. That is not to say that I back away from or deemphasize conflicting or multiple perspectives. I simply do not use differentiation as a springboard for generating broader normative claims.

---

[339] Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The qualitative report*, *13*(4), 544-559.
[340] Yin, R. K. (2009). Case study research: Design and methods, 4th. *Thousand Oaks*.
[341] Baxter & Jack (2008).

Case Selection Strategy

Prior to embarking on this study, I had been interested in the relationship between innovation and deviance. In particular, I was fascinated by how we make sense of and respond to behaviors that seem simultaneously socially productive or beneficial and potentially transgressive or infringing. These early explorations led me to a theoretical framework for studying the phenomenon based, in large part, on Merton's structural strain theory. As a result of conversations regarding these early manifestations, culminating in the dissertation proposal defense, I heeded the advice of my committee to narrow the scope of my exploration into innovative deviance to a single case, fairly quickly settling on the mass digitization project and the subsequent emergence and evolution of HathiTrust. By binding the case in this way, I am able to research a complex phenomenon while maintaining a reasonable scope.

Stake recommends that cases should be selected for their anticipated ability to help the researcher better understand the phenomenon in question (Stake, 2006). He presents three main criteria for selecting cases: 1) the case should be relevant to the phenomena or thing studied; 2) the case should be representative of diversity across contexts; and 3) the case should provide opportunities to learn about complexity and contexts.

Mass digitization and HathiTrust are important and timely examples of a particular set of controversies arising at the intersection of copyright, technological change, and emerging social practice. This example provides a glimpse into processes of sociotechnical transformation and enables the synthesis and application of a new analytic approach combining innovative deviance and sensemaking theories. The mass digitization project and HathiTrust have drawn the attention, ire, and/or admiration of members of academia, industry, the government, the judiciary, as well as diverse set of stakeholder advocacy groups. Furthermore, because the mass digitization project involves numerous partners, and HathiTrust now consists of over 100 partners, this case enables exploration of embedded diversity (within, between, and across partners) in the context of the Google project and HathiTrust as well as in comparison to other parallel and/or competing large-scale digitization projects (such as those spearheaded by the Internet Archive and Open Content Alliance).

Research Questions

This study seeks to answer the following three research questions:

- RQ1: How and why did the University of Michigan engage in mass digitization of in-copyright works and how was its sensemaking and decision-making reflected in intra-organizational practices, processes, mechanisms, policies, and tensions?
- RQ2: How and why did these (conceptions, decisions, practices, processes, mechanisms, policies, and tensions) prompt the genesis of HathiTrust and how have they continued to evolve over time and in response to internal and external factors?
- RQ3: How might HathiTrust's emergence and evolution deepen understandings of processes of sociotechnical transformation and inform copyright policy debates around technological change?

Data Sources and Analysis

The primary data for the study were derived from semi-structured interviews and documents. Participants were selected using purposeful sampling, supplemented by a snowball recruiting strategy, focusing on current and past employees and advisors of HathiTrust partner institutions, participants in competing and/or parallel digitization efforts, and critics of mass digitization and/or HathiTrust. I made use of existing networks to identify and recruit key informants. I also sought out marginal, retired, and possibly disaffected figures as well. Additional participants with special knowledge of the legal and policy implications of mass digitization were also recruited.

Recruitment took place via in-person meetings as well as email solicitations. Interviews were conducted in person, over the phone, and using Skype (voice only) at a time and length convenient to my participants. The interviews typically lasted one hour and many participants were interviewed multiple times. All interviews were recorded with the consent of participants on an HTC Inspire using the Smart Voice application or, if conducted via Skype, using Amolto Call Recorder software. Interview recordings were transcribed by the researcher or Scribie, a transcription service provider. A total of thirty-one individuals were interviewed, representing twelve different institutional/organizational affiliations.

While the identities of many of the figures associated with HathiTrust and other large-scale digitization efforts are a matter of public record, I have taken care to anonymize the responses of my participants to the greatest extent possible. In particular, I have opted to associate quotes, statements, and perspectives with somewhat vague reference to the participant's

role, such as, for example, "lead architect," "HathiTrust member," "librarian," "technologist," and "legal commentator." In addition, I have made every attempt to make direct quotes non-attributable and general statements a synthesis of multiple responses. That said, readers may be familiar enough with HathiTrust and its members to surmise sources of content. The informed consent document used for this study is attached as Appendix A. The University of Michigan Behavioral Sciences Institutional Review Board reviewed this study and determined that it is exempt from IRB oversight.

In addition to interviews, documents provided another key source of data for this study. Documents analyzed for this study included publicly available artifacts (HathiTrust's website and web archive including press releases, statistics, and other items; court filings, orders, briefs, and opinions; agreements and contracts made available through freedom of information laws; articles, reports, and public statements contained in news articles and published presentations) and private intra-organizational communications (memos, emails, meeting notes) where provided by participants.

Data analysis, including coding, began with the first data collected and ran throughout the project. I extracted themes from my data by coding passages relevant to my research questions; the themes were then organized into categories. Early analysis assisted me in refining and formulating subsequent interview questions and in testing the theory and narrative through a process of member checking as these elements emerged. Interview transcriptions, notes, and documentary data were analyzed using an iterative, inductive, open coding process combined with qualitative memoing to identify important themes and concepts that emerge from the data.[342] Coding and cross-coding of interview transcripts and other artifacts were done by hand.

The story of HathiTrust called for data analysis of embedded units (distinguishing based upon institutions, participant roles, public vs. private status, academic vs. nonacademic and so forth) situated within the larger case. The data was therefore analyzed *within* the subunits separately, *between* the subunits, and *across* all of the subunits to offer a richer and more fully illuminated description and explanation of the phenomenon in its organizational context.[343]

Given the relevance of theme- and pattern-based elements as well as temporal elements in the study of knowledge infrastructure development, findings from the study have been

[342] Miles & Huberman (1994).
[343] Baxter & Jack (2008); Yin (2003).

organized as a diachronic story tracing the development and evolution of HathiTrust from its pre-origins in the mass digitization project through the present day, punctuated by synchronic flourishes tying the narrative back to themes and theoretical frameworks discussed in the literature review.

Limitations

A case study approach to research, and single-case studies in particular, are often criticized for their lack of causal generalizability to larger populations via repeatability or replicability. However, where the purpose of the research is to describe the experience of an individual or organization, develop interpretations or explanations of that experience, or develop courses of action and to make decisions appropriate for this particular individual or organization, then the case study method is an extremely useful one. Thus, while case studies do not offer statistical generalizability, when applied successfully they do enable analytic generalizability,[344] a rich understanding of the relevant contextual factors involved in the phenomenon of interest, and transferability. As already discussed, relevant literature and research from related fields (STS, law, sociology, and organizational science) promote constructivist case studies as ideally suited to my focus of inquiry. Findings from this study are not intended to provide replicable or predictive capacity but are rather intended to describe and explain, in a deeply contextual sense, the processes of sociotechnical transformation arising from the co-evolution of copyright laws, technologies, and emerging social practices shaping and shaped by mass digitization and the development of HathiTrust.

Qualitative studies have also occasioned criticism for their perceived lack of rigor or trustworthiness. Scholars have written extensively on both the challenges of qualitative research design and strategies for establishing credibility, transferability, dependability, confirmability, and trustworthiness.[345] General guidelines for critically appraising qualitative research have also been published.[346] This research employed design principles intended to diminish issues related

[344] Yin (1994).
[345] Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *ECTJ*, *29*(2), 75-91; Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry* (Vol. 75). Sage; Lindlof, T. R., & Taylor, B. C. (2002). Asking, listening, and telling. *Qualitative communication research methods*, 170-208.
[346] Forchuk, C., & Roberts, J. (1993). How to critique qualitative research articles. *Canadian Journal of Nursing Research*, *25*, 47-47; Mays, N., & Pope, C. (2000). Assessing quality in qualitative research. *British medical journal*, *320*(7226), 50.

to lack of rigor, credibility, or trustworthiness by including numerous and varied data sources which enabled me to triangulate between data sources, explore the phenomenon from multiple perspectives, and anchor findings. In addition, I used principles of idea convergence, confirmation of findings, and a process of member checking to ensure that my data was of a high quality.

Another potential limitation is one that is common to many empirical studies: missing data. When working with historical and retrospective data, there is some inherent difficult in figuring out the extent to which artifacts and pieces of data are missing. The problem of "unknown unknowns" is an unavoidable risk in every study. The more troubling issue with respect to missing data, however, is the "known unknowns."

This study was conducted while the copyright infringement litigation involving HathiTrust was active. As a result, I faced a challenge in that some of my potential participants were reluctant to be interviewed due to potential risks associated with the pending litigation. While reluctant and absentee participants formed the distinct minority, the possibility cannot be ignored that potentially important perspectives will be inadvertently left out of this analysis. I attempted to ameliorate this issue, wherever possible, by triangulating existing participants' accounts against publicly available documents to fill in those gaps.

In addition, as a researcher with institutional ties to HathiTrust, and a professional and sometimes personal relationship with some of my research participants, I must acknowledge the risks associated with my own close proximity to the focus of my research. As I have proceeded with this research, I have taken opportunities to pause and reflect on whether or not I might be unduly privileging a perspective or occluding a perspective, whether I am dispassionate and disinterested enough to represent the understandings of my participants without clouding them with my own biases, whether I am confident enough to tell the story I see emerging from the data without feeling constrained by my relationships with some of my participants (some of whom are also members of my dissertation committee). I have attempted to mitigate these risks by seeking guidance and mentorship from seasoned researchers who are both familiar with my work and personality and also completely without any interest in HathiTrust or the University of Michigan. In addition, I have used strategies of member checking, idea convergence, and confirmation of findings to avoid biases and/or selective inattendance that might otherwise problematize the findings and conclusions of this research.

Conclusion

In this chapter I detailed the features of qualitative case study research approaches conducted under a constructivist paradigm, explained my path to the HathiTrust case, and detailed sources of data, methods of analysis, and limits to this study. I conducted a total of 42 interviews with individuals associated with HathiTrust, parallel or competing digitization projects, critics and detractors, and individuals with special knowledge and expertise. In addition, I also collected and analyzed numerous documents generated throughout HathiTrust's pre-and-early history and well as contemporaneous artifacts. These data are the primary inputs of the chapters to follow.

Too little is known about the processes and mechanisms of sociotechnical transformation. A case study approach yields the best information for beginning to fill that gap and lay some of the groundwork for other forms of research (surveys, experiments, and so forth) to follow in the future. HathiTrust provides a kind of "natural experiment" that allows me to study the phenomenon. It is an "existence proof" of important aspects of this topic of sociotechnical transformation.

Chapter V: Pre-History of HathiTrust

This study explores processes of transformation in sociotechnical systems by tracing the emergence and evolution of HathiTrust. As previously discussed in Chapter II, large-scale digitization efforts had been ongoing since the late 1970s but faced significant technological, financial, and social challenges. In fact, due to the limited scope of these earlier digitization efforts, copyright law never surfaced as a significant concern in most of these early endeavors. The Google Library Project represented a significant departure from earlier large scale digitization efforts.

The story of HathiTrust begins not with its official launch in 2008, but with Google's mass digitization project ("MDP"), discussions for which began as early as December, 2002. As subsequent chapters will describe, HathiTrust emerged out of the Google MDP and particularly the partnership between Google and the University of Michigan. This chapter first describes some of the principle contours of the Google Library Project and highlights some of the main tensions that arose among various stakeholders, particularly at the intersections among academic research libraries and information technology firms.

Next, the chapter delves into sensemaking around the MDP, using public statements and contracts (where available) to make some general observations about the sensemaking of MDP partner institutions before focusing in greater depth on the sensemaking and decision-making around the University of Michigan-Google agreement. Interview data is used to draw out the main justifications offered in support of Michigan's behavioral commitment — its decision to partner with Google in the digitization of its entire print collection. These justifications include pragmatic, ideological, reputational, as well as copyright concerns and reflect the convergence of socio-structural forces and individual and organizational factors. The innovative deviance framework is applied as a way of helping further describe and explain the University of Michigan's sensemaking and decision making process with respect to the mass digitization of millions of copyright-protected works in its library collections. In addition, special attention is

drawn to section 4.4.2 of the UM-Google agreement; this provision is significant not only because it enabled cross-institutional collaboration around the newly digitized corpus — making HathiTrust possible — but because it demonstrates the role of policy-making as a way of safeguarding space for future (potentially transformative) innovation and generativity.

Google Library Project

To contextualize the emergence of HathiTrust, it is necessary to step back and briefly discuss some important precursors, particularly the "Google Book Search Project," the broad umbrella term encompassing the "Partner Program" and the "Google Library Project" and the mass digitization project. Under the Partner Program, publishers who hold copyrights in a work could authorize Google to scan and index their books so that users' search queries would produce information about the book such as bibliographic information, page number(s) containing the search term, and information about where to purchase the book from booksellers and/or from the publisher directly. In addition, under this Program publishers who permitted Google to display ads on the pages of their books would also share in the contextual advertising revenue. As the Partner Program is carried out pursuant to an agreement with rightsholders, it does not raise copyright issues with respect to potentially unauthorized and/or infringing activity. The Partner Program began in 2005 and has been widely used ever since.

Under the Library Project, Google scanned and indexed materials from partner libraries, made the full-text of public domain works available for browsing and/or reading, and made "snippets" (a few sentences of text around the search term) available for books subject to copyright protection. As with the Partner Program, basic bibliographic information about the book as well as information regarding bookstores and libraries that sell or lend the book is provided to users of Google search.[347] The Library Project involved the scanning of both public domain and in-copyright works and therefore the project raised concerns amongst various stakeholders regarding the legitimacy and legality of the scanning; the digitization has been the

[347] Google (2016). "Google Books Library Project: What does a Google Books Library Project book look like?" available at http://books.google.com/googlebooks/library/index.html

basis of a major copyright infringement lawsuit that, over a decade later, is still a potentially live controversy. [348]

The Google Library Project began in 2004 with the announcement of Google's agreements with several major research libraries. The so-called "Google 5" included Harvard University, Stanford University, Oxford University, the University of Michigan, and the New York Public Library. The Library Project quickly grew to include a number of other libraries. Over time, additional partners have joined but exact numbers are difficult to pin down because many aspects of the project, including basic information identifying partners, have remained hidden from view.

For example, Google's own website indicates that it has over 40 partner libraries around the world but it does not provide an exhaustive list. [349] Further confounding a precise understanding of the project, Google entered into separate agreements with each of its partners. These agreements defined the parameters of the partnership including identifying what materials the library would permit Google to scan and what Google and the library might do, or be prohibited from doing, with the resulting digital copies. All but a handful of the contracts remain sealed under confidentially agreements. Except in the few instances where state freedom of information laws have been invoked to overcome the non-disclosure agreements between public institutions and Google, or where confidentially was breached, basic information about the project, such as what portions of the collection have been digitized, what the ownership or disposition of those digital copies might be, and what uses might be permitted, is spotty. Thus there is a persistent and significant haze of uncertainty and ambiguity around the details of the Google Library Project.[350]

Despite having access to significantly incomplete data on the project, there are a few basic facts we know. For example, we know that Google pays all of the costs associated with the digitization, aside from the library personnel expenses associated with pulling and re-stacking books. Counsel representing Google in its lawsuit with the Authors Guild testified before the Second Circuit Court of Appeals that, as of December, 2014, Google had invested upwards of

---

[348] The Authors Guild has filed a petition for Certiorari with the U.S. Supreme Court. *Authors Guild, Inc. v. Google, Inc,* 804 F. 3d 202, Court of Appeals, 2nd Circuit, October 16, 2015; *Authors Guild, Inc. v. Google, Inc.* 954 F. Supp. 2d 282, Dist. Court, SDNY, Nov. 14, 2013.

[349] http://books.google.com/googlebooks/library/partners.html

[350] The publicly available agreements include: University of Michigan's agreement; University of California's agreement; University of Wisconsin-Madison agreement; University of Texas-Austin agreement.

$120 million on the project.[351] Furthermore, we can glean some basic information from institutional press releases and other public statements commonly published when the partnership is announced. The table below provides information identifying known partners, the date they joined the project, and some information about the (at least initial) scope of their participation. (*Table 1. Google Library Project*)

Table 1. Google Library Project

Name	Joined	Scope
University of Michigan[352]	12/2004	PD & in-©; UM retains digital copy; UM can use copy internally and in offering cooperative web-based services (§4.4.2)
Harvard University[353]	12/2004	PD only; Harvard retains digital copy
New York Public Library[354]	12/2004	PD only
University of Oxford[355]	12/2004	PD only
Stanford University[356]	12/2004	PD first (pre-1964) then in-©
University of California[357]	8/2006	PD works only; UC retains copy for internal use; distribution of text files to other libraries limited to 10%
Complutense University of Madrid[358]	9/2006	PD only; C.U.M. retains digital copy; cooperation possible

[351] *Authors Guild v. Google* Oral argument before the Second Circuit on December 3, 2014, Synapsis available at robertjbernsteinblog, https://robertjbernsteinblog.wordpress.com/2014/12/06/authors-guild-v-google-oral-argument-before-second-circuit-on-december-3-2014/

[352] University of Michigan Library – Google Digitization Project: A Brief Overview, Dec. 13, 2004, available at http://www.lib.umich.edu/michigan-digitization-project/michigan-digitization-project-complete-list-resources; U-M Library/Google Cooperative Agreement, available at ; U-M Library/Google Amended Agreement, May 2009, available at http://www.lib.umich.edu/michigan-digitization-project/michigan-digitization-project-complete-list-resources ; Further Information on U-M Google Amended Agreement available at http://www.lib.umich.edu/michigan-digitization-project/michigan-digitization-project-complete-list-resources

[353] Harvard University Library, Harvard-Google Project, http://hul.harvard.edu/hgproject/index.html

[354] http://newsbreaks.infotoday.com/NewsBreaks/Google-and-Research-Libraries-Launch-Massive-Digitization-Project-16307.asp

[355] University of Oxford Bodleian Libraries, Oxford Google Books Project, available at http://www.bodleian.ox.ac.uk/dbooks

[356] Stanford University Libraries, 2005 Statement of Support & Participation, available at http://lib.stanford.edu/google-books/statement-support-participation-2005

[357] http://www.cdlib.org/cdlinfo/2006/08/10/uc-libraries-partner-with-google-to-digitize-books/ Broken links to press release and additional information about the partnership. But see University of California Digital Library, UC Libraries Mass Digitization Projects available at http://www.cdlib.org/services/collections/massdig/faq.html

[358] Jimenez, Ricardo Acebes, The partnership project between Complutense University of Madrid Library and Google Book Search, Fourth UNICA Scholarly Communication Seminar 2008, May 15-16, 2008, Charles University in Prague.

University of Wisconsin-Madison & Wisconsin Historical Society[359]	10/2006	PD works & university-owned works; links to Google[360]
University of Virginia[361]	11/2006	PD & in-©
University of Texas-Austin[362]	1/2007	PD & in-©
Library of Catalonia[363]	Early 2007	PD only; links to Google
Princeton University[364]	2/2007	PD only
Bavarian State Library[365]	3/2007	PD only
Cantonal and University Library of Lausanne[366]	5/2007	PD only
Ghent University[367]	5/2007	PD only
Committee on Institutional Cooperation (CIC)[368]	6/2007	PD & in-©; links to Google
Keio University[369]	7/2007	PD only
Cornell University[370]	8/2007	PD & in-©; Cornell retains digital copy

[359] University of Wisconsin-Madison Libraries, Overview of the Google Library Initiative, available at http://proxy.library.wisc.edu/digitization/faq.html#primary.
[360] UW/Google Agreement was amended in 2009 to permit sharing across libraries and linking with HathiTrust copies in addition to Google copies. See Press Release, UW-Madison Expands Agreement with Google, July 9, 2009, available at http://proxy.library.wisc.edu/digitization/press.html
[361] University of Virginia, U.Va. Library Joins the Google Books Library Project, November 14, 2006, available at http://news.virginia.edu/content/uva-library-joins-google-books-library-project
[362] University of Texas at Austin, The University of Texas Libraries Partner with Google to Digitize Books, January 19, 2007, available at http://www.lib.utexas.edu/about/news/google/
[363] Technologia, Google digitalize 35 mil libros de la Biblioteca de Catalunya libres de derechos de autor, available at http://www.lavanguardia.com/internet-y-tecnologia/noticias/20090727/53753696854/google-digitaliza-35-mil-libros-de-la-biblioteca-de-catalunya-libres-de-derechos-de-autor.html
[364] Princeton University, Library Joins Google project to make books available online, February 5, 2007, available at http://www.princeton.edu/main/news/archive/S16/84/71S02/index.xml?section=topstories
[365] http://www.imageware.de/en/anwenderberichte/AWB-Loesungen/690-2/
[366] http://www.google.ca/googlebooks/library/partners.html
[367] http://www.google.ca/googlebooks/library/partners.html
[368] Committee on Institutional Cooperation, Google Book Search Project Introduction, available at http://www.cic.net/projects/library/book-search/introduction. CIC-Google Agreement available at http://www.cic.net/projects/library/book-search/cic-google-agreement
[369] Keio University Library, Keio University's collection digitized by Google Books Library Project, October, 2013.
[370] Cornell University Library, Google Books Search Library Project, available at https://www.library.cornell.edu/google-book-search-library-project

Columbia University[371]	**12/2007**	**PD only; Columbia retains digital copy; links to Google**
Austrian National Library[372]	?	PD only; ANL retains digital copy
Municipal Library of Lyon[373]	7/2008?	PD only?; library retains a digital copy

Based on the information contained in public documents summarized in Table 1, we can see that the project appears to have been rolled out in two main periods — the initial launch period in the Fall of 2004, and then a second period during late 2006-2007 — and that the second roll-out was primarily comprised of non-U.S. partners. Perhaps most importantly, we can observe that the project had acquired critical mass — a wealthy global benefactor in Google and lots of prestigious partner institutions — making it difficult to stop.

In addition, it appears that the vast majority of partners limited the digitization to public domain works, presumably to avoid potential copyright liability. The ownership interests in the digital copies varies from institution to institution with many institutions retaining copies that Google provide and others linking to or eschewing ownership. We can speculate that the disposition of ownership of the copies may also be tied to copyright-related concerns, although it is likely that technological and/or infrastructural limitations also play a role.

The Library Project has generated significant criticism from a variety of stakeholders on a number of fronts. The most obvious criticism was raised by some authors, author-advocacy groups, and publishers claiming that the Project infringes copyright rights. Shortly after the Library Project got underway, two major lawsuits were filed against Google. On September 20, 2005, the Authors Guild and several individual authors filed a class action lawsuit against Google for copyright infringement, requesting damages and injunctive relief. On October 19, 2005, five publishers (McGraw-Hill, Pearson, Penguin, Simon & Schuster, and John Wiley & Sons) sued Google for copyright infringement, requesting only injunctive relief. (A subsequent lawsuit, filed by the Authors Guild against HathiTrust, makes similar arguments and will be

[371] Columbia University Libraries, Columbia University Libraries Becomes Newest Partner in Google Books Search Library Project, December 13, 2007, available at http://library.columbia.edu/news/libraries/2007/20071213_google.html

[372] Austrian National Library, Austrian Books Online, available at http://www.onb.ac.at/ev/about/austrianbooksonline.htm

[373] ActuaLitte, Google partenaire numerique official de la biblioteque de Lyon, July 12, 2008, available at https://www.actualitte.com/bibliotheques/google-partenaire-numerique-officiel-de-la-bibliotheque-de-lyon-3396.htm

discussed in subsequent chapters of this work.[374]) In both cases, the plaintiffs based their claims on Google's copying the entire text of copyrighted works for use in conjunction with its search database and providing snippets of text to users in response to user search queries. Authors' and publishers' complaints generally revolve around issues of control and financial remuneration. The Library Project, and digitization more generally, is perceived as a threat to the established business models many authors and publishers have come to rely upon for their livelihoods.

The criticisms of authors and publishers were rebutted on several grounds. First, to the extent that copyright law is being used to bolster a declining method of conducting business, those arguments stand in clear opposition to provisions of the Copyright Act.[375] Second, to the extent that copyright law functions as a trade regulation aimed at preventing market harms from befalling copyright owners, it is not clear that the Library Project is undermining the market in authors' works. If a work is in-copyright but out-of-print, as many of the digitized works are, the argument is particularly flimsy because readers have few meaningful alternative venues for accessing the work. Relatedly, it is difficult to comprehend a rational basis whereby an author seeks to restrict users from accessing a digital copy of their work when the user has free access to a print version (albeit at a higher transaction cost). Finally, there are issues related to "orphan works" — works that are likely in-copyright but the rightsholder's identity is not reasonably ascertainable. Restricting access to these works seems, by definition, to be a victimless infringement where all plaintiffs are hypothetical and no one would have standing to sue.[376]

Authors and publishers are not the only groups who have been critical of the Library Project. While many legal scholars have expressed support for the endeavor, others have expressed concern. For example, shortly after the Google Library Project was announced, Siva Vaidhyanathan, a cultural historian and media scholar and professor of Media Studies and Law at the University of Virginia, raised a number of concerns including the privacy of patron records,

[374] See also Centivany, A. (2015).

[375] 17 U.S.C. §102(b) states "In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work."

[376] The orphan works problem will be discussed in greater depth in a future work that, borrowing from real property law, characterizes the orphan works problem as a problem of absentee landlords. Real property law (i.e. law governing land rights and use) provides mechanisms whereby an individual may gain rightful access, and even ownership, of a parcel of land through adversarial use however no such provisions exist in the context of intellectual property. In addition, while one of the claims in the Authors Guild's suit against HathiTrust involved orphan works, the courts declined to adjudicate those claims for lack of ripeness (although I believe one could object on the basis of standing as well).

risks of privatization, and property concerns tied to the tremendous investment in proprietary formats. Chief among his concerns, however, were the intellectual property issues raised by the project, namely that the project risks destabilizing a system already out of equilibrium.[377] Under this view, while it is true that the Google Library Project raised significant substantive concerns around specific legal principles, the larger concern was its impact on the rule of law more generally. The fact that the existing copyright regime is out of balance or miscalibrated with respect to technological change provides a justification against engaging in activities that are likely to exacerbate the disequilibrium. This concern seems to recognize the sorts of concerns expressed by social disorganization and structural strain theories. Instead of accepting that the disequilibrium is a trigger for normative competition or innovative deviance, however, Vaidhyanathan seems to argue that we should focus on bolstering and buttressing inner and outer containment factors — the individual and organizational buffers that maintain social order in the face of structural strain.

Framed a slightly different way, Vaidhyanathan's perspective may be indicative of a reluctance among legal scholars to break new legal ground in such a public and potentially high-stakes manner until sufficient groundwork has been laid to enable a court to make a ruling favorable to their viewpoint. Borrowing from a related history, many in the intellectual property and technology law community were still reeling from the Supreme Court's ruling in *Eldred v. Ashcroft*.[378] Eldred, lead petitioner in the case, was an Internet publisher who challenged the constitutionality of the Sonny Bono Copyright Term Extension Act of 1998. He was joined by a number of other publishers who relied upon the public domain for their work.[379] The Sonny Bono Copyright Term Extension Act provided rightsholders an additional twenty years of copyright protection and applied both prospectively to new works and retrospectively to works already in existence. Under the revision, the new copyright term amounted to the life of the author plus seventy years or, in the case of works-for-hire, anonymous, or pseudonymous works, ninety-five years from the date of first publication or one hundred and twenty years from the date of creation. The petitioners argued that extending the term of copyright project was an unconstitutional divestment of the public's interests in public domain works.

---

[377] Vaidhyanathan, S. (2005). A Risky Gamble with Google. *Chronicle of Higher Education*, *52*(15). See also Vaidhyanathan, S. (2006). Googlization of Everything and the Future of Copyright, The. *UC DAViS l. reV., 40*, 1207.
[378] *Eldred v. Ashcroft*, 123 S. Ct. 769 (2003).
[379] Pub.L. 105-298. Sections amended: 17 U.S.C. §§108, 203(a)(2), 301(c), 302, 303, 304(c)(2).

Passage of the Act was extensively lobbied for by members of the content industry. The Walt Disney Company in particular was a staunch supporter of the Act as it sought to delay its copyright in Mickey Mouse from expiring.[380] As its title suggests, Sonny Bono's estate was also a key supporter. Bono's widow testified before Congress that "Sonny wanted the term of copyright protection to last forever" but, as that would violate the Constitution, she suggested Congress consider Jack Valenti's proposal that copyright lasts "forever less one day."[381] The Act was also vehemently opposed by many law scholars, public interest advocacy groups, and technology firms. Ultimately, the twenty-year extension was passed.

Lawrence Lessig, a renowned intellectual property and technology law scholar and lead counsel for Eldred, expressed regret over his handling of the case. In an article entitled "How I Lost the Big One," Lessig explained how it was his own mistake and failure that lost this incredibly important case. In his view, he had focused too heavily on logic and reasoning and too little on impassioned and persuasive rhetoric.[382] As a result, countless works of authorship would remain under lock and key for another two decades and the public suffers. Others criticized Lessig for having underestimated his opposition, being overly confident in the correctness of his interpretation of the law and his beliefs in what was "right," and thus failing to lay sufficient groundwork for a successful argument. It was reasonable for some legal scholars and commentators, fresh off the disappointment of *Eldred v. Ashcroft*, to be hesitant about the Google Library Project's brazen willingness to tempt fate by opening a new portal for rightsholders to secure their own interests against "encroachment" of the public domain.

Library Perspectives

Within the library community, a number of issues and objections were raised about the Library Project including concerns around disintermediation, the social implications and community-borne resentment, and objections from competing digitization projects and the open access community. Each will be briefly discussed in turn.

Some librarians have expressed concerns that mass digitization may lead to disintermediation. The idea is that patrons may increasingly turn to Internet-based resources and physical libraries may become underutilized and, as a result, their financial support may become

[380] As a result the Act was nicknamed by some "The Mickey Mouse Protection Act."
[381] Testimony of Mary Bono, "H9952" *Congressional Record*. Government Printing Office. October 7, 1998.
[382] Lessig, L. (2004). How I Lost the Big One. *Legal Affairs. Retrieved January, 15*, 2007.

increasingly difficult to justify. Concerns around disintermediation have been particularly vociferous in the context of smaller municipal libraries and libraries in developing countries, but even within the ARL community fears of disintermediation as a result of large-scale digitization are being voiced.

A librarian employed by possibly the nation's largest and most prestigious university described the climate of library budgetary concerns as follows:

> "Now, we're so driven by costs. Everything needs to be transactionalized and so we've articulated the costs of every transaction in the library. We've got the costs of what the space is, how your taxes are being used in libraries, how much money the university is spending, how much money it costs them to maintain what some people might see as mausoleums, big buildings in the middle of campus that would be a great place for a student center or classroom space. So there is a constant threat to libraries. This is made even clearer when you consider that academic libraries, in relation to other units on campus, often employ the greatest number of staff. If a university is looking for permanent savings, where do look first? They look to cut staff. All libraries are under threat, some town libraries have unfortunately lost the battle, but even university and college libraries are under constant threat."

The concerns raised by this study participant were made even more palpable by the knowledge that the interview was taking place on their last day of work. Due to what they called a "slash and burn" restructuring at the library, this particular librarian decided to (or perhaps felt pressured to) resign and look for opportunities elsewhere, possibly in a different sector.

The majority of librarians interviewed for this study were not as concerned about the potential of disintermediation, despite sharing a general concern around budgetary issues. A more general perspective was that digitization had, in general, a positive impact on the role of libraries and librarians. The availability of digital content tended to draw more attention to the collection, bring more people into the library, and actually bolster interest, support, and use of the brick-and-mortar library space.

In addition to concerns around disintermediation, another criticism about the Google Library Project stemmed from resentment about being "passed over." Some members of the library community voiced discontent that Google had not approached or chosen their institution for partnership in this ground-breaking project. As previously mentioned, a number of libraries had been actively engaged in digitization efforts. Raj Reddy and colleagues at Carnegie Mellon University had been working on their Million Books Project, which sought to create a free-to-

read, searchable collection of one million books available to everyone over the Internet.[383] Cornell University had been a leader in digitization and, in particular, standard-setting efforts. The University of California system had established the California Digital Library, linking its users to digital information culled from its various campus libraries. Many of these efforts were undertaken with the support of public and private (nonprofit and for-profit) funds. The National Science Foundation, the Alfred P. Sloan Foundation, Xerox, Microsoft, Yahoo!, and the Internet Archive invested resources, money, and expertise in many of these early collaborative digitization efforts.

When murmurs of Google's project began filtering through the library community, it prompted a mixed reaction. For some, it generated anticipation and excitement. It raised the hackles of others; there was a concern that Google would throw its weight around and quickly dominate the digitization playing field. In addition, when the Google 5 were announced, some of the early innovators in large-scale digitization were disappointed or resentful that their institution had not been selected.

A librarian at the University of California noted the tensions and responses within the library community when news of the Library Project began to come out:

> "Another perspective or window on those relationships was the facts surrounding the Google 5. University of California is, of course, in Google's backyard and we weren't part of that project and I think that to some degree the formation of the Open Content Alliance was a response to the Google project."[384]

The concerns and resentment were, in the case of the University of California, the Internet Archive, and other partners, channelled toward a competing digitization project, undertaken under the umbrella of the "Open Content Alliance." This endeavor will be discussed in subsequent sections.

One can only speculate about Google's sensemaking around the selection of its initial partners. Based on the institutions that comprised the Google 5, Harvard University, Oxford University, Stanford University, University of Michigan, and the New York Public Library, some possible conclusions might be drawn. For example, institutional reputation and prestige may have played a role. Related to this, the collections housed by the various partners (their size,

[383] Reddy, R. and StClair, G. (2001). "The Million Book Digital Library Project," Dec. 1, 2001, available at http://www.rr.cs.cmu.edu/mbdl.htm

[384] Interview with digitization pioneer, transcript on file with author.

quality, strengths, special collections etc.) likely had an impact of Google's selection. Geographical diversity may have played some role. Institutional culture, particularly whether the institution was risk-averse or risk-taking, overly bureaucratic or willing to fast-track initiatives, and conservative or exceptionally innovative may have played a role. Finally, it seems plausible that social relationships and the personalities of key participants probably played some role. For example, Larry Page, the co-founder of Google, is an alum of the University of Michigan and had a personal role in developing the project at that site.

To the extent that there was disappointment or resentment amongst members of the library community for being "left out" of the Google Library Project, those energies were fairly quickly channelled into support for alternative projects, perhaps most notably the Open Content Alliance. The Open Content Alliance formed as a collaborative digitization effort, spearheaded by Brewster Kahle and the Internet Archive, united around making digital content openly accessible to everyone over the Internet. Kahle described its origins, and its relationship to the Google project, as follows:

> "We started hearing secret murmurs of secret negotiations going on between Google and libraries, that there was going to be a digitization project. I was receiving an award at a library conference, the Coalition for Networked Information, so I got to make a speech and I basically asked: 'If we build it will you come? If we make it so you can digitize a book for $0.10 per page will you support this?' Carole Moore, the Librarian at the University of Toronto, said she wanted to work with the Internet Archive to digitize all of the library's public domain works. I cried for help because I suspected that, working with Google, libraries would go and make a closed system. Sloan funded what became the start of the Open Content Alliance, and Yahoo! chipped in and eventually Microsoft decided to pay $0.10 per page to really get the digitization going. Microsoft eventually put in ten million dollars through the $0.10 per page thing. And it turned out that Google and the libraries *were* making a closed system. It would be accessible by some people, people in large research institutions, but not the public."

Despite his misgivings about the Google Library Project, Kahle does credit it with "getting people off their butts" and motivating a tremendous amount of activity around digitization.

Carole Moore has since retired from the University of Toronto but she did participate in this study, as did a number of current U of T librarians and Internet Archive employees who still run the scanning center housed on the 7th floor of Robarts Library at the University of Toronto.

In conducting this study and, in particular, researching the technical aspects of the project and its workflows, the open vs. closed dichotomy becomes somewhat striking. I had the opportunity to observe the Internet Archive's scanning center at the University of Toronto, to observe and photograph the Scribe machines (scanners), to observe the scanning process, and to interview its employees. (Figure 15)

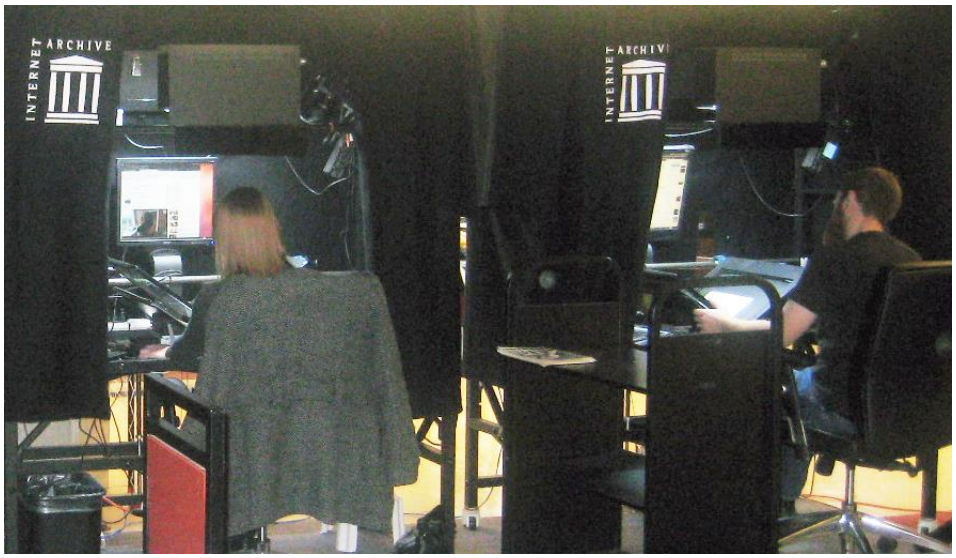


Figure 15. Employees scanning books at the Internet Archive's scanning center, Robarts Library, University of Toronto.

The scanning center on the seventh floor of Robarts Library has approximately two dozen scanning stations. In recent years, the scope of the digitization at Toronto has been scaled way back. Only a couple of the scanners are in use during a typical work day. But during the digitization heyday at Toronto, all of the stations were filled by full-time employees working in multiple shifts. The scanning center is located on a secure floor of the library which means that only certain library staff would have access to the facilities. The restricted access is more likely a consequence of the other tenants of the space, namely library IT, rather than an indication of the Internet Archive's wishes to around privacy and security.

By contrast, Google's scanning centers are secret, even from its institutional partners. A University Librarian at one of Google's partners recalled visiting a scanning center but the recollection was vague as he had only visited it once. He also noted concern about his obligations under the university's non-disclosure agreement and therefore exercised extreme caution in discussing his (admittedly foggy) recollections.

Carole Moore, the former University Librarian at the University of Toronto, described her decision to partner with the Internet Archive and house a scanning center as follows:

> "Many of the larger libraries were discussing and considering the Google contract, sort of talking to each other but sort of not talking to each other. We at the University of Toronto were talking actively to the University of California at the time, which was very anti-Google because they wanted a much more open approach. We were talking about how we would or could collaborate. It was sort of open, but everybody was keeping their own counsel in terms of what they would do in their own institution. But in Canada, all grants provided to public institutions are always under the condition that the digital content is open freely to everyone. The public tradition of Canada, the public investment and the contractual requirement gave us this tradition that things would be open. So the Google conditions did not sit well with us at all. We started experimenting with Brewster, with out-of-copyright books, and we were learning along the way. We went from one (scanner) machine that didn't work all that well, to two, to ten. At one point we had 26 machines and were doing two or three shifts per day."

The University of Toronto, as one of the Internet Archive's main partners in the Open Content Alliance, moved forward with the digitization of public domain works, content to figure out the technical aspects project as they went. The positive relationship between Carole Moore and Brewster Kahle played a significant role in the relatively smooth success of that partnership. But the Open Content Alliance faced some internal challenges with some of its other institutional partners.

For example, some other members of the Open Content Alliance were troubled by the lack of organizational infrastructure to support the endeavor. The California Digital Library, for example, had a relatively long history of digitization and was already well-integrated and embedded into a complex institutional arrangement as the digital library for the entire 12-campus University of California system. Some OCA members from the CDL found the process of adapting to the looser, more emergent, organizational structure of the Open Content Alliance challenging:

> "The Open Content Alliance was defined as an alliance of organizations but it never really developed any organizational infrastructure. I think that was somewhat frustrating to many of the libraries that were part of that initiative. Many of us really wanted to form some organizational infrastructure around the partnership and that was challenging with the Internet Archive because it's a very different sort of organization. It's led by one brilliant individual and was not founded as a true collaborative organization. Efforts to turn it into a more collaborative organization were not successful. We discussed it but were never able to move it to an organization that had a coherent shared leadership and decision-making authority. The Open Content Alliance didn't really exist organizationally. It was just a series of funded activities that were loosely

> connected primarily by the funding source (Microsoft) and by the Internet Archive as a galvanizing force."[385]

It was apparent, at least to some members of the library community, that in addition to the technical challenges of large-scale digitization, social and organizational aspects also posed some potentially serious obstacles. Despite shared core values of openness and shared goals of making the cultural record accessible in digital format online, forging effective collaborations and building momentum for collective action proved to be a significant stumbling block.

In addition, securing continuous funding for digitization was a constant source of strain. While Microsoft and others had contributed significant funds to the Internet Archive's efforts, they also put certain conditions on how those projects would develop including, for example, the selection of which materials could be scanned. If there were specific collections a library wanted scanned, they often had to secure their own sources of funding through grants or other forms of support.

By contrast, Google covered virtually all of the costs associated with scanning, and they scanned everything the library would permit. Thus, while the University of California had a strong commitment to open-access and had been, at least by some accounts, resented having been left out of the initial roll-out of the Google Library Project, it did ultimately join the project in 2006. The University of California continued, for a time, with the Open Content Alliance even after it partnered with Google but, at least for Kahle, UC's move was somewhat of a hit to the open access cause. Describing one of the key administrators at the University of California, Kahle said:

> "He was one of the most articulate and really thoughtful people in this whole area. He backed the Internet Archive and the Open Content Alliance and was really articulate as to why you'd want to do this. He's a really smart guy. But then he went and backed Google and I think it was monetary. I think it was just, they offered to pay for it. It's something the libraries have always wanted to do and they look at us (the Internet Archive) and say, 'How are you going to get there from here?' So I think it was 'Take the money where you can.' And also, he thought, and I think he was right, that someday these books will be freed up. But it hasn't happened. They're still sitting on the public domain. It doesn't make any sense."

Kahle's view of the Google Library Project as being a fundamentally closed system benefitting the academic research community to the exclusion of the general public is disputed

[385] Interview participants from the California Digital Library, transcript on file with author.

by others. Many people argue, for example, that the public domain works are, in fact, free-to-read online. That said, however, Google's agreements with its partners require those institutions to put technical protection measures in place to prevent bulk downloads of content. In Kahle's view, these provisions unnecessarily restrict the public domain which should be downloadable and minable as well as readable.

In addition, the terms of Google's agreements with partner institutions continues to exclude the Internet Archive from collaboration. Kahle has asked Google Library Project partners for access to the public domain scans for inclusion in the Internet Archive's database but those requests have been rejected. Never one to take rejection lying down, Kahle and others have pushed back against Google's restrictions. Kahle recounted a particular example involving the late Internet entrepreneur, visionary, and open access hacktivist, Aaron Swartz:

> "Aaron Swartz, who by the way worked at the Internet Archive for a period of time, orchestrated a set of his friends to slowly download Google's public domain scans and upload them to the Internet Archive. 800,000 of them were done this way. The Internet Archive attributes the copies to Google and the libraries they came from originally, and we did our best to OCR them, but they are in the public domain so they are posted on the Internet Archive site. Google, to its credit, did not assert copyright on these digitized materials. We've gotten unsupportive comments from one of the Google libraries but very supportive comments from another for having these materials up on the Internet Archive. I found it an interesting Rorschach test of the libraries and librarians."

Amongst the librarians I interviewed, this revelation about Swartz raised modest disappointment rather than outrage. As one participant reflected:

> "It's a tough one. There are those people who say 'These are public domain books and so they should be freely available to everyone else.' But I'm saying that Google is making a huge capital investment that we can't make ourselves for digitizing these books and I'm going to be respectful of that. And so, I don't really approve of what Swartz did. I wish he had gone and digitized those 800,000 books himself or convinced Google to change its practices."

This sentiment, while understandable, is not really practicable. The technical, financial, and organizational challenges of large-scale digitization proves prohibitive for many established, well-endowed institutions, let alone a (albeit incredibly dedicated and talented) college student.

To summarize, some of the key issues raised in contemplation of Google's mass digitization project were: 1) copyright infringement, 2) harm to authors and publishers, 3) exacerbating copyright disequilibrium and anomie, 4) opening up vulnerabilities for further

encroachments on the public domain and fair use by rightsholders, 5) potential for disintermediation, 6) tremendous costs associated with digitization and the overall financial strain most libraries operate under, 7) organizational and institutional challenges of collaboration and collective action, and 8) ideological choices and tensions along the closed vs. open spectrum. This list reflects both socio-structural top-down factors and bottom-up social or relational concerns. Whether and how the Library Project might ameliorate or exacerbate those issues was, at the early stage, anyone's guess. Stakeholders on all sides of the issue pressed forward into the unknown, bracing for whatever they might find (and create).

Sensemaking and Decision-making around the Library Project

As already noted, we have few hard facts about the Library Project because so much of it remains confidential. In this section, using publically available data, I begin with some general observations about the sensemaking of Library Project partners. While it is difficult to draw conclusions in a data vacuum, we can use the publicly available statements and press releases that so often accompany institutional participation to speculate about the sensemaking of many of the Project's partners. These public statements include basic information, such as the date particular libraries joined the project and key terms of the agreements with Google. I follow up with a more detailed description of the University of Michigan's sensemaking and decision-making process, drawing on public documents as well as in-depth interviews with key decision-makers.

Press releases by no means communicate the full and complete picture of institutional sensemaking but they can serve as important artefacts of the sensemaking process. One of the reasons for this is that public statements (such as press releases) illustrate the high degree, binding strength of the decision as institutions make their choice to participate in the MDP highly visible and irrevocable. Furthermore, as Weick and others have noted, the strength of the behavioural commitment tends to correlate to the relative intensity of the proffered justification for the action.[386] We can mine press releases for evidence of some early justifications for joining the Project and use them as a touchstone for understanding how the Project was conceived of at the outset and how it evolved and transformed over time, leading to the emergence of organizational forms and knowledge infrastructures like HathiTrust.

[386] Weick (1995:11-12).

The most commonly asserted justifications for participating in the mass digitization project tend to focus on the *preservation* of print materials, promoting the internal *educational and research mission* of the institution and, particularly in the case of libraries outside of the United States, *communicating cultural heritage*. Improving *access* is another oft-touted benefit although its value as a meaningful category is somewhat dispersed. "Access" is a term that seems to represent a number of distinct concerns around *dissemination, delivery, discoverability, social justice,* and *administrative efficiency*. Interestingly, despite Google paying for the tremendous costs associated with digitizing the content and in some cases making improvements to the quality of scans that were then shared with its partners through batch updates, few institutions cited *financial* concerns or *pragmatism (economic and technological)* as a justification. One institution, the University of Michigan, cited the *transformative* potential of the project — this will be discussed in a subsequent section. Perhaps most surprisingly, another institution cited as its sole justification the commemoration of its "150th anniversary."

These glimpses into the initial sensemaking around the Library Project, particularly with respect to the ways in which the justifications clustered around a relatively small number of themes, may be indicative of the social context in which libraries' decision-making occurred and the co-construction of meaning around the Project. Shared norms and expectations around preservation, access, educational/research mission, and (for non-U.S. libraries in particular) communicating cultural heritage influenced the rationalizations developed for entering the partnership and provided a process of legitimation.

In addition to providing a legitimizing mechanism, justifications offered in press releases may also have a predictive influence on how the Project ultimately developed. As the world's first mass digitization project, there was no obvious roadmap dictating the features that would emerge. But adopting the "library" moniker and offering a series of justifications closely tied to traditional library practices, norms, and values imbued the decision with what Weick calls tenacity. Tenacious justifications tend to prefigure subsequent perception and action, which means they often become self-confirming, verging on self-fulfilling prophecies.[387] The clustering of justifications provides some social cover against potential risk and provides clues and context for potential commonalities that could accommodate and support collective action.

[387] Weick (1995:13-14).

Another aspect of the early sensemaking around the Library Project which does not find an obvious analog in Weick's work deals with the concomitant legal implications of the choice. In addition to providing justification reflective and responsive to the library community, one can only assume that another purpose of the statements and press releases were to begin building legitimacy in the event of future litigation. In the legal context, we call this "creating good facts," building a narrative that can be referred back to as a method of producing selective attention and building resilience against criticism. One way in which this was accomplished was by providing justifications that emphasized libraries' special and cherished roles in society (i.e. foster education and research) and signal key library functions (i.e. preservation and access) can constrain sensemaking and foster confident action with respect to the Library Project.

Viewed with a more critical eye, the specific language adopted also reflects some insights into particular institutions' sensemaking about the potential legal risks and rationale for the endeavor. Cornell University, for example, provided a justification that focused almost entirely on discoverability, thereby positioning itself favorably in relation to existing case law — in *Kelly v. Arriba Soft Corp.,* the Ninth Circuit held copying for indexing purposes, to facilitate search, to be a fair use. Furthermore the absence of reference to preservation in Cornell University's public statements is also significant. As previously mentioned, Cornell was a leader in establishing standards for digital preservation and held the view that Google's scans were not of preservation quality. Therefore unlike the majority of other Library Project partners, Cornell's justification avoided all reference to preservation, signalling its position that a defence to copyright infringement on that basis was, in its view, unlikely to be successful.

University of Michigan's press release was the only one discovered that justified its participation, in part, on the "transformative" potential of mass digitization. Certainly this word choice references the broad transformational potentials of digitizing an entire library. It also provides an artful reference to copyright law's fair use doctrine which often turns on whether a given use transforms the original. But we need not rely solely upon the University of Michigan's press release for understandings of its sensemaking and decision-making process because the UM-Google Cooperative Agreement is one of the few that has become publically available. The remainder of this chapter focuses on the University of Michigan's sensemaking and decision-making around the Library Project, drawing upon data generated from in-depth interviews with key participants and also discusses the UM-Google contract and its relationship to sensemaking.

While large-scale digitization projects had been undertaken long before Google came onto the scene, earlier efforts paled in comparison to the sheer scale, speed, and magnitude of the Google Library Project. It was an effort full of potential and rife with uncertainty. These are difficult circumstances in which to make decisions and engage in contract negotiations, particularly when considering the typically conservative nature of most academic institutions butting up against the typically innovative and risk-taking nature of Google. It is somewhat remarkable that agreements between Google and university librarians, chief information officers, legal counsel, provosts, presidents, and regents were reached at all.

The potential copyright risks could have only exacerbated an already rife decision-making process. As previously discussed, copyright law gives authors a number of exclusive rights over their works including, for example, the right to make copies, to publicly distribute the work, and so forth.[388] The Act also contains various limitations of those rights including, for example, an exception for fair use, for a number of library and archive-oriented uses, and so forth.[389] Authors' exclusive rights and the limits on those rights reflect copyright's means-end formula; they reflect the bargain between the public's interest and the interest of rightsholders and they operate in tension with each other.

While the balance between rightsholders' interest and the public's interest may seem straightforward, in practice it can be surprisingly unclear and messy. One scholar referred to the Copyright Act as a "swollen, barnacle-encrusted collection of incomprehensible prose."[390] As a practical matter, it is often difficult to know what is legally permissible and what is not. The extent to which a university might successfully invoke fair use for digitizing its entire corpus was uncertain at best, not least because the specific uses the University of Michigan might end up making with respect to the copies (and which would ultimately be evaluated in a fair use determination) emerged over time rather than being decided upon from the get-go.

Notwithstanding the tremendous uncertainty, one section of the Act that is relatively clear, however, is section five, which lays out the remedies for copyright infringement. Here we learn that remedies for infringement include, amongst other things, injunction, impounding and destruction of infringing copies, and damages of up to $150,000 per instance of infringement. As of 2012-2013, the University of Michigan Library contained approximately 8 million volumes.

[388] 17 U.S. C. §106.
[389] Limits on the exclusive rights are contained in 17 U.S.C. §§107-122.
[390] Litman, J. (2010). Real copyright reform. *Iowa Law Review, 96*(1), 09-018.

Even if we assume a large portion of those volumes are in the public domain, UM's liability could still hypothetically reach the trillions of dollars range. Although, as a public state institution, the University may be insulated from liability for monetary damages under the doctrine of sovereign immunity, injunction, impounding and destruction of infringing copies would remain available.

So how does a typically risk-averse institution like the University of Michigan decide to engage in an activity that is not only guaranteed to raise the ire of some rightsholders but also potentially (if the sovereign immunity and fair use arguments fail) expose the institution to damages for copyright liability and/or the embarrassment of an injunction that prevents the University from keeping (and using) the scans? In interviewing several of the key decision-makers about the sensemaking and decision-making process, a number of primary issues and concerns surfaced which demonstrated that pragmatic, reputational, and ideological factors played a role, in addition to copyright-related factors.

Large-scale digitization efforts had been contemplated and attempted before, with varying levels of success. More often than not, digitizers were waging a constant battle against financial pressures and the creep of technological obsolescence. By offering to cover virtually all of the costs, complete the project on an extremely fast timeline, and perhaps provide technical reassurances against obsolescence through batch updates, Google erased many of the concerns that had plagued earlier large-scale digitization efforts.

Several of my participants described the decision-making process as essentially a pragmatic cost-benefit analysis. The potential costs if the University of Michigan was sued and lost were less than the potential benefits if they did not get sued, or if they got sued and won, and the decision-makers were confident that the facts and the law would be on their side.

But it wasn't a sure thing. Lots of people recognized its potential but nobody really knew how the Google Library Project was going to turn out. As one participant said,

> "People thought, 'Hey, this is going to be good. It's going to be world-changing. We're not exactly sure how it's going to do it and what the exact benefits are, be we know somehow it's going to be good."

Another participant said:

> "We didn't have everything all figured out from the get-go. We knew that this was a great opportunity and we wanted to seize it but we weren't exactly sure

> what we were going to end up doing with the scans. I had great hopes for where this could go."

There were dreams and visions at this early stage, but the benefits were still uncertain and perhaps inchoate. This uncertainty undercuts the potential for conducting a true cost-benefit analysis: "it should always be remembered that intangible benefits are weak bases for rational judgments" because it is difficult to know what value to assign them.[391] True cost-benefit analyses have a tremendous difficulty accounting for uncertain and intangible benefits, particularly those that concern public goods like those contemplated by the mass digitization project. Thus, the initial justifications around the decision to partner with Google and digitize the entire library were based on the conviction that mass digitization is a better way, because it offers tremendous potential benefits to society, although that conviction may have been largely aspirational rather than factual.

Technological and moral ideology also played roles in the decision to partner with Google. A common theme that frequently resurfaced during my interviews was that digitization was viewed as an inevitability. Participants in my study talked about mass digitization not as a goal but as a given. The move towards a digital environment is going to happen, it is the future, and it is not up for debate. One of my participants said for libraries, "it's like digitization is written into our organizational DNA. It is what we have to do." Another participant reflected:

> "I can't recall anybody in the discussion saying 'Oh, we don't have to do it this way. This digital thing is overblown. We're not going to have to confront these issues.' It's like everybody knew they had to confront the issues and they did. They had to and they did."

This sort of technological determinism has faced criticism in the literature. Kling, for example, cautions against being too easily seduced by the pro-technology viewpoint which can often conflate technological progress with social progress. Particularly when a technological advance ignores key aspects of the social environment in which it operates, technological advance can have a deleterious impact on society.

One participant responded to Kling's brand of skepticism the following way:

> "What I took away from Kling was, saying 'It's a better way' is a pretty good argument. If you can really claim it's a better way and people believe it's a better

[391] King, J. L., & Schrems, E. L. (1978). Cost-benefit analysis in information systems development and operation. *ACM Computing Surveys (CSUR), 10*(1), 19-34.

> way, that's most of the battle. And I think, without gilding the lily here, digital access to most stuff is a better way."

Certainly some authors, authors associations, and publishers were claiming otherwise, that digital access could potentially harm some aspects or segments of the social world of creative and intellectual expression. By in large, the people who played a central role in U of M's decision to partner with Google did not find those arguments compelling. "The fact that the Google Library Project causes some people to grow concerned about their livelihood is ultimately a moral argument, not an economic one." Fears that mass digitization may undermine existing business models is not a sufficient reason to abstain from doing it, particularly when the facts do not support a finding of market harm to rightsholders.

As one participant noted, "there was this strong belief in the inherent rightness of the position of making stuff accessible to people, of making the benefits of those things accessible to society." We don't generally permit people to hold back the progress of society simply because they feel entitled to, grown accustomed to, or become dependent on continued enjoyment of the benefits that accrued to them under the old regime:

> "God dammit, I want there to be a mechanism where almost everybody in the world has access to almost everything that has ever been published in electronic form at zero marginal cost, perhaps with some subscription fee, but a fairly small one. That is what I think the world ought to look like. For academic work, I think that marginal cost and the subscription fee should probably both be zero. The Google project showed me a feasible path to get there, not a complete path, but the starting point: 'Okay, let's digitize a whole bunch of stuff so that all that prevents it from being available in the way I'd like it to be available is law and custom.' I also hoped that, in time, the existence of this corpus of work would cause various actors in the world want to figure out a way to use it and to use it well and that sort of pressure, especially from the youth, would eventually lead to arrangements and outcomes that would make the work readable as well as searchable. I was optimistic that if we, as a society, have valuable assets, then we, as a society, will figure out how to use them. That was the utopian goal."

This utopian vision of what might be possible was a key factor in the sensemaking around the Library Project. These visionaries understood, however, that "it was going to take a long time before we figure out how to get there," particularly with respect to the legal and promotional arrangements that such a vision would require.

When Larry Page approached Bill Gosling, the University Librarian for the University of Michigan, with an offer to create a digital back-up copy of its entire library collection, and cover

virtually all of the costs, to some it seemed like an opportunity Michigan couldn't pass up. It also fed into Michigan's sense of ego and exceptionalism.

When Google, one of the world's most dynamic and innovative companies, came forward it offered a way to bolster the University's sense of exceptionalism and reputation, and its unique role in the State's economic future. There was strong institutional support from the key players in the essential University sectors including the Library, Information Technology, Office of General Counsel, the Provost's Office, as well as from the President and the Regents. One of those players summarized the sense as follows: "There was a very strong feeling of Michigan exceptionalism on the part of key players that this is the kind of thing that Michigan does and we should do it." Beyond being a good thing for the institution, there was a sense that Google was a rising tide that could potentially raise all of the state's ships.

While pragmatic, ideological, and reputational concerns played a role in Michigan's decision to partner with Google, the decision did require some careful deliberation, particularly with respect to the copyright issues. As one participant noted:

> "There was a fair amount of turmoil and contention about copyright issues. One the one hand, these were not risk-taking people who wanted to be sued by everybody on earth, but they weren't particularly afraid of being sued because they thought that this was a better way."

Based on my interviews, key figures in the Office of the General Counsel developed a legal rationale to support and define U of M's participation in the Google Library Project that the key decision-makers found convincing. The rationale was based on the concept of the "dark archive" which is a term of art borrowed from archival science that describes a collection which is held in private, without any opportunities for access. As described by one of my participants, in the context of the mass digitization project, the dark archive doctrine provided that "the right of first use extends to our making a digital copy of everything we own."

The dark archive rationale is not a copyright doctrine, although there are traces of it in §117. That section applied only to computer software and entitles the purchaser of software to make a backup copy which they do not use, i.e. it remains "dark," unless and until the original copy is damaged, destroyed, corrupted, etc. Obviously the library's print collection is not computer software so it appears this rationale is some murky blend of traditional archival practice and various other unidentified mitigating factors (such as fair use and sovereign immunity). Although I was not granted interviews with members of U of M's Office of General

Counsel, based on my interviews with other central figures, the overall experience of those involved suggests that the dark archive rationale was not really interrogated and the decision to join with Google was fairly simple, straightforward, and not overly deliberated upon.

Another central figure in Michigan's partnership with Google described the deliberations around the copyright implications as follows:

> "On one level the copyright issues are crucial. On another level, we didn't pay much attention to them because we decided very early on that, given what we were going to do, we were not at significant risk for a finding of infringement or damages from infringement. I think we have the right to scan in-copyright works and use them for research purposes, for preservation purposes, for all students with disabilities to read them. We've already been doing that. We believed we could make a good faith case based on fair use and library privileges in the Act to defend the uses that we were making. I'm sure we did the legal analysis, but I don't remember doing it."

The point about being able to make a "good faith case based on fair use" is significant because, as mention in Chapter II, libraries may be shielded from monetary damages where their activity, later determined to be infringing, was undertaken with a good faith belief that it constituted a fair use. This raises some interesting questions around the standard of care and, in particular, whether the good faith belief must be both subjective and reasonable.

As another participant recalled:

> "One of the things that I think was really laudable was the bravery of Michigan's President (Mary Sue Coleman). I don't know whether she herself really thought it through but she was basically unafraid. The digitization project resonated with her. It was a risk she was willing to take. She said, 'We're going to go ahead and do this. We're going to partner with Google. We're going to scan all these books. We're going to create this thing.' If you were trying to identify a signature of her presidency, I think this is it."

Another participant reasoned:

> "I argued in favor of partnering with Google because it was a move that would force theories. Either people would be silent about it and they would be okay with it or it would force a fair use case that would be on favorable terms for us, assuming we did it right. And some of this is retrospective, and there is probably a warm glow of looking back knowing how things played out, but at the time I really do remember being very concerned that we either use fair use or we lose it. We were looking at the question prospectively rather than just reactively. And our motives were impeccable. There was no profit incentive. This was about both preservation and creating a corpus. And pragmatically, it was a trajectory that the University was on at our own pace, but even though at the time we were pursuing

> one of the most aggressive digitization efforts in academia, at the pace we were going, we would have gotten where Google is now in about 100 years. So it was seen it as a way of accelerating what we were already doing.
>
> Short of licensing something, there is no way to guarantee you won't become a test case for fair use. The only way that you can determine that your use was, in fact, definitively a fair use, is to have a judge tell you that. Part of the challenge around copyright cases is, for the most part, publishers pick cases that they think they will win, and then use those decisions to narrow the scope of fair use. And the Google Library Project felt to me, at least intuitively, like, 'Man, if we're going to have a discussion about fair use, this is the project to have a discussion of fair use around.'"

It is likely that the prospect of protection under sovereign immunity provision played some role in Michigan's decision-making process although the participants in my study indicated it did not weigh heavily on the ultimate decision:

> "We wanted to have the fight on the terms of the fight, not because we have sovereign immunity and can't be held liable for infringement. It really served as a safety valve in the event that everything went down in flames at least they couldn't get damages."

As incredible as it might sound, given the worst-case scenarios offered by the Copyright Act, the decision to make a back-up copy of the entire Library was one instance, consistent with a long trail of precedents and likely antecedents, of Michigan's "selfless audacity."[392]

As described by one of lead architects of U of M's decision to join the Google Library Project, the thinking at the time was that:

---

[392] This phrase was borrowed from the following piece publishing in the Chronicle of Higher Education concerning Michigan's decision to merge its press with its library, but it could just as easily describe a number of other similarly situated endeavors:

> "The University of Michigan's bold decision two years ago to merge its press with its library, and to publish all future books online, free of charge, offers tremendous hope and a way out of our predicament. What is so inspiring about Michigan's experiment is its selfless audacity, its resolve to produce good, free books without waiting for other institutions to reciprocate. A pragmatic, calculated business plan might have looked something like a Start treaty: We, Michigan, will undertake years of negotiations with presses and libraries at other institutions, cautiously acting in concert, committing ourselves and a circumscribed group of signatories to verifiable targets for sharing our publications without charge, while striking protocols to ensure that what we give away is commensurate with what we receive from others.
>
> Instead, Michigan acted unilaterally, with no assurances of reciprocity, in a fit of altruism."

Geffert, B. (2011). Libraries, publishers and a plea for a shotgun wedding. *The Chronicle of Higher Education, 20*. See also, Fister, B. (2001). "Selfless Audacity" Means Creating a Sustainable Not-a-Business Model | Peer to Peer Review, March 24, 2011, available at http://lj.libraryjournal.com/2011/03/opinion/peer-to-peer-review/selfless-audacity-means-creating-a-sustainable-not-a-business-model-peer-to-peer-review/#

> "This is probably the showdown that we've all known had to happen. And if we lose, it's not over. And if we win, it probably is over. And I didn't ever hear it said, but I think there were quite a few people who thought that 'this is the last chance for people who are really opposed to us digitizing the stuff at all to prohibit us from doing that."

The legal risks were real, and potentially serious, but so were the pragmatic, ideological, and reputational considerations. The conviction that digitization was both a better way and an inevitability provided support to the committed action and, combined with the tremendous inchoate potential of a digital copy of the library, formed the basis of justifications subsequently offered to explain and legitimize the decision.

The University's sensemaking around the decision to join the Google Library Project signals a number of potentially important observations. First, social relationships formed a foundation of the behavioral commitment to partner with Google. This was true between U of M and Google; Larry Page, a co-founder of Google, was also an alum of U of M. More importantly, it was true within U of M; a small but tightly knit cluster of key players within U of M who shared the culture of Michigan exceptionalism controlled the salience of the mass digitization problem (ultimately a copyright problem) and facilitated with confidence selecting the tradeoff of vigorous action over careful deliberation. As Weick notes, a choice amongst alternatives will ultimately come down to the control of salience, deciding whether deliberation or action is more important. As was the case with the Google Library Project, this decision is often irreversible: "If you choose in favor of accurate sensing, you reduce your capability to take strong action." Moreover, looking before one leaps may result in not seeing anything, and therefore may result in inaction. If one just jumps in, on the other hand, action is guaranteed to generate outcomes that ultimately provide the raw material for seeing something.[393] All solutions, however, do not require massive action and so, again, it is a question of salience. If a problem is construed such that only a relatively small corrective action is required, then detail and accuracy are crucial."[394] If a problem is construed on a grand scale, as many social movement-oriented problems are, then an action-oriented approach is better suited. Based on my interviews with participants involved in Michigan's decision to join the Google Library Project,

[393] Weick (1995:53).
[394] Weick (1995: 50).

it was readily apparent that the problem(s) mass digitization was going to solve were construed on a grand scale: digitization is a better way, it is the future, it is inevitable.

At the same time, however, with respect to copyright law there was no bridge between "here" and "there." There was no obviously legitimate path for digitizing the entire print collection. The mismatch between the legitimate goals of digitization and the lack of available legitimate means for accomplishing the goal produced a structural strain that pushed firms and institutions like Google and UM toward innovative deviance. The possibilities of the MDP highlighted a mismatch between the goal of copyright and the articulated available means designed to achieve that goal.

Innovative deviance offers a new analytic framework for understanding how and why the University of Michigan and other institutions engaged in mass digitization of in-copyright works. Michigan engaged in behavior that, in the view of many if not most, promoted the goal of copyright law — the cultural and intellectual enrichment of society — but, with respect to copyright law as it existed at the time, it used unorthodox, potentially illegitimate, potentially infringing means to do so. Under this framework, UM's decision behavior was not the result of opportunistic or irrational decision making or technological determinism, but rather was a natural consequence of the imbalance in copyright's means-end formulation brought about by significant changes in information technology. Technological change strains the structural foundations of copyright law. It creates an imbalance between the law's means and end and raises serious doubts about the existing law's continuing functionality, credibility, and relevance. Breaking the law becomes easier, perhaps unavoidable, and more easily justified.

This view is consistent with the interpretation of key participants in UM's decision-making process:

> "One of the ways in which the law gets changed is that it gets broken. And I would argue that the law's already broken, and what mechanisms of breaking the law do is reveal how the law can't be sustained under the new reality."

Borrowing from the social disorganization theorists, in the grand scheme of things, breaking the law or poking at the wounds of a dying regime is a form of normative competition. Reorganizing around new norms that accord with a new technological reality is essential to processes of sociotechnical transformation. Sometimes the existing laws and other institutional forms are resilient enough to be able to stretch and adapt to the contours of emerging

technologies and behaviors. In those cases innovative deviance can become a process or mechanism by which the law and institutions learn to change. This theme will reappear in subsequent chapters as the story of HathiTrust's emergence and evolution continues.

Post-Decision Considerations

Once the decision was made to join the project, a second set of concerns surfaced around how the partnership and Project would be defined. The UM-Google Cooperative Agreement is one of the handful of publicly available Library Project contracts. It is significant not only for what it reveals about the University of Michigan's sensemaking about the Project, but also for its role in subsequent library contracts and the eventual emergence of HathiTrust. This section begins with some general observations about the relationship between contract negotiation and sensemaking before highlighting the basic copyright issues and a few of the key terms of UM's contract. The contract provides important clues about how some of the uncertainties of the Project were managed, cordoned off and, in some cases, preserved.

Formal contracts like the UM-Google agreement operate in tension with the sensemaking process. In particular, they impose a certain degree of rationalizing and attempt a certain degree of prescience that may not realistically reflect the way most sensemaking occurs. Contracts are designed to reduce uncertainty, to design around all imagined pitfalls, to force the parties to slow down and deliberate about their actions and decisions. Contracts are made in anticipation of action, to constrain the possibility space of future activity, to protect against harm. They are a mechanism for *prospectively* opening and closing spaces of engagement and activity. In contrast, sensemaking is fundamentally *retrospective*. Contracts also serve important social roles by channeling behavior and providing evidence of terms of the agreement as ways of signaling to those outside the contract that a meaningful agreement exists. Contracts are, by their very design, distinct from the dynamic, revisionary, retrospective process of sensemaking they co-exist with. This section will first discuss a few of the key features of the UM-Google agreement and then draw upon interview-generated data to flesh out the sensemaking picture.

A close read of the UM-Google agreement reveals that it contains several significant terms. First, we learn that UM was the first institution (and remains one of the few) that permitted Google to scan its entire collection — public domain and in-copyright materials. Second, we learn that UM retained ownership of its digital copies whereas many other library partners refused ownership (presumably to reduce exposure to copyright liability or simplify

technical infrastructure issues). Perhaps most significantly, however, is provision §4.4.2 which reads in pertinent part:

> **4.4.2 Use of U of M Digital Copy in Cooperative Web Services.** U of M shall have to right to use the U of M Digital Copy, in whole or in part at U of M's sole discretion, as part of services offered in cooperation with partner research libraries such as the institutions in the Digital Library Federation.

Thus, this provision permitted UM to share its digital copies with other institutions and use them in the provision of cooperative web-based services. The importance of this provision is perhaps best described by the librarian who crafted it:

> "I wanted to make sure we had a provision that we could use to take what we believed to be the only comprehensive library digitization effort, the Michigan one, and leverage it for collective action around print management, management of the collective collection. §4.4.2 gave Michigan the right to use its copies as part of services offered in cooperation with partner research libraries. Michigan was the only institution that negotiated this clause and this is the clause that made HathiTrust possible."

Absent §4.4.2, it was doubtful that research libraries would have been legally permitted to share or collaborate around the digital corpus. We would have ended up with Google's range of services and up to forty siloed back-up copies (one for each participating library). While such digital siloes are not without value, it is the capacity of the libraries to cooperate and collaborate around this material, and build something by, of, and for the academic research world that was so potentially transformational. Without §4.4.2, there would have been no HathiTrust, no collective action around the collective collection.

After the UM-Google agreement was disclosed, and other institutions recognized §4.4.2, interviewees noted that Google shifted its approach by preemptively including §4.4.2-like privileges in all subsequent agreements. Similarly, based on interview data, pre-existing agreements were amended to also include the clause although there remain some questions about the legal effect of those revisions:

> "Many of the libraries' amended agreements with Google were dependent on the Google Books Settlement going through. When the court rejected the Google Books settlement, the libraries' amended agreements with Google were also rejected."

This would suggest that the terms of the original library-Google agreements would still control, however, as a practical matter, Google has discretion to enforce the terms of those

agreements and has shown no indication that they intend to restrict library collaboration. In other words, it appears that Google is allowing §4.4.2 to stand, effectively permitting libraries to engage in cooperative activities.

It can be tempting, in retrospect, to normalize the evolution of new sociotechnical forms. In hindsight, the Library Project, its agreements with various institutions including the University of Michigan, and the subsequent emergence of HathiTrust can seem like a logical, linear follow-on to previous large-scale digitization efforts. This research reveals how tenuous, conflicted, and emergent these developments were. With respect to §4.4.2 in particular, this research suggests that its inclusion in the UM agreement was largely value-driven rather than goal-driven. Participants at UM appreciated the potential power of the resources generated by the MDP — a digital back-up copy of its entire library — but its potential beyond serving as a back-up was largely inchoate and aspirational. Participants were not exactly sure what they would or could do with the content, but they knew they wanted to be able to decide collectively. §4.4.2 was a policy safeguard for the value of collectivism and for the MDP's inchoate potential.

Due in part to its questionable, uncertain, and/or murky purpose, negotiations around the inclusion of §4.4.2 were somewhat contentious. Some representatives from Google, particularly those who were invested in the technical design elements of the MDP, viewed the policy as a potential weak link in its still-emerging business plan. UM's lead architect of the provision described the negotiations in the following way:

> "I got very close to the person responsible for the digitization effort at Google and I can recall a number of conversations with them where they would essentially say: 'What the fuck did we do? I think we just gave away our business here!' And I would say: 'No, no, you did not at all. Google will find ways to capitalize on this that will not be undermined by another copy being out there.' And then they would say: 'But what is this 'Digital Library Federation?!' And I would say: 'No, no, it's just 'like' the Digital Library Federation. Don't get distracted by that.' The fact that we were also dealing with Larry Page and with the General Counsel – they had a bigger picture view and were not troubled at all by this the way that some product managers were."

Google was contributing significant resources to the MDP; they needed the libraries to provide access to the print materials but they wanted to avoid inadvertently creating a competitor in the provision of web-based services around the digital corpus.

Although representatives of Google declined to be interviewed for this study, citing on-going litigation around the MDP, one can speculate as to the reasons the company's co-founder

and general counsel were not overly concerned. Interviewees were quick to point out the ways in which the libraries and Google perform distinct functions, with different goals, and approach digitized material differently. Libraries regarded the digitized corpus, in some ways, as an extension of their print collections. While it presented new opportunities for search and discovery, remote access, and service opportunities for blind and disabled users, these new possibilities arose very much within the library environment. There was a strong centripetal force pulling the library scans back to the core missions of preservation, access, records management and ensuring high-quality metadata.

Google, on the other hand, had its own set of priorities that were likely different than those of the library. Even Google's Book Search, arguably the nearest cousin of the library community, offered different functionality and used quality measures calibrated for purpose distinct from those of libraries:

> "If you searched for something, Proust's "Remembrance of Things Past," in Google Books the three volumes were not together. And journals were scattered and not understood in a coherent way. In contrast, libraries would provide a serial record or a multi-volume unit title of a record which said "here are all the parts and here's how they relate to each other and the print holdings." If you had a thousand-volume journal that went back to the seventeenth century, the library would show you number one, two, three, rather than "your hit occurred in these 25 places in some uncoordinated way."

Discoverability and record management were just a few of the ways in which Google and libraries differed. Overall, there was a compelling sense that research libraries and Google operated in distinct, non-competing spheres. Libraries care a tremendous deal about metadata, record management, and core library missions around preservation and access as evidenced by the library-partner press release data. Google cares a tremendous amount about improving and expanding its computational power and reach via enhanced search algorithms and the development and provision of innovative new services like Google translate. The MDP was a means for each to achieve their particular ends. §4.4.2 made sense in the context of libraries traditional functions and did not seem to pose a serious threat to Google's interests or market dominance.

Hindsight bias can pose a significant obstacle to accurate tracing of the creation and subsequent impacts of a policy such as §4.4.2. There is a strong tendency for people to understand the Library Project and HathiTrust as it currently exists and retrospectively make

sense of its emergence and evolution as a rational, planned, fairly linear, perhaps even unavoidable progression from point A to point B. In reality, however, that could not be further from the truth.

When the UM-Google Agreement was entered into, and section §4.4.2 was drafted, UM was operating on the belief that the digitized corpus would be a dark archive – a digital back-up copy of the library with little to no provision for access.

> "The legal rationale supporting our decision to digitize our entire collection – in-copyright works as well as public domain – was based on the dark archive principle. We believed we were permitted to make a back-up copy of everything we own for ourselves. Whether we do anything with it or let other people see it is where we believed the copyright questions would come in."

It may, in retrospect, seem implausible or unbelievable that a sophisticated and savvy institution like the University of Michigan engaged in the mass digitization of its library to create a dark archive, but it was an understandable position to take given the state of copyright law as it was understood to exist at the time.

This is not to suggest, however, that there was no inkling that the digitized corpus might someday, somehow, under a modified set of legal and organizational facts, become something more than a dark archive. There was certainly a utopian vision of creating a universal digital library underlying §4.4.2, but it was deeply buried and very much inchoate when the agreement was made. No one, not even the architect of the provision, knew the extent of its significance and meaning in advance. §4.4.2 was, at that point, a matter of embedded generativity, safeguarding the possibility of future transformation within the academic research library world by preserving open spaces for collaboration and new, innovative, and unanticipated uses of what was assuredly an unparalleled digital resource.

Having lived through some of the earlier large-scale digitization efforts (and failures), the architect of §4.4.2 recognized that primary obstacles to the success of large-scale digitization projects were not technological or resource-based, but were rather social and/or institutional: "We needed to find a way to use the MDP to leverage collective action around the collective collection." Although it was not deterministic, the inclusion of this clause in the UM-Google Cooperative Agreement reflects important values and opportunities for generativity embedded the contract itself.

While §4.4.2 opened space for new, often unanticipated, forms of innovation and generativity, it is important to note that the clause also foreclosed some opportunities. The Internet Archive, for example, has been largely excluded from participation in HathiTrust because of §4.4.2 and other aspects of the UM-Google Agreement. As described by one of the lead architects of the Agreement:

> "The University of Michigan's agreement with Google allows us to share the scans with other library consortia and indeed part of the reason that the HathiTrust exists is to meet that requirement. It is a library consortium. The Internet Archive is not part of a library consortium."

Relatedly, the Agreement has fairly rigid restrictions about sharing and collaboration even with respect to the public domain materials:

> "Under the Agreement we would not have been permitted to open the public domain for mass use. We agreed with Google that we would take measures to prevent potential competitive usage of the scans – we argued that there wasn't actually that much economic value in the public domain scans – but nevertheless, we agreed that we would take measures to basically throttle mass downloading of the collection.
>
> Giving the Internet Archive permission to mass download all of the public domain materials to host on its site would not have been consistent with our agreement with Google. My personal view is that Google is being wrongheaded and they should have opened up the public domain but we have to wait for all these damn lawsuits to resolve before we can start being rational with them again."

Recognizing that sensemaking involves post hoc justification and reorganizing and reconstructing interpretations of events to resolve multiple, often conflicting interpretations of a decision's meaning(s), there may be more to the Internet Archive's exclusion than what appears on the surface.

Conclusion

Large-scale digitization efforts had been undertaken prior to the Google Library Project, but those efforts were relatively underfunded, ad hoc, and prone to obsolescence. They also paled in magnitude and scale in comparison to the Google Library Project, an important project in the prehistory of HathiTrust. A complete understanding of the Project is complicated by the persistent haze of uncertainty and confidentiality that surrounds the MDP. The overarching lack of information about the project, combined with its sheer magnitude and obvious copyright implications, led many stakeholders and commentators to voice concerns about the Project.

These concerns include: copyright concerns of rightsholders; concerns of legal scholars around the potential harm to the rule of law and/or risks of overzealous forays into fair use; concerns of librarians around disintermediation, reputational issues, and the perceived "closed" nature of the Project (both with respect to the general lack of transparency around the details of the project and in terms of digitized content).

Uncertainty and concern around the Project also produced a spectrum of responses from parallel or competing digitization efforts. The efforts of the Open Content Alliance, consisting of the Internet Archive, University of Toronto, University of California, amongst others, with funding support from a number of technology firms and foundations, were offered as a counterpoint to the Library Project. The Open Content Alliance example demonstrated how some of the perceived problems of the MDP were resolved, e.g. through an unwavering commitment to open access principles, but nevertheless many of the challenges plaguing earlier large-scale digitization efforts persisted, e.g. the projects were largely ad hoc, with limited funding, and lacking binding organizational infrastructure.

Notwithstanding the haze of uncertainty around the Library Project, this chapter described how public documents such as press releases and contracts, to the extent they are available, provide important clues about the sensemaking and decision-making processes of Library Project partners. For example, the press releases themselves are important and powerful artifacts of the binding commitment of institutional partners and lend legitimacy and tenacity to the project as institutional justifications cluster around shared library norms and practices such as preservation, support of academic/research mission, and improving access. The press releases also provide important clues about how institutions perceive the copyright implications of their participation — reference to preservation, educational use, and even transformation demonstrate a sensitivity to copyright and produce a selective attention that guards against possible future criticisms by controlling salience around the purpose of the Project.

The University of Michigan's sensemaking and decision-making around the Project was explored in detail, drawing upon public documents, in-depth interviews with study participants, and the UM-Google Cooperative Agreement. In concert, these sources of data suggest that pragmatic, reputational, ideological, and copyright concerns factored into the decision to partner with Google in digitizing its entire print library. In other words, the decision was the product of broad socio-structural pressures and individual social relationships amongst key decision-makers

at Michigan and Google. The scale, speed, and financial backing of Google, paired with UM's willingness to undertake a comprehensive retrospective conversion effort (including public domain and in-copyright works) and develop the organizational infrastructure to withstand the copyright risks and support collective action around the resulting resource, pushed the MDP past the obstacles that had hampered many previous large-scale digitization efforts. Digitizing everything as fast as possible and securing the possibility of cooperative work around the corpus was an imperfect strategy in some respects, but it was able to successfully fend off the creep of siloed obsolescence and that had threatened, endangered, and extinguished many previous digitization attempts.

Relatedly, we saw the importance of social relationships as providing a strong basis for engaging in vigorous, risky, action without becoming encumbered by a lengthy deliberative process likely to tap the institutional energy needed to implement and justify the project. These relationships controlled salience around the issue of mass digitization and supported justifications that focused on the broad social goals of the project (digitization was a better way to promote the progress of society) as well as digitization's inevitability. The simultaneously goal-focused but means-eschewing nature of the project, combined with its perceived inevitability, aligns with sociological theories about the relationship between technological change, structural strain, and deviance. The University of Michigan's digitization of its entire library collection, including the in-copyright works, was an example of innovative deviance, the creative use of potentially infringing or illegitimate means to accomplish the overarching goals of copyright law and academia. When existing laws collide with emerging technologies and social practices, innovative deviance can provide a mechanism by which law and institutions learn to change.

In addition, it highlights the importance and potential of policy, as exemplified by the UM-Google contract. In particular, §4.4.2 had a tremendous impact on library mass digitization and the subsequent emergence and evolution of HathiTrust not because it was determinative but because it embedded a source of value-driven generativity into the system. §4.4.2 did not cause, create, or produce the HathiTrust we know and understand today. Rather, it embedded within the MDP a source of generativity capable of opening (and closing) spaces of social practice and technical design in new, innovative, often unanticipated, and potential transformative ways. Thus, this research suggests that an increased focus on policy can help bridge "the gap between

what we need to do socially and what we can do technically" that has been a primary concern and challenge for many researchers studying the intersections of law, technological change, and emerging social practice.

Policies like §4.4.2 can also stabilize and coalesce future sensemaking and decision-making around emergent sociotechnical forms. The UM-Google agreement provided important clues and signals about individual and organizational sensemaking and decision-making around the MDP. Policy provides a window into what the parties believed was important, what concerns (or lack of concerns) they might have had, and what sorts of justifications might underlie those beliefs and decisions. In addition, when the UM-Google agreement became public, its power became demonstrable; the policy altered the ways in which other institutions negotiated, enacted, and engaged in digitization. The irreversible and visible commitment between UM and Google channeled future sensemaking and decision-making around the MDP, tightening these processes and making them more predictable, orderly, organized, and self-reaffirming. The policy became more than a simple artifact of a binding commitment between UM and Google. It became a reserve for leveraging and generating collective action.

Chapter VI: Genesis of HathiTrust

Continuing the story that started with the beginning of Google Library Project in the last chapter, this chapter describes and explains how and why HathiTrust emerged from the University of Michigan's participation in that Project. This chapter discusses how HathiTrust developed initially as a solution to an instrumental technical problem. It then discusses sensemaking and decision-making around the establishment of a shared digital repository and the eventual launch of HathiTrust in 2008.

Solving an Instrumental Problem

Once the decision was made to join the Google Library Project, scanning of the University of Michigan's library got underway. Google set up a pilot facility to begin the scanning process in a building on campus. Once the project got going, they leased an industrial facility on the outskirts of Ann Arbor where the bulk of the scanning took place. The location, the process, the technologies used; virtually all aspects of the scanning project were confidential, even to key administrators at UM.

Google's trucks would pick up the books from the library, shelf-by-shelf, stack-by-stack, and transport them to the scanning center. When new books arrived at the center, they went through a triage process whereby books whose physical condition might make them unsuitable for scanning were rejected. The rest of the books would proceed through the scanning process in assembly line fashion. Once scanning was complete, the rejected and scanned books would be reunited on the book cart, in their original order, and returned to the library with alacrity and determined efficiency.

Given the scale of the project and the sophistication of Google, the technical origins of the project were shockingly of the "garage-inventor" variety. According to one of my participants: "The particular scanning gadget was invented by Larry Page personally, at his mother's kitchen table." In addition, Google invented special library book carts that could

function easily among library stacks and corridors as well as out of the building and into the back of the particular kind of truck Google operated:

> "Google actually did something you never would have believed anybody could do. They really reinvented the library carts. The library carts that we had all been using for thousands of years were not optimal for this process and so they invented a new size and shape of library cart and stamped out hundreds of them, maybe thousands, and they fit just so in a particular kind of truck and were integrated into this process."

While technical aspects of Google's ingenuity were noteworthy to observers at the University of Michigan, it was the human-engineering throughput that they found particularly impressive:

> "There were more than dozens, probably not as many as hundreds, of work stations. What did a work station look like? It looked like a place where you set a book and photograph it, scan it. And we know (not from what I saw, good heavens!) but from the fact that you saw so many thumbs and fingers on scans, that human beings turned the pages. And the thing was very well organized in the sort of operations engineering sense. Things stayed in order so it was easy to get them back into the carts and then back to the libraries they came from and put back on the shelves. It minimizes costs. And they were processing thousands of books each day.
>
> From the very beginning, my reaction to Google has been that these people are completely unafraid of, and believe in exploiting at every margin, scale and that's exactly what this place (the scanning center) told you. It was big, lots of people doing lots of things, absolutely everybody knowing what they were up to, and really impressive throughput. If librarians had organized a mass digitization facility, it would have been only a third as fast."

At the height of Michigan's digitization, Google was scanning approximately 30,000 volumes each week, with an average turnaround time for any given work of about one week. To put this in context, a participant from Cornell University, another early leader in large-scale digitization efforts, remarked that when Cornell joined the Google Books Project several years later, Google was able to scan in two weeks the equivalent of what Cornell had scanned in the previous decade. The scale and speed of this project was unlike anything that had ever been done before.

The Library Project very quickly became an embarrassment of riches for the University of Michigan. More than that, however, concerns were mounting at Michigan about how to protect the valuable resources flooding in. U of M was getting the back-up copy of the library it

had asked for but it was not prepared to deal with the instrumental problems this back-up raised. Michigan urgently needed a place to put the scans and it lacked the technical infrastructure to securely store them.

Relatively quickly, Michigan developed an initial infrastructure for storing the digital scans. As one participant described it, the infrastructure developed as a "forcing function of the thing itself." The first instance of the digital back-up copy of the library was created fairly quickly, supported by funds provided by the Provost's Office. But according to participant interviews, everyone knew that, to do it responsibly, there had to be a second instance located offsite "so that problems that hit you aren't likely to hit them." The University of Michigan was desperate to create a second instance, somewhere else.

John Wilkin, at this time an Associate University Librarian at Michigan and a lead architect of U of M's digitization efforts, suggested that U of M create a second instance and simply give it away to another institution that would take it. Under his leadership and expertise, Michigan had already created the infrastructure to support the initial back-up copy. Now he proposed adding more storage offsite and letting other institutions in to use it. This position found support within the key decision-making hierarchy at Michigan. As another participant reflected:

> "What John Wilkin argued for was, I think, exactly right. If every institution tries to do their own version, it won't be done well. But if we have a single infrastructure, we can do it at a high quality and we can afford to bring in other people. Michigan is already supporting this thing quite well; we just need another instance somewhere else."

The problem with creating the second instance was neither technical nor social; it was financial. Michigan did not have the funds to cover the full expense of creating another back-up. It needed to find additional financial support and so began to think through a model in which different institutions build on, and contribute funds and content to, Michigan's system as a shared digital repository. The way to attract partners, in the view of participants at Michigan, was to build a quality infrastructure that was cheaper to co-maintain than build from scratch.

As one of the lead architects described:

> "The infrastructure had to be done right. It had to be done in a way that people looked at it and said to themselves, 'This is something we can't not do, but we can't afford to do it on our own, and we don't need to do it on our own. We can partner with these guys and it will get taken care of."

The evolution of conceptualizing the products of the Library Project from a second digital back-up copy, held as a dark archive, to a shared trusted digital repository was therefore not predetermined or predestined. Rather, the concept of and need for a shared digital repository emerged as a response to a series of instrumental problems set in motion by the mass digitization of Michigan's print collection.

Sensemaking and Decision-Making around the Shared Digital Repository

Michigan turned to its partners, the Committee on Institutional Cooperation ("CIC"), to ask for support in creating the shared digital repository. The CIC was viewed as a source of social capital which Michigan could draw upon to accomplish its task. As one of Michigan's decision-makers described: "We're good at sharing with each other. We like building things together. We recognize the advantages of economies of scale."

By this time, other institutions had joined the Google Library Project and were facing similar dilemmas with respect to secure storage of the scans. For example, the University of Wisconsin-Madison, a member of the CIC, joined the Google Library Project in October, 2006. Meanwhile, the CIC Librarians and Chief Information Officers had also been talking to Google about joining the project as a consortium but, according to study participants, there was a tremendous degree of nervousness amongst CIC institutions about the copyright implications of digitization. In particular, the CIC was struggling with questions regarding ownership and possession of digital copies of in-copyright works.[395]

While the CIC seemed primed to move on creating a shared digital repository in theory, it was not a priority in the same way it was for Michigan. As one study participant described it: "There was no notion in the CIC that the digital corpus would really become a huge collection, at least anytime soon, and so there was no forcing function for most of the CIC about this." As a result:

> "Discussions about the creation of a CIC shared digital repository were vague, and they were moving quite slowly. What would the shared digital repository be? Would it be a CIC project? Would it be a project of some university? Were there

[395] Ultimately, the CIC did join the Google Library Project in June, 2007 and, under the terms of its agreement, Google held copies of the CIC digital scans in escrow.

> other universities involved? Would it be a project of a consortium of universities? How are we going to determine the governance, write the bylaws, and so forth?

While conversations amongst the CIC regarding the shared digital repository were moving along very slowly, tens of thousands of scans continued to pour into Michigan's coffers each week. A serious disequilibrium resulted between Michigan and the other CIC institutions. For Michigan, the need to create a solution to its back-up problem was urgent and coupled with a desire to do it collectively at scale. In comparison, the other CIC institutions simply did not share Michigan's sense of urgency and thus the conversations about creating a shared digital repository were moving very slowly.

Reflecting or perhaps compounding tensions between Michigan and the CIC, several participants observed a clash of cultures between technologists and librarians during discussions about a potential shared digital repository. From the librarians' perspective, the hesitant slow-moving decision-making process reflected a culture of collectivism that was integral to the identity of libraries and which libraries had thrived upon for centuries. It also reflected a sense of respect and gravity for the serious role libraries play in society. For example, part of the reason libraries enjoy certain privileges under copyright law is because they can be trusted to perform their important function in a responsible, lawful manner.

In contrast, from the perspective of the technologists, the CIC librarians' deliberations about the creation of a shared digital repository were "pecking this thing to death." Another participant referred to the discussions as a "Zeno's paradox situation." Zeno was a Greek philosopher from the fourth century B.C. famous for the *reduction ad absurdium* method of reasoning whereby a hypothesis may be disproven or contradicted by carrying the reasoning to its increasingly absurd ends. In the context of the shared digital repository, this particular participant was suggesting that the CIC librarians' task of creating the shared digital repository was being broken down into an infinite number of smaller tasks, making completion of the ultimate task impossible:

> "We were 99% of the way there but the rest of the way was very clearly going to be something that we weren't going to be able to accomplish because everybody was splitting that last 1%. We had library directors saying, 'Yeah, it seems kind of pricey, maybe we shouldn't have two copies of this. The redundancy thing gains us something but we can save money if we don't do that.' But we at Michigan had already committed to that path! It was very clear to us that we

> needed to have two copies and a back-up to make it viable. And this was supposed to be the meeting where we made the final commitment!"

Michigan needed the repository to move forward but it didn't have the money to do it on its own. The CIC was caught up on the last remaining details (the 1%). It became clear to many participants that the negotiations were stuck.

The failure to reach consensus could be understood as a difference of opinion regarding salience of the mass digitization project and a future shared digital repository. Differences in sensemaking, particularly with respect to how the CIC and Michigan framed the central problem was in turn, reflected in different (ultimately incongruous) approaches to decision-making. The CIC favored a more deliberative, detail-oriented approach because it construed the essential problem and challenge on a smaller, fine-grained, scale. The CIC viewed the decision making problem as one of uncertainty and thus sought increased quantities of information to provide the source of a solution. It was essentially looking before it leapt. The CIC therefore was reasonable to prioritize accuracy and careful planning over quick action.

Michigan construed the salient features of the problem differently. It faced an immediate, significant challenge that had, in its view, a fairly straightforward and simple solution. The challenge holding up the creation of the shared digital repository was not uncertainty, but rather confusion and thus Michigan focused its efforts on reigning in the multiplicity of meanings to reduce the criticality. It was essentially encouraging the CIC to jump out into the unknown because, after all, it is difficult to predict how things will turn out ahead of time and at least this way outcomes will be generated that provide the raw material for future action. Given Michigan's perspective, it was reasonable for it to favor a vigorous action-oriented approach to decision-making.

At the end of a long day of CIC discussions and debates about the shared digital repository — the meeting where the CIC was supposed to finalize its behavioral commitment to act but was instead trapped in a Zeno's paradox — a decision had not been reached. For Michigan, inaction was not a viable outcome. If the shared digital repository was not going to move forward as a CIC initiative, Michigan had to look elsewhere for its solution. A key figure at Michigan described his position the following way:

> "I was something of a bull in a china shop. I hadn't been a University Librarian for very long. I didn't know the secret handshakes. I was a former Provost. I think I was a suspicious character in the CIC and I think that actually served the

whole project well. I tried to be friendly, and we did give a lot, but I was unwilling to be hamstrung by the norm of unanimity that meant so much to my CIC colleagues."

Less than twenty-four hours later, Michigan had its deal. The shared digital repository would be created without the CIC. After the anti-climactic conclusion of the CIC meeting, Michigan called upon an old friend, Indiana University, for help. In a matter of hours and a couple of brief phone calls between librarians and CIOs at the two institutions, an agreement was reached that they, Michigan and Indiana, would move forward with the shared digital repository on their own.

Indiana's CIO described the decision-making process as follows:

"I got a call from the CIO of Michigan saying, 'Our Librarian is going to call you because the CIC librarians are really struggling to figure this out.' Then Michigan's Librarian calls while I'm changing planes in Chicago. He knew that I didn't have a lot of time and he said:

> 'The shared digital repository governance is fucked. This is not going to happen. I can find about $600,000 per year at Michigan. Can Indiana find about $300,000 per year? We'll tell the CIC that we're going to sort this thing out — we'll be the operators of the shared digital repository and the CIC can be its first client. And down the line, we'll get this moved to something else, but this is the only way to get it done.'

And I said, 'Well, I'm very intrigued. Just let me consult my Librarian. By the next morning my Librarian had gotten the $300,000 per year and I had squared things away with general counsel. By noon the next day, I called Michigan back and said 'Indiana is in. We will sign on for the $300,000 for several years to get this thing launched.'

I told my guys in research technologies, 'Go do whatever Michigan wants. If they say: 'We're using this. We're going to use that architecture. We're going to buy from whatever vendor,' and if you think it will work and it's reasonably sensible, just do it. And they stood up and literally turned that thing on in 30 or 60 days.

And I have to credit the strength of Indiana University's IT organization because that was a bit of a countercultural moment in higher education. In higher education, even in administrative and staff positions, everybody gets a vote and everybody gets a say and you have to reach agreement on things."

The two institutions had what was described by many involved as a "charmed relationship." There was a tremendous amount of personal capital and trustful capital built up between Michigan and Indiana on many different institutional layers:

> "The charmed relationship isn't structural but personal. We have a lot of personal connections of people who have confidence in each other and in creating good outcomes together. We could jump out into the unknown, without everything figured out in advance, and trust that we would both make smart decisions and solve the obvious emergent problems together."

In interviewing the key figures at both institutions about the relationship, a number of explanations were offered in an attempt to demystify the nature of the bond. For example, participants cited a shared organizational temperament as one basis for the closeness between Michigan and Indiana: "Both institutions have people in key leadership positions who were more interested in making things happen. Not just studying it, but making it happen." Relatedly, participants cited close personal and professional bonds between key administrators at both universities. The Chief Information Officers,[396] the University Librarians, the Provosts, General Counsels, Presidents, were all committed to shared values and these commitments served as powerful links between figures at each institution. One participant explained, "We are kindred spirits and we complement each other." Another participant said, we share the sense that "great public research universities have to act now or risk becoming less relevant. That is what drives us." Michigan and Indiana also share a history of successful collaborations and projects. Sakai,[397] a learning management system that has been adopted by over 350 colleges and universities around the world, and the more recent Unizin Consortium[398], working to create a digital learning landscape built around cloud-based infrastructure based on open technology standards, are but two examples.

The tremendous social and organizational capital shared by Michigan and Indiana supported vigorous decision-making and a willingness to jump out into the unknown. One of the participants said:

> "Institutions feel like they have to be able to answer every possible foreseeable question before they take the first leap. And so that reservoir of personal capital really helps a lot. It's remarkable that from a phone call at about 3:30pm in

[396] The University of Michigan did not have a CIO but an Associate Provost for Academic Information and Instructional Technology Affairs and, later, Vice Provost for Academic Information.
[397] Sakai Project, https://sakaiproject.org/
[398] Unizin, http://unizin.org/

> O'Hare airport to noon the next day, Indiana University would dig up and make an exceptional $300,000 per year commitment and take on the ire of some of our colleagues by stepping out and saying: 'We will be the operator of the shared digital repository.'"

When Michigan and Indiana returned to the CIC meetings the following day, and informed its members of their plan to create the shared digital repository, shockwaves rippled throughout the room. One participant described the announcement, saying:

> "Oh my God, one day, the CIC is going to do this and the next day, it's just Michigan and Indiana. You can imagine, I mean, whoa, that was like, 'Hey, what happened here?!' It was a bomb."

For those in attendance at the meeting when the announcement was made, the episode was shocking. Stories of that meeting still persist today. Many people were very unhappy about it at the time and some of those feelings linger on:

> "Librarians have a very collectivist culture and for someone to break out and do something this way was not only debatable as a strategy, it violated cultural norms of how librarians tend to do things and it violated the governance structure of the CIC."

The person who actually made the announcement recalled it this way:

> "We said, 'Indiana and Michigan are going to cover the entire costs between the two institutions and if the CIC institutions want to come in now, they can be secondary partners and will pay for part but will not have a seat at the table in the same way.' And it was a catastrophic falling out. One of the library directors turned his back on the table. Literally turned his back to me, and lots of people were very unhappy about it."

Sensemaking around the decision to push ahead with the shared digital repository offers a useful perspective into that decision and provides a conceptual link to the innovative deviance framework. Michigan engaged in act of innovative deviance by breaking from the shared norms and practices of the library community and the CIC governing structure to strike out on its own, with the help of Indiana. The impetus of this act was a perceived mismatch between the shared goals of the library community — establishing a shared digital repository — and the means of accomplishing that goal — the CIC's highly detail-oriented deliberative process that privileged egalitarian consensus-building over rigorous action. By breaking ranks with the CIC and forming a bi-lateral partnership with Indiana University mid-negotiation, Michigan used

illegitimate means to achieve an accepted legitimate goal, ensuring that the digital repository which it so desperately wanted was built.

A short time later, the key participants from Michigan and Indiana met in Indianapolis to discuss strategies for moving ahead with the shared digital repository. Over the course of a day, the group chose a name for the repository, identified a strategy for getting buy-in from additional institutional partners, sketched out basic details for what the repository should look like and how it should operate and agreed on which aspects of the project could be shelved until some future date.

Creation and Launch of HathiTrust

Rudyard Kipling's short story, "Toomai of the Elephants," begins with the following short poem:

I will remember what I was, I am sick of rope and chain–
I will remember my old strength and all my forest affairs.
I will not sell my back to man for a bundle of sugar-cane:
I will go out to my own kind, and the wood-folk in their lairs.
I will go out until the day, until the morning break–
Out to the wind's untainted kiss, the water's clean caress;
I will forget my ankle-ring and snap my picket stake.
I will revisit my lost loves, and playmates masterless! [399]

It was Kipling's story that inspired the name "HathiTrust." During the Indianapolis meeting just described, Michigan's Librarian had Kipling's story floating around in his head and, when it came time to select a name, he suggested it as a source. Initially, he thought they should name it "Toomai," but that is the name of the human elephant-handler in the story and what he was drawn to, as the poem suggests, was the elephant or, more precisely, the idea of the elephant. He explained his sensemaking as follows:

> "I somehow liked the idea of the library and libraries as being an elephant. Elephants never forget. Elephants are big and strong. Elephants carry heavy burdens. And it's inspired by Kipling, so it's a semi-classical reference, shows that we're sort of literate."

[399] Kipling, R., & Howe, I. (1982). *The Portable Kipling*: Penguin Books.

While it did not surface in his description, the elephant in Kipling's story also represents liberation, breaking free from constraints and returning to one's true nature. One can speculate as to how the decision to digitize the entire library may have also seemed like a step toward liberating the ideas, knowledge, and learning contained in the pages of its books in line with the utopian vision of making all of the world's books accessible online to everyone.

While the elephant in Kipling's story was a source of inspiration, "elephant" was not perceived as a compelling name for a shared digital repository. The Hindi word for "elephant," which is "hathi," seemed like a much more appealing alternative.

As Michigan's Librarian explained his idea behind adopting the "hathi" moniker, Indiana's CIO was online conducting a domain name search. After discovering that the "hathi" domain was already taken, they brainstormed for a few minutes, tacking on other words, until they found one that both made sense and was available as a domain name. Everyone agreed that "HathiTrust" was an improvement on the original concept: "We have all these works that are in trust for the future, guarded by, carried by, in my imagination, this wise, infinitely long-lived elephant." The shared digital repository now had a name: HathiTrust.

At this point, the thing that would soon be formally introduced to the world as HathiTrust functioned as a technical solution to Michigan's instrumental problem: it was a safe, secure place to store digital scans. Some of the more challenging aspects of its development still lay ahead. Namely, how Michigan and Indiana might navigate the organizational and institutional fallout stemming from their innovative deviance and begin building consensus and partnership around HathiTrust.

As one of the lead architects described, once the technical problem was solved:

> "My first reaction was 'What will all of the people who were involved with this do? Well, they'll hate us. They'll hate Michigan. Anybody we try to bring in will hate us because we're so hegemonic. So I wasn't worried about the technical side. Michigan and Indiana had that covered. I was worried about the organizational side."

With respect to the CIC institutions, once the initial shock of Michigan and Indiana's announcement wore off, and discussions within the CIC resumed, the CIC did agree to come in as a partner. But attitudes toward Michigan and Indiana remained mixed for some time. Some CIC members recalled thinking: "there's Michigan again, taking the lead on things and not waiting for everybody else. Why can't those people in Ann Arbor be more collaborative?"

Participants from Michigan and Indiana noted that the resentment they felt from certain quarters of the CIC for going forward with the digital repository without asking, in their words, 'Mother, may I?"

One of the ways that Michigan and Indiana sought to diffuse some of the backlash was to make it very clear that "HathiTrust would always view the CIC as a founding member." Even though the CIC initially came on board as a secondary partner, the first client of Michigan and Indiana's repository, the founders noted:

> "We ultimately gave them a seat on the board and on the executive committee, and that turned into two seats in time. So I think they've gotten everything they would have gotten, but it was the thing that caused them to move forward."

There was initially no forcing function for the CIC with respect to creating the shared digital repository. Michigan and Indiana's innovative deviance ultimately became that forcing function. The act of circumventing or abandoning CIC norms and governance and proceed independently toward the goal of creating the digital repository was the bomb that caused them to move forward.

Another thread of Michigan's early organizational strategy around HathiTrust was to bring in additional institutional partners outside the CIC. The University of California, in particular, was pursued quite vigorously. Michigan's interest in the UC system was motivated by reputational and geographical concerns. It was also tied to personal relationships and investments of social capital that had been underway for years prior.

Conversations between individuals at the University of Michigan and the University of California had been ongoing with respect to using the Google scans as a lever for collective action around print management of the "collective collection." Those discussions had even gotten as far as an ultimately unsuccessful grant proposal to the Andrew W. Mellon Foundation, co-authored by a Michigan librarian and a key administrator at the California Digital Library entitled: "Mass Digitization of Research Library Holdings and the Public Good."[400] In the memory of one of those participants, discussions between Michigan and California stalled after the proposal got a "serious rebuff" from Mellon, a disappointment compounded by the tragic demise of a key UC administrator and subsequent "collapse in the University of California administration." The upheaval at the UC led to organizational restructuring which pulled the

[400] On file with author.

CDL administrator who had co-authored the grant proposal into UCOP (UC's Office of the President). Needless to say, the collaboration was sidetracked.

When Michigan and Indiana were contemplating HathiTrust, a newfound attention was paid to the UC. Key participants hoped the prior collaborations retained sufficient warmth to spark a reconnection. Sensemaking around a possible UC partnership was described by one of the key figures at Michigan as follows:

> "We need to bring in the University of California because the CIC produces about 10% of the PhDs, and the University of California produces another 10% of the PhDs. If we got 20% of PhD construction it will be very hard for the others not to join. Once the two sort of biggest institutionalized players are in, we'll get there."

When the Michigan librarian responsible for those early discussions with the CDL reached out to them anew about the possibility of joining the nascent HathiTrust, he discovered that, not only had his former collaborator moved on, the CDL had a completely new roster of personnel. All of the social ties Michigan hoped to capitalize upon had unraveled. Moreover, there had been apparently zero communication (or zero organizational memory) within the CDL regarding the mass digitization grant proposal:

> "The former head of the CDL hadn't shared our conversations with anybody else at CDL so when I came back a couple of years later and started trying to move things forward with CDL, I shared with them what we had written and they were all stunned that the conversations had gone so far and they didn't know."

Gaining the University of California's commitment was a formidable challenge beyond the broken social ties. As previously described, the University of California and the California Digital Library had been leaders in early digitization efforts. They had been an active participant in the Open Content Alliance. They had partnered with Microsoft, Yahoo!, the Sloan Foundation, and the Internet Archive prior to partnering with Google on the Library Project. Sure, Michigan was an innovator and leader in digitization, but the UC viewed itself in some ways as in a tier to itself. A participant at the CDL reflected that the CDL saw itself as "the intersection of the Venn diagram of digitization:"

> "We had a great sense of the big picture of what people were working on, how far they were, what kind of challenges they had, how they're thinking about access and preservation in particular. We really were in the center of the communication and social side of digitization efforts."

The University of California's self-appraisal and objective positioning within the social ecology of digitization had a number of implications, positive and negative, with respect to Michigan's desire to bring them on as partners of HathiTrust. On the positive side, CDL had experienced frustration over the lack of organizational infrastructure on some of its prior collaborations, perhaps most notably with the Open Content Alliance. HathiTrust seemed to offer a solution to this problem. In addition, while some of its previous partnerships (with Google and the Internet Archive) provided a platform or locale for access and preservation, participants at CDL noted:

> "Neither situation really served the needs of research libraries. We were accomplishing digitization but we were not accomplishing the infrastructural aspects the libraries needed."

Relatedly, there was also a sense that private companies could not be trusted or relied upon to manage the data in a way that was consistent with library values or practices:

> "The academy traditionally tries to solve problems like each one of us are an island. But the digital goes to scale, it favors scale. Either we figure out how to create scale ourselves in ways that we can steer in our interest, and take some advantage of the economics of it, or others will create scale and they will manage it in ways that are not in our interest."

Weighing against a possible partnership in HathiTrust, the University of California intended to host its own content and was working on creating a technical infrastructure to support its institutional goals. The group working on developing a CDL-based preservation scheme for dealing with the Google scans was not impressed by Michigan's project which they initially viewed as a regional, CIC project.

Furthermore, the UC system is large and, by many accounts, highly bureaucratic. Consensus is often difficult to come by. Eventually, the CDL did approach the University of California's governing board and advocated that the University of California join HathiTrust but the board was not immediately convinced. They asked the CDL to write an internal memo explaining why the University of California should join. Meanwhile, the Michigan librarian who had been spearheading the partnership wrote a formal letter entreating the University of California to join HathiTrust which he delivered to his old collaborator (now at the UCOP). The University of California still did not act.

In the view of some people at Michigan, the situation was getting dire:

> "It was easy to get the CDL people to join because this was right up their bailiwick. But it was clear to me from the start that this wasn't going to go anywhere unless we got Berkeley and UCLA on board. They are by far the biggest pieces of the UC system in terms of campuses, and they have stopped things repeatedly in the past. If Berkeley and UCLA gang up they are essentially invincible. So, we didn't necessarily need them to say, 'We're in. We love it.' But we at least had to get them to say, 'We won't fight it.' That took about a year."

The University of California did, eventually, decide to partner in the shared digital repository that would eventually launch as HathiTrust. Its sensemaking around the decision was essentially two-fold. The first reason was financial. Like Michigan, and the growing list of Google Library partners, the University of California faced the technical problem of creating the infrastructural support for its back-up copy. While the UC had the technical ability to meet this challenge, money to fund the project was another question. Thus, according to my interviews, the primary reason the UC decided to join HathiTrust was economic.

The second reason for joining revolved around issues of salience and control. Although the UC sat on Michigan's offer for some time, word eventually began to trickle through that Michigan, Indiana, and the CIC were about to announce the launch of HathiTrust. If the University of California wanted to join as a founding member, have a seat on the board and executive committee, and take an active role in shaping the path of HathiTrust, they needed to make a decision. As it had done with the CIC's deliberation process, Michigan forced the University of California's hand by saying: 'We are moving forward with or without you. If you join us now, we'll give you a seat at the table, but if you wait, you won't get that level of status within the organization.' The strategy was a gamble that paid off. It prompted a binding behavioral commitment on the part of the UC and enabled Michigan to, once again, push the repository forward despite a stagnating decision-making process.

But this was, ultimately, a short-term strategy for generating acquiescence to build critical mass around HathiTrust's creation. For HathiTrust to have lasting success, its founders recognized that Michigan could not operate the repository as a dictatorship. Michigan needed to play the long game and that meant putting in place measures that would ensure that control of HathiTrust would be given over to the collective. As one of my participants very eloquently described:

> "The library community is very catty. Because they've been deprived of power for so long they engage in horizontal violence at the local level. So, the number one complaint would be that Michigan is doing this thing that really benefits us so that they can control us. This was going to be a huge issue. And so we had to give HathiTrust over to the members of the community, so that they could settle upon what HathiTrust might become. We couldn't say 'This is the direction it's going to go' because, even if we were right, it would be prima facie evidence that we were drunk with power, and mad, and taking them where they didn't want to go. This is all part of that old aphorism: 'Those who lead the revolution shouldn't be the ones who build the subsequent state.' And I think that was right in this case. We had a vision, which was that we really needed to back-up our digital scans, but the rest had to be settled by the library community."

When the press release went out on October 13, 2008, announcing the official launch of HathiTrust, it was introduced as a shared digital repository jointly founded by the 12-university consortium known as the CIC and the 11 libraries of the University of California system. There was no specific mention of Michigan or Indiana (other than as CIC members). Much to the frustration of Michigan's public relations department, which wanted HathiTrust to bear Michigan branding, Michigan's fingerprints were already receding from the HathiTrust creation story.

Receding into the emerging HathiTrust collective was a crucial foundation for HathiTrust's future success. A librarian at Michigan who joined the organization just after HathiTrust's initial launch reflected:

> "When you're at Michigan, you see what's going on here. It wasn't until I was at a CIC meeting and saw people with HathiTrust stickers on their computers and heard them referring to HathiTrust as 'We' rather than as 'Michigan' that I realized there was already this broad sense of collective action being expressed around HathiTrust. It was really an amazing thing to see."

While a sense of collective ownership and control over HathiTrust began to emerge from the outset of its launch, the next chapter describes how the transition to collective governance followed a more measured and deliberative process. HathiTrust's lead architects continued to play an important role in its early development as the organization continued to grow and transform from a Michigan initiative to a semi-autonomous collective organization.

Conclusion

The period between the UM's decision to join the Google Library Project in late 2004 and the official launch of HathiTrust in 2008 reveals several key insights into the processes of sociotechnical transformation. For example, tracing the evolution of the mass digitization

project from something initially conceived of as a dark archive to a collective shared digital repository demonstrates that change and transformation emerges through complex, dynamic, non-linear, often unanticipated actions and interactions amongst technology, social practice, and law.

The overwhelming influx of scans, upwards of 30,000 volumes per week, vastly outpaced all earlier large-scale digitization efforts. The scale and speed of the project revealed a failure to fully prepare for and anticipate the need for a secure, multi-located technical infrastructure for storing the digitized content. The infrastructure which would eventually form the technical basis of HathiTrust emerged as a technical solution to an instrumental problem. But the technical and financial issues facing Michigan were quickly overshadowed by the organizational challenges it faced.

This chapter demonstrates how the innovative deviance framework can provide meaningful clues into the organizational aspects of HathiTrust's emergence. While Michigan and the CIC shared a common goal of creating a collective digital repository, a disequilibrium arose with respect to the means of accomplishing the goal. Michigan favored a rigorous action-oriented decision-making approach whereas the CIC's approach was deliberative, focused on fine-grained details, guided by principles of egalitarianism that had been core values of the library community for centuries. When Michigan decided to circumvent or abandon the CIC's governance and create a shared digital repository with Indiana University, it was engaging in innovative deviance — making creative use of illegitimate means to achieve a desired legitimate end. That innovative deviance, in turn, became the forcing function for the rest of the CIC to get on board with the shared digital repository. It was, as one of my participants articulately described, the bomb that got things moving forward.

While the disequilibrium and organizational strain prompted Michigan's innovative deviance, the willingness of Indiana University to partner with Michigan in this risky behaviour reflects the importance of sensemaking and close personal bonds and affinity. In particular, this research showed that shared values reflected a confluence of salience around the core issues. In addition, strong and trustful personal capital provided the foundation necessary to engage in potential risky behavior guaranteed to raise the ire of colleagues at other institutions. Relatedly, and perhaps most importantly, this work demonstrates the fundamental importance of similarity

in decision-making approaches. Both Michigan and Indiana favored vigorous goal-oriented action over a process-based look before you leap strategy.

Social relationships and differences in decision-making approaches also played key roles in terms of Michigan's negotiations with the University of California. Despite having worked closely on digitization issues in the past, a change in personnel dissipated much of the social capital and common that had been established in years prior. In addition, differences in institutional self-perception about what its proper role should be in the transformative process influenced the negotiations. Despite coming to the Google project years after Michigan, the University of California situated itself at the intersection of the Venn diagram of all the various digitization efforts and projects. Its expertise and leadership made it reluctant to sign on to something it perceived to be a "Michigan" or "CIC" project. For the shared digital repository to be "the digital repository," Michigan believed it needed the UC system. Michigan courted the UC for a year, trying to convince it to join, before ultimately gambling with an ultimatum: join us now as a full partner with the power to shape the initial evolution of the repository, or take a chance on forging your own path. The gamble paid off and the University of California agreed to join Michigan, Indiana, and the other CIC institutions.

The final key takeaway from this chapter deals with the organizational fall-out of innovative deviance. As participants noted, once critical mass was established to ensure that HathiTrust could have a successful launch, the main issue became how Michigan could ameliorate some of the relational damage caused by its "bomb" and ultimatum. Its strategy, which proved to be incredibly effective, was to erase its organizational fingerprints from HathiTrust. Thus, when HathiTrust was launched it was introduced as a joint initiative of the University of California and the CIC. Moreover, by its launch, its origination as a technical solution to an instrumental problem had been subsumed in a broader values-oriented goal deeply tied to core library functions of the providing access to and preservation of knowledge. The press release for HathiTrust's launch may even provide some of the tenacious justifications necessary for coordinated sensemaking around what HathiTrust could or should become. It described HathiTrust as a collective organization that coordinates preservation and use of digitized content for collective action within the library community on long-standing issues, a retrospective justification that, as the following chapter describes, quickly morphed into a self-fulfilling prophecy.

Chapter VII: Reification of HathiTrust

> "Real isn't how you are made. It's a thing that happens to you. … It doesn't happen all at once. You become. It takes a long time."
>
> -Margery Williams Bianco,
> *The Velveteen Rabbit.*[401]

This previous chapters describe and explain how HathiTrust came to be. This chapter describes and discusses an important phase in HathiTrust becoming the HathiTrust we know today. This is the part of the story where HathiTrust is reified, where it becomes real, where it creates a new reality. The process of reification is essentially a social process and therefore this chapter will focus on the social aspects of HathiTrust's development and evolution.[402]

The term reification has a number of other connotations beyond the "social process of becoming real." Particularly in Marxist and postmodern philosophy, reification is criticized for its tendency to oversimplify and collapse an object into a single phenomenon or form that can then be studied, controlled, etc., instead of acknowledging that the meaning of an object emerges from an unbounded, continuously evolving set of mutually-constitutive processes, experiences, and participations.

My approach to this phase of HathiTrust's development borrows both Weick's sense of reification — the social process of becoming real — and the more critically-oriented interpretation of reification just described. The process of HathiTrust becoming the thing we know it to be today did grow out of a collective sensemaking process and this chapter describes and explains key facets of that process. But in also revealing the dark history of HathiTrust's emergence and evolution, and the murky, still unresolved and contested intersections, this work also pushes back against a static, monolithic understanding of what HathiTrust is.

Picking up where the story left off in the last chapter, the first section of this chapter describes HathiTrust's early sensemaking and decision-making in forming its governance

[401] Bianco (1926).
[402] Weick (1995: 23).

structure and describes the evolution of its path, purpose, and organizational trajectory. The second section shifts into a description of how HathiTrust grew and transformed in scope and functionality from a shared digital repository to a more complex, multi-faceted service organization. This chapter concludes with a discussion of a milestone event in the evolution of HathiTrust, its Constitutional Convention, an event which formally marked HathiTrust's transition into a (semi-)autonomous collective organization.

HathiTrust's Early Governance

At its inception, the governance structure of HathiTrust consisted of an Executive Director — an Associate University Librarian at Michigan who had served a vital role in Michigan's digitization efforts and was a lead architect of HathiTrust — and a small executive committee comprised of the University Librarians and Chief Information Officers at the University of Michigan and Indiana University. When HathiTrust was officially launched, representatives from the CIC and the University of California also joined the Executive Committee.

While few things were certain in terms of the broader social implications of mass digitization and the creation of HathiTrust, the Executive Committee had three clear priorities from the outset. First, it wanted to grow the membership of HathiTrust:

> "We recognized upfront, even though it was something new to all of us, that it had such potential and was going to get huge. It was clear, almost from the beginning, as everyone started wanting to sign on. We decided, let the membership grow."

Much as Michigan knew additional partners would be needed for the shared digital repository to be a success, the Executive Committee knew its membership would need to grow for HathiTrust to reach its full potential (whatever that may turn out to be).

Second, the Committee agreed that legal ownership of HathiTrust would remain with the University of Michigan. In other words, while HathiTrust was launched jointly by the CIC and the University of California, it originated as a project of the University of Michigan and thus its status, as a legal entity, remained with Michigan. The possibility was there that, at some point, Michigan might transfer legal ownership to some other institution or otherwise undertake action to make HathiTrust a fully sovereign entity. But there was never much interest or organizational momentum in pursuing that end:

> "HathiTrust was initially structured this way as a matter of convenience but, over time, there was never a real push to transition legal ownership of it to some other entity. There was a sense that Michigan was better at handling risk, or that it is better for the other member institutions if Michigan (rather than they) bore the risks."

While participants noted that the doctrine of sovereign immunity was not discussed when the mass digitization project was first undertaken, it was discussed by the Executive Committee as an additional justification for Michigan retaining legal ownership of HathiTrust. Recall that, as a state institution sovereign immunity was perceived as limiting Michigan's potential liability for monetary damages from copyright infringement. Private institutions, on the other hand, fall outside the protections of sovereign immunity and thus, as several participants speculated, Michigan's ownership "gives the members of HathiTrust some comfort."

In accordance with its legal status, Michigan's Office of the General Counsel entered into bilateral agreements with each of HathiTrust's member institutions. These agreements set forth the essential terms of what activities and uses would be permissible with regard to the HathiTrust collection. The vision of HathiTrust evolved rather quickly. At the outset:

> "HathiTrust was envisioned in the beginning as a storage locker on the outskirts of town for digital scans. Michigan would own the storage locker, and there would be some basic rules, but what you do in your storage locker is your business."

This was not, however, the model that was ultimately adopted. The membership's autonomy with respect to what could be done with HathiTrust's collection was limited by what Michigan allowed, and what Michigan allowed was primarily a function of its contractual obligations under the UM-Google Cooperative Agreement and risk assessments made by Michigan's Office of General Counsel. Ultimately, as one of the founders of HathiTrust reflected, "Michigan's permissions were pretty limited." Michigan put a ceiling on what uses were possible and individual member institutions could adopt their own policies within that space in accordance with their own objectives and levels of risk tolerance.

The third thing the Executive Committee agreed upon at the outset what that control over HathiTrust should not remain in the hands of its progenitors, now embodied in the Executive Committee, but should instead be passed on to the collective membership. It was understood that HathiTrust should be "a broad collective thing, not "us" telling "you" what to do." One of the participants in my study explained:

> "We knew we had to find a way to shift the control of this organization from the CIC and UC because the academic library community will hate us in the same way they would have hated Michigan and Indiana."

With respect to Michigan's role, the move to create an innovative initiative and then essentially give it over to a broader collective was part of its institutional identity:

> "Michigan is powerful in its way, but the way we exert our power is less overt than other institutions. Our bias is, we'll get this ball rolling but we're not going to stick around to control things."

One of the challenges that was not discussed by participants but which may be obvious in hindsight is that there is an inherent tension between the Executive Committee's second and third goals. Michigan simultaneously retaining legal ownership of HathiTrust and giving control of the organization over to the collective may be a source of friction where the risk tolerance of Michigan is incompatible with the organizational goals of HathiTrust. At the time these goals were agreed upon however, in fall, 2008, none of the participants viewed this as a challenge, in part, because Michigan was still very much an innovative deviant with a relatively high and institutionally homogenous risk tolerance.

Setting aside that tension for the time being, the transition of control from the Executive Committee to the broader HathiTrust membership raised a number of pragmatic challenges. In fact, HathiTrust's inaugural Executive Director reflected that "the biggest change in HathiTrust has been its move to a more formally defined shared governance." Two concerns, in particular, seemed to dominate the Committee's sensemaking around this issue.

First, the Executive Committee was concerned that giving control of HathiTrust over to the collective too quickly might prematurely stunt, and possible extinguish, the energy and momentum necessary to establish HathiTrust as a viable, resilient entity. On an instrumental level, the Committee knew it would take time to get HathiTrust off the ground and therefore it wanted to reserve control for a period of time. One of the Committee members explained it this way:

> "We will get there, but it can't happen overnight because we are doing stuff and we need to keep doing stuff. It wasn't blackmail or holding anything hostage. It was the plan. We don't have good models for this kind of thing and we needed time to figure stuff out. We need time to make sure this thing is solid and reliable, that the right technology is there, and the right people are working on it between Michigan and Indiana."

The Executive Committee foresaw a series of tasks ahead that would need to be accomplished in order to get HathiTrust up and running. It needed to figure out how the infrastructure would work. It needed to have a model for attracting and supporting (technically, organizationally, and financially) its growing membership. It needed to determine what kinds of services it might offer, and so forth.

Fresh off the previous decision-making standstills that had bogged down negotiations around the creation of the shared digital repository and HathiTrust, there was a compelling sense that a smaller, less formal, and looser governing structure like the Executive Committee could more easily maneuver the multitude of decisions on the horizon. One of the participants described it as follows:

> "It was good that we didn't really have to have a real formal organization at the beginning. The Executive Director put the issues on the table, provided guidance in dealing with them, and ultimately did most of the work. The Executive Committee was nimble. We almost always agreed on everything and we did a lot very quickly. We didn't try to dot all of the "i"s and cross all of the "t"s the very first day. If we had, we would have been distracted because, as we discovered, doing that requires a tremendous amount of work."

Delaying transition to a formal collective governance structure was therefore seen as preserving or buffering a much-needed incubation period during which time HathiTrust could develop, innovate, and collaborate on the products and services it might offer, and build and test a robust supportive infrastructure.

In addition to concerns around stymieing HathiTrust's growth and potential, a second justification emerged for the small Executive Committee retaining initial control. Specifically, some members of the Committee expressed concern that, if HathiTrust were turned over too quickly to the collective, it might make it vulnerable to usurpation by other institutions. The fear was that other powerful institutions might seize the opportunity to assume a dominant role in HathiTrust's emerging governance and undermine the potential for a truly collective, democratic governance model. As one of my participants explained:

> "The question became one of control. Who is going to get to call the shots? And we knew right from the start that certain institutions, Harvard comes to mind, are going to want to control this from the get-go: 'Either we control it or we don't play.' That's an extreme version of Harvard's position, but that was sort of their position. My guess is that if HathiTrust had been started at Harvard, there would not have been discussions about how that institution needed to get out of the way for HathiTrust to be successful. It would have been a Harvard thing from the get-

> go: 'We're Harvard. Don't you want to join?' They wouldn't try to stop HathiTrust but they might withhold membership as their most extreme action. But we knew we would get enough people to sign on that it would be a going concern and eventually everybody would kind of go along."

Delaying the transition to collective governance was justified as a way of enabling the Executive Committee to safeguard future collectivism against institutions who may seek early membership as a way of vesting a controlling interest in the organization.

Therefore, in order to grow, cultivate, and protect the membership, and have time to "get the ball rolling" on the multitude of decisions ranging from the day-to-day operations to defining its broader goals, a target date was set, at which time the Executive Committee was pass control of HathiTrust over to the collective. Michigan and Indiana had committed funds to support HathiTrust for a period of three years, and thus three years from the date of its official launch, HathiTrust's members would convene for a meeting — a Constitutional Convention — whereby the membership would take control of HathiTrust:

> "It was a very loose structure at first, but the idea was always there that, in three years, we will have a Constitutional Convention, a date certain by which we would come out with a map to a real governance structure, with a governing board, bylaws, all of those good things. This new governing structure for HathiTrust would be implemented and the existing controlling structure would step back, back away, and say: 'You guys take it.' The Convention would gather all of the participants, the supporters, the members of HathiTrust. By then we will have a better idea of cost, what the issues are, and have a better ability to look at the kinds of projects that might form around this kind of a collection."

The Constitutional Convention was therefore envisioned as an organizational rite of passage marking HathiTrust's transition into a full-fledged collective. Between this initial sensemaking and decision-making by the Executive Committee and the Constitutional Convention three years later, HathiTrust went through a tremendous period of growth and transition which will now be described.

Emerging Features

In retrospect, providing access to the HathiTrust collection may seem self-evident, but as earlier sections of this thesis explains, the digital corpus largely comprised of the Google Library Project scans was initially conceived of as a preservation repository or a dark archive. Because

there was no purported intention to make the scans accessible, the initial design of the infrastructure that became HathiTrust was not oriented around facilitating access.

However, once the instrumental challenge of securely storing the digital scans was solved, the attention of HathiTrust's progenitors quickly turned to questions of access. As one participant described:

> "In the early days, we were really thinking about preservation. That's what got this going. Over time, it began to occur to us, the excess power this resource had, and we began to build in access as well."

Beyond preservation, the key uses that emerged from this perceived "excess power" were access for non-consumptive research such as full-text search and copyright rights determination, improved access for print-disabled patrons, and the beginnings of enhanced access to so-called orphan works. Before describing sensemaking around those emerging uses, this section will address some broader issues, challenges, and tensions raised around preservation and access.

Relationship and tension between preservation and access

In the library and archive community, preservation and access are often difficult to parse. Practitioners view them as mutually constitutive. Access is not particularly meaningful if you cannot rely on an item being preserved and organized in a way that makes it discoverable and useable. In addition, as one of my participants who is a lead archivist explained:

> "Preservation without access isn't preservation. The idea that somehow you can have material locked away and maintained in a pristine form with no access just isn't very practical in terms of preservation needs for no other reason than repositories aren't gonna wanna spend a lot of money preserving stuff that no one can take a look at."

While, as a matter of librarian and archival practice, preservation and access may be functionally entangled, in terms of copyright law, they receive different treatment. As Chapter II described, libraries and archives enjoy specific exemptions around making copies for preservation purposes that are explicitly distinct from specific exemptions around making copies for lending and patron use. Once a library makes a lawful copy for preservation purposes, it is not obvious what, if any, conditions might trigger lawful access to the copy, which is to suggest that while preservation and access may be deeply entangled with respect to social practice, they are held separate with respect to copyright law.

Nevertheless, once the institutions had safely stored the scans in the shared digital repository, new uses and justifications for accessing the scans quickly emerged. The lead architect of a competing digitization effort speculated that the transition from a preservation justification to preservation-plus-access justification may have been more deliberate than emergent:

> "I'm not saying that HathiTrust's progenitors are being disingenuous, but if the talk focuses on preservation it skirts all sorts of issues that are thorny. You can kind of prime a "motherhood and apple pie" perception about what you're doing and you don't have to deal with the other issues. If you're really about preservation, then the preservationist wants a physical object — these books that were coming out of these libraries were safe and many of them were held in off-site repositories. So the physical materials were quite safe."

Reading between the lines a bit, the participant here is problematizing preservation as a valid justification for engaging in the digitization of materials for two reasons. The first reason is that the source material, the books, were already well-preserved. They were not damaged, deteriorating, lost or stolen, stored in a format that had become obsolete as §108(c) requires.[403] In addition, many preservationists believed that the Google scans were nowhere close to preservation quality. One of the participants in the study whose institution was an early leader in digitization and joined the Google Library Project in 2007 said:

> "We never viewed them as being preservation scans. Michigan always used a preservation argument but we never did. I still don't think the quality of the Google scans would meet anyone's standards for preservation access. Certainly not the standards we cite."

Interviews with participants at Michigan suggest that preservation was a key concern not necessarily for the specific reasons articulated in the Copyright Act but for a more general reason that libraries' print collections were all dying a slow death from "acid overdose." Interestingly, *digital preservation*, rather than preservation of the print collection was also offered as a justification:

> "You can't put all of this money into scanning this stuff and then not store it in a way that you don't lose it. That's insane! So that had to be fixed."

Preservation as a justification for undertaking the original digitization and preservation as a justification for creating the shared digital repository and, later, HathiTrust were being

[403] Some of the books were in various states of slow deterioration, "dying from acid" as one librarian put it.

conflated. The conflation might reflect disingenuous motives, as the earlier participant suggested. But it may also reflect the emerging reality that, as a shared digital repository, HathiTrust was merging digital content of varying provenance — much of it was scans of public domain material generated from their own institutional digitization efforts and/or collaborations with the Internet Archive — and merging institutional justifications — many HathiTrust members cited digital preservation as a key motivation in joining HathiTrust.

While speculation as to the sincerity of motivations is a potential concern, an arguably more troubling implication is that the Copyright Act's distinctions between preservation and access in §108 are wildly mismatched to the experiences and practices of actual librarians and archivists. Digital technologies make the distinctions between preservation and access even less meaningful for libraries and archives, particularly when we consider that access comes in many guises which serve distinct purposes. For example, a human accessing the digitized work to read or download it is arguably quite different than uses such as a machine accessing the digitized work to index and mine its contents or translate it into a format which a blind or print-disabled patron can access.

These distinctions became triggers for subsequent decision-making by the Executive Committee around questions of access. It didn't take long before new questions emerged about what uses might be legally possible, and socially viable, with respect to the digitized corpus. The HathiTrust Executive Committee honed in on the possible ways they facilitate access to public domain content and shine a light on the dark archive. The additional uses that emerged with respect to the copyrighted works during this period were non-consumptive uses such as full-text search and access for copyright determination, and the provision of copies for print-disabled patrons. The University of Michigan began a project that would have provided enhanced access to so-called "orphan works" but that project was suspended before any works were actually released.

Access to Public Domain Content

One of the first key milestones in the evolution of HathiTrust was the creation of its graphical user interface. The technical ability to view content gave rise to the (perhaps obvious) follow-up question: "Which content will we provide access to and what kind of access will we provide?"

The decision to provide access to the digitized public domain materials was uncontroversial within the Executive Committee. Neither the Copyright Act nor the contract with Google seemed to forbid HathiTrust from making those materials publicly available on HathiTrust.org. Providing the technical and organizational infrastructure to link preservation and access was fairly simple and its implementation was relatively straight-forward. As a result HathiTrust members are able to link to, read, and download the full text of public domain works. In addition, members of the general public can read the full text of public domain works and have limited download ability (works may be downloaded a few pages at a time).

Restrictions on the public's access to digitized public domain content stems from the agreements Library Project partners entered into with Google. These restrictions have and have garnered criticism. Brewster Kahle, the founder of the Internet Archive, has been a particularly vocal detractor. Chief among his complaints are the download restrictions placed on the public domain materials which makes it very difficult (perhaps impossible) for non-HathiTrust members to conduct non-consumptive research on the corpus:

> "I don't think organizations and libraries like HathiTrust got the timing right. We've got an Internet and the computer technology that allows research to be done in a very different way now. Let's let it happen. In fact, let's not just *let* it happen, let's *make* it happen."

In addition, Kahle argues that the public domain works are secured in ways that make their content less discoverable by non-Google search engines: "Global search engines are unable to crawl and index them in new and different ways and therefore they can't be referenced and pointed to in that style." At the very least, says Kahle, the public domain should be opened up for machines to read: "Most of the things reading these works are machines, not people. They're machines!"

Kahle has communicated his concerns directly to representatives of Google HathiTrust, recounting:

> "I have literally, in public, gotten on my knees and begged, hands clasped in front of me, in front of Michigan's Librarian and the head of the Google Library Project, and begged them to open the public domain. Literally.
>
> They said 'No.'"

When I interviewed Michigan's Librarian and asked about this interaction, he acknowledged Kahle's account but explained that the restrictions in Google's contract limit

Michigan's ability to share the scans, even the scans of public domain works. The contract restricts sharing to library consortia and makes it subject to the effective implementation of download restrictions. The Internet Archive does not satisfy either of these requirements. The Librarian concluded by saying:

> "My personal view — and I have no trouble saying this; I've said it in enough places so if I haven't been quoted it's because nobody's listening — is that Google's view of this was wrongheaded. They should have just opened up the public domain works. That's still my view."

Other parties to the contract negotiations with Google reflected that both the agreement itself and the subsequent emergence and evolution of HathiTrust were imperfect. The realities of the situation reflect compromise and satisficing but it was a vast improvement over what had come before:

> "Is this going to be the best version of this thing that could possibly be? Probably not. Is it going to happen now rather than in the distant future? Yeah. Yeah, it can if we work on it. So, that's really how it happened."

Despite its somewhat controversial handling of the public domain, the emerging version of HathiTrust — a trusted digital repository with access provisions for public domain works — had tremendous appeal to the academic research library community. Several of the participants in this study indicated that if HathiTrust had stopped there, and not moved forward with additional forms of access and service, it would have been a tremendous success. Indeed, numerous institutions were prompted to join HathiTrust on this basis alone.

In December 2010, HathiTrust gained its first member outside of the initial membership of the CIC and the University of California. According to participants at that institution:

> "We joined because we felt it was important, and the opportunity was there. HathiTrust was looking to expand membership at that point. We had been partners in different mass digitization projects at that point, with Google and Microsoft, and we were looking for a means of preserving those works and making the most value out of those. By the time we joined, HathiTrust had moved beyond preservation. There really was something to look at and see and use. It wasn't simply a dark archive or anything of that nature.
>
> Where the emphasis was still towards preservation was in HathiTrust's design of the cost basis for joining. It was still based at that time on the amount of material an institution was depositing HathiTrust, so essentially the cost of the storage. At that point, what you pay for affects the way you see something as well. So it was

still seen at that point within the library community as a preservation repository. But it was starting to change.

We did not digitize any of our in-copyright materials, so even if HathiTrust were to be sued for copyright infringement, it would not have had a direct negative effect directly on us. Of course it could affect us in other ways in terms of reputation of the university — how we are seen by our faculty and others — but that wasn't a serious concern to us."

As additional members joined HathiTrust, they echoed the sentiment that digital preservation and access to the public domain were primary drivers:

"Membership in HathiTrust was appealing because it enhanced our digital preservation strategy. We had a local copy, an Internet Archive copy, and now the HathiTrust copy as well. The Internet Archive offered to transfer its holdings to the HathiTrust but not every partner offers to do that. Membership also provided holdings expansion because anyone can have access to the public domain materials and anyone can search across the copyrighted materials. Hathi also makes it possible to put a link to the full-text in our catalog and therefore provides a very enriching tool, in addition to the three-part preservation strategy."

Another early HathiTrust member recalled:

"We joined HathiTrust because we wanted a third, safe place to store our digital copies, in addition to the Internet Archive and Google. All of our scans were in the public domain and so HathiTrust gave us easier access to them. Prior to that the only way we could get to them was to link through Google books and while I don't think that Google's the Devil, they are a business and make business choices. We wanted to make sure that our records were safe and that if we were linking to a resource that the resource would be around."

From the perspective of prospective members, a primary source of HathiTrust's value stemmed from its ability to offer digital preservation and access services related to public domain content. In addition, this quote suggests that its identity as an organization of, by, and for the academic research community was also an important factor.

Relatedly, many members noted the considerable goodwill and enthusiasm for HathiTrust's mission: "To contribute to the common good by collecting, organizing, preserving, communicating, and sharing the record of human knowledge."[404] HathiTrust was becoming more than a preservation-focused repository to something with a broader vision and purpose,

[404] HathiTrust (2011). "HathiTrust Constitutional Convention Meeting Notes," p. 9, available at http://www.hathitrust.org/constitutional_convention2011

more closely resembling a digital library. HathiTrust was becoming a brand. As one of my participants described:

> "It didn't take long after HathiTrust was born for people to realize that this way was much better than continuing down the road of the shared digital repository. HathiTrust proved itself hardly before the ink was dry. We had other libraries clamoring to join in. As new partners came in, it became "the thing." What we all wanted all along was for HathiTrust to be the digital library for academic libraries going into the future and with this many partners joining, we got that solidarity and that brand."

The reconfigurations of HathiTrust were tied into pre-existing social relationships and institutional frameworks within the library community. HathiTrust was developing and evolving organically, according to the desires, needs, objectives, and concerns of its stakeholders. The development of HathiTrust and the growth of its constituency were self-affirming and reinforcing. Although formal control resided in the Executive Committee, these aspects of development and evolution were already pointing toward its reification as a collective. As Weick and others have described, the development and reification of an organization is embedded in social relationships and processes. In retrospect, the social aspects of organizational development can become less visible as they become more fully integrated into our seeing and understanding of the organization. As the social aspects and relationships "tighten" through shared sensemaking and collective action, decisions become more predictable, orderly, organized, and self-reaffirming. This process of reification is ultimate a social one that lends both form and substance to a developing organization. In understanding how and why HathiTrust developed and evolved during this early period, we should therefore look to social factors for guidance.

HathiTrust's emergence and evolution shaped and was shaped by the social environment in which it operated:

> "Academic librarians know what they know, which is traditional academic librarianship. Some of them may be very enlightened and forward looking about this, but realistically what they're in the business of doing is academic libraries the way they've been done. There was a very strong centripetal force pulling things back toward the center. Even if you got people who were entrepreneurially minded, and early adopters, and risk takers and so forth, they were still from the library community. So, most of the people who governed the details of how HathiTrust developed were mindful of library issues. They were thinking in terms of what academic libraries do, what academic libraries are for. There were also, of course, reinforcing influences from the provosts, the faculty, and the other

> academic community members who themselves were stakeholders in university libraries which were then stakeholders in HathiTrust. So, in terms of what HathiTrust would become, it's just a matter of who's doing it and what their values are, what their consciousness is, and what they care about. The library community was very clearly going to be running this, and it did. I think it's a good thing, but it was also an unavoidable thing.

That said, however, the social environment in which HathiTrust operated was not homogenous. Major research libraries such as the University of Toronto and digital library alternatives such as the Internet Archive disputed HathiTrust's emerging characterization as a "good thing." The perspectives of these participants offer a contrasting view of HathiTrust's role and impact on this broader transforming ecology:

> "The question was: What were the libraries going to do? When they went completely secret, that was kind of an indication that something strange was going on. What we ended up with was HathiTrust which is kind of like a JSTOR: it had funding from the Mellon foundation, and JSTOR is financially very successful, and it helped lead to the death of Aaron Swartz.
>
> The idea of making a subscription service for libraries is not a very "Google" thing to think of. It *is* a very "major research library" thing to do. It comes in the tradition of OCLC or JSTOR or RSTOR. And if you take the big libraries that Google was dealing with, it's not an unnatural outcome. And if you take the libraries that *didn't* go along with Google: Library of Congress, University of Toronto, Boston Public Library, they have leadership that are wary of that style that is becoming very common in higher education libraries. Higher education libraries built Elsevier. As much as the libraries complain about Elsevier, it only exists because those major research libraries funded Elsevier at a phenomenal level.
>
> So, where did that come from? The University of Michigan, the University of California, and the others didn't have to go and make a centralized system. It's just the kind of thing that those libraries do and I don't think it serves the public."

Arguably, some of the ways that HathiTrust serves the public was in pushing the boundaries of fair use with respect to digital content. During this formative period in HathiTrust's early history, it undertook several important initiatives, some of which formed the basis of the copyright infringement lawsuit discussed in the next chapter. The following sections briefly discuss some of the sensemaking around those initiatives.

ACCESS FOR NON-CONSUMPTIVE USES

In the period between HathiTrust's launch and the Constitutional Convention, the organization (and its members via the organization) undertook several initiatives related to non-consumptive research. Three that will be briefly mentioned are the Copyright Review Management System (CRMS) project spearheaded by the University of Michigan under a grant by the Institute of Museum and Library Services (IMLS), full-text search which was a Hathi-wide service, and the HathiTrust Research Center which was launched in the spring of 2011 as a collaboration of the University of Illinois, Champaign-Urbana and Indiana University.

The CRMS project was undertaken to fill in the massive gaps in copyright determinacy resulting from the ad hoc registration and arcane reporting practices of the Copyright Office. As described by the primary investigator on the grant:

> "Through this project we can identify the author death date, and if something is old enough, we say that it can be opened, but even if it's not we know when to set a trigger for us to be open in the future. We emphasize finding, seeking the public domain, but were really emphasizing the idea of copyright determinacy. More information is helpful. When we identify a work as being in the public domain, that conclusory information is provided to HathiTrust to determine whether something will be opened."

The CRMS project is thus an example of non-consumptive research because the primary interest in a work is not the information contained in it, but rather the information about it. Furthermore, the copyright review process is one that would be virtually impossible as an administrative manner without access to the digital scans:

> "A few other institutions have tried small-scale copyright rights determination pilot projects without HathiTrust, pulling the books by hand, and it's basically impossible to do that work, even on a small scale. You need some kind of interface and access to the digital scans to be able to do this work."

As described by another librarian involved in the project:

> "It would not be possible to do the work, copyright rights determination work, as an administrative matter unless we have the reviewers having collective access to the scans that are in the HathiTrust. The fact that there is this collection, this cross-institutional collection that, with appropriate security procedures, we can make reviews efficiently because of the scans rather than pulling individual books. It allows for a high level of efficiency and it also allows us to have documentation of each of the reviews, and we know that remote reviewers, because of the double-blind process, to different people at different places will randomly review the same assigned work, we couldn't do it unless there was

> something like HathiTrust, some mechanism for sharing the scans in the same system. We couldn't have that level of administrative possibility and consistency. The HathiTrust framework and collection allows for us to do more work because, if we didn't have that tool, we would be doing this by hand, or frankly not at all because it's not feasible."

The CRMS project was initiated by Michigan in 2007, prior to HathiTrust's official launch. As it proceeded, the project continued to be a Michigan initiative but it arguably benefitted the other HathiTrust members and the general public as works that were determined to have entered the public domain and/or information enabling a trigger date for future transition to the public domain were set. This project also set an important progressive precedent for subsequent projects. As the operations manager of HathiTrust noted:

> "Getting the Copyright Review Management System operational was a big deal because when that went into production the scans became something besides just a big block of dark stuff that we can't do anything with. There is really something there. It's malleable. You can flip the switch from closed to open and I think that is a huge part of what we're doing — to enable the greatest access to these materials, and different kinds of access."

The CRMS project was simultaneously serving an important social function, but investigating facts about protected works to build the fount of public knowledge and, in some cases, access. But it also highlighted the striking insufficiencies in the way that the Copyright Office managed and communicated basic information about protected works. In a shocking number of instances, works remained underutilized simply because basic information about author death dates and location of first publication was both crucial (in terms of copyright determinacy) and incredibly difficult to acquire. Therefore, while performing a significant substantive function, the CRMS project also highlighted just how big the problem is.

To manage the scope of the problem, participants intended to create a review process that was distributed amongst various partner institutions. One of the mechanisms they were working on to facilitate a distributed CRMS was "toolkit:"

> One of the things that we're trying to do is create a toolkit that other partner institutions can use to do their own rights determinations. So we're creating and vetting the methodology that other institutions can adopt to help share the work that way. Earlier in this project I thought that we could develop these toolkits and make it completely distributed and everybody could do their piece but I think to do this responsibly you actually need to have a committed investment somewhere; there needs to be a committed home.

At least for now, that home is the University of Michigan.

Another challenge for the project has been its incomplete or underutilized record-keeping and communication practices. As it currently stands, all records and documentation from the project are held solely by the University of Michigan. An unintended consequence is that it makes tracing the impact of the project very difficult because there is no visible marker on works in HathiTrust that shows that the work has gone through the rights review process. As the principal investigator describes:

> "We don't actually know who is using these works, how many people are using them, or how they're getting used. When people use works that were opened by virtue of the copyright rights determination process, they're not aware of that because there's not a mark on the record that says it was reviewed. I'd really like to have records that indicate that something was reviewed even if it's closed so that people are aware and they don't ask us again, but also so that there's recognition in the value of the review being done."

It is far more difficult to demonstrate and communicate the value of the CRMS project to those not involved when there are not these digital breadcrumbs or traces that can be followed. Works that have gone through the rights determination process are not marked as such.

Another early form of non-consumptive use was full-text. The digitized content was OCRed and indexed which made the full-text of the entire corpus searchable. As described by HathiTrust's day-to-day manager, establishing full-text search was a huge milestone because it distinguished HathiTrust from most other library catalogs whose search returns basic bibliographic information about a work:

> "It becomes the sort of living breathing thing where you can search for the name of your grandfather or grandmother and get results and find out information about them."

In terms of sensemaking about adding full-text search functionality, one of the participants in that decision described the situation as follows:

> "We created a shared repository that was partitioned into fire-walled chunks that each institution could, under the "dark archive" doctrine, dump their stuff into. And then there was a second layer of discussion which had to do with fair use and the sentiment was that stuff about the stuff is fair use. So the number of instances of the word 'dog' across the HathiTrust corpus could not be copyrighted. It could be copyrighted if somebody wrote something about it, but the copyright holders on the documents couldn't say 'You can't look at my instance of dog and count it.' Of course you can take that to a fuzzy point but that was part of the process of changing in the law and social understanding and so forth. It was pretty clear that

> text analysis was going to be possible against this corpus and the limits of this were not clear, but most of the low-hanging fruit and immediate stuff was clear enough that people who knew about this weren't worried about it."

Again, the sensemaking reflected in this justification references expectations around copyright but does not necessarily reflect a strict doctrinal understanding of the law. Copyright infringement does not require that the infringing use result in a protectable work, e.g. that the number of instances of the word 'dog' be copyrighted. Rather, copyright is a strict liability tort and fair use functions to make otherwise infringing uses non-infringing. The adjudication of full-text search is addressed in more detail in Chapter VIII of this thesis.

The last example of non-consumptive use I will mention here stems from the launch of the HathiTrust Research Center (HTRC) in April, 2011. A service jointly offered by the University of Illinois and Indiana University, the HTRC aims to support computational non-consumptive research across the corpus for educational purposes.

> "HathiTrust assembled a working group to develop a call for proposals for the research center, a request for proposals, and Indiana and Illinois submitted a proposal for that; faculty of those institutions submitted the proposal. They wanted to establish this research center so that is what they've done. It still got quite a ways to go, but the way it works is that faculty request data sets and so forth from HathiTrust. We distribute data sets, the research center has all the materials at least those right now that are in the public domain, and they offer access to it in an enclosed environment."

In terms of the evolution of HathiTrust, it was important to the Executive Committee to take advantage of the affordances of HathiTrust's status as an academic organization and "push into a new service model around what had started out as a preservation model." The contract with Google was somewhat constraining in terms of supporting computational research, but so long as HTRC could institute mechanisms to vet projects for their academic research potential, the contract would not be prohibitive.

> "The establishment of the HathiTrust Research Center was an incredibly important milestone because it showed that not only do we offer the basic level of access to what we have, but we enable new kinds of research. In thinking about the future of HathiTrust, it means a lot to know that were in that space, were in that direction of cutting edge research, and enabling new uses. If we were just a static collection, we gathered the stuff and you could click through it, it would be fantastic, but it makes it a whole different level to have that research center, that research component.

> But, the research center is still trying to figure out its role. Some research has been done, but I don't think that it has exploded to the point that you might anticipate in the future."

One of the participants in this study described an incredibly frustrating and disappointing experience of a faculty member who was attempting to do basic metadata text mining of the corpus. Ultimately, employees at Indiana University stepped in to provide helpful assistance but the participant was left with the following impression:

> My theory is that HathiTrust has a great brand with worldwide recognition and cachet, but underneath it's not a highly sophisticated technology infrastructure. It's trying to evolve into something else and it's not there."

Access for Print-Disabled Patrons

The University of Michigan was an early leader in digitizing works for purposes of making them available to print-disabled patrons. It had been engaging in this activity prior to the Google Library Project. Participants reflected that accessibility was a long-standing important issue for the library and was a driving force behind digitization efforts.

When HathiTrust was launched, Michigan imported its policies on accessibility to HathiTrust which, in turn, became the instrument for "serving the works up to other institutions" for use in the provision of services for print-disabled patrons. As one participant from Michigan noted, "HathiTrust is where we store our stuff, so that's where we get it from to provide this service, but Michigan is doing it. HathiTrust was sort of incidental to Michigan."

In terms of the broader academic research library community, access for print-disabled patrons was not a priority:

> "It wasn't an issue that was being discussed within the community as a whole, at conferences and other things like that. It wasn't a driving force for digitization. It was very much a follow-on to digitization rather than a driving force. It was one of the purposes we came to figure out after the fact. Now that we've got all the scans, what are the other uses we can make of them? Access for the print-disabled emerged as a good option after the fact."

Member institutions are able to opt in to the service and a number of institutions have taken advantage of the service. Overall, response to the program has been nearly uniformly positive. One participant commended "Michigan's really wonderful work for increasing the access for individuals with reading disabilities." Another remarked that one program was, in his view, one of the most revolutionary aspects of HathiTrust:

> "What I think was really revolutionary is access for the print disabled. Doing something like this hadn't occurred to me until HathiTrust implemented it. But, boy, what a game changer that is! Rather than having to get a book, and take it to a special office on campus, and have them put it on Kurt's file machine, and have it show up a week or two weeks later … to suddenly to be able to have access to 11 million volumes within a day or two, and not being restricted in what you're reading, that is a real groundbreaker."

The program for print-disabled patrons has had a positive impact on accessibility issues more broadly. In recent years, for example, the Internet Archive has been actively digitizing current books and making them available to the blind and dyslexic. Kahle notes:

> "We're taking books that were published in the twentieth century and lending them and there hasn't really been a problem with that because we're nonprofit, and libraries are nonprofit, and these books are all by and large not very easy to obtain."

Access to Orphan Works

The third category of access which HathiTrust pursued during this time dealt with so-called orphan works — works that are in-copyright but the rightsholder(s) are unidentifiable, indeterminate, or uncontactable. Again, using HathiTrust as an instrument for cross-institutional collaboration, Michigan and a number of HathiTrust members partnered on the Orphan Works Project which launched in April, 2011.

This was another initiative that was largely advanced by Michigan and there was some speculation that the timing of the Orphan Works Project may have been expedited in light of the Constitutional Convention scheduled for later that fall, 2011. One participant recalls the Executive Committee and Michigan's General Counsel's Office deciding:

> "Yeah, this is one of the things HathiTrust is going to do. Let's go ahead and get this in place because once the new control structure is in place they might be too conservative to do this. And we think this is something they need to do."

Given the fact that Michigan attracted a few other partners on the Project, it may not have seemed as contentious as the just-quoted participant feared. In fact, one of Michigan's early partners reflected almost nonchalantly on the decision to tackle orphan works:

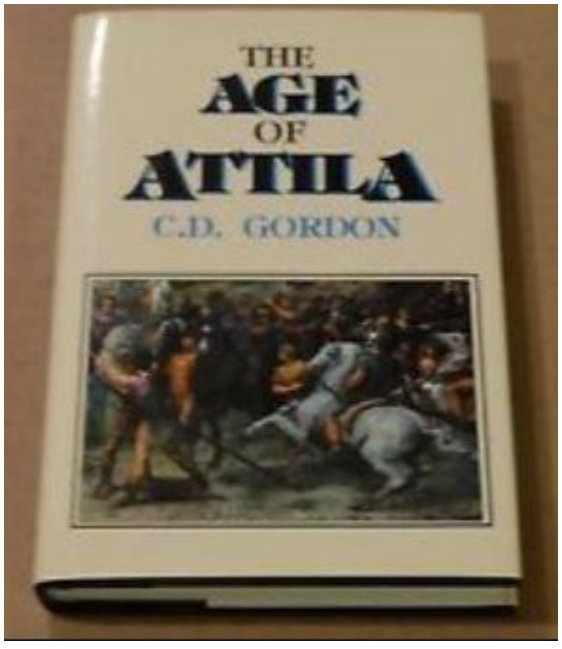
> "When the Orphan Works Project came along, we said, 'Yeah. That sounds pretty neat. Let's try doing that as an experiment.' And that's why we joined on with that. There were four or five schools that said that they would participate in the Orphan Works Project."

As a consequence of the lawsuit discussed in the next chapter, the Project was suspended before digital access was actually granted to any of the purported orphan works. It was therefore significant as a possible trigger for the lawsuit. In addition, the Orphan Works Project was significant in terms of how it foregrounds tensions between copyright law and social practice and problematizes sensemaking around initiatives that push the boundaries of access with respect to in-copyright works.

By all accounts the Orphan Works Project emerged from the digitization project rather than being a driving force in digitization. As one of the lead architects of the Project describes:

> "When we started thinking about what fair uses we might make of the scans, orphan works stand out as a putatively significant body of materials. For me there was a sense of the need to push the question. There is a large body materials in copyright but a significant percentage of it has no associated rightsholder. As a consequence we end up squandering a significant resource with copyright protection."

There was very little data or first-hand experience on how significant the body of orphans might be. So, a number of pilot orphan works investigations were undertaken. For example, one participant recalled an investigation involving a book published by the University of Michigan Press in 1960 called *The Age of Attila: Fifth-Century Byzantium and the Barbarians:*[405] *(Figure 16)*



[405] http://catalog.hathitrust.org/Record/000538413

Figure 16. Cover of "The Age of Attila," C.D. Gordon,

> "There was a book, I think it was called *The Age of Attila*. It has some relevance today for classical studies but it was never a big earner and will never become a big earner. And so we said we want to put it back in print. We have a scholar who wants to make some editorial changes, and the contract gives you some percentage of the royalties, which we would honor. We put the 1956 book online with a Creative Commons license and the rightsholder said 'yes' without any hesitation. I think we all believe that most scholars are interested in their scholarship being shared rather than profiting significantly."

In terms of copyright policy, the Orphan Works Project seemed to make a lot of sense. Presumably if rightsholders really cared about their exclusive rights they would not make themselves so difficult to discover, seek permission from, or pay royalties to. In the above example, the rightsholder ended up being fairly easy to track down, in part because copyright in the work had been renewed and thus the provenance was easier to trace. In many instances, however, rightsholders are far more difficult to discover. Unfortunately, the Copyright Act is indifferent to the difficulty or ease with which a rightsholder might be identifiable or locatable. The onus is entirely on the user; rightsholders are free to be as visible or invisible, attentive or inattentive, as they wish.

One of the participants in the Orphan Works Project provided the following telling anecdote about the on-the-ground realities of identifying rightsholders:

> "One of my last experiences before the Orphan Works Project was canceled was at the print house where we were going to republish a work whose copyright had been renewed. And I will say that the press had a very cavalier attitude about this. No contact had been had with the rightsholder for decades and therefore we didn't need to worry about the terms of the contract. And I insisted that we pull out all the stops to try to find the author's heirs. We went through obituary statements in local newspapers and found that the person had moved around Canada and ultimately that his heirs were predominantly in Western Canada. Some church announcement or something like that led to finding the son who told us that the wife was still alive and we put it all together. It took easily in excess of 100 work hours, over three months, to locate the rightsholder but we did, and the heirs were pleased to see the work be made available openly online. It was hard and yet ultimately possible to locate the rightsholders."

This anecdote suggests that if enough resources are thrown at the question, the rightsholder will ultimately be discoverable. "It may be a very hard problem, but it's ultimately a doable problem."

As a practical matter, it is not feasible for libraries to spend hundreds of hours tracking down absentee rightsholders in a single work that probably has a very limited potential audience. Therefore, participants in the Orphan Works Project turned their attention toward developing reasonable mechanisms and processes for vetting potential orphans and defining standards of due diligence with respect to the search process. As it turned out, this was not an easy task. A participant involved in managing the day-to-day operations of the project described its basic inner-workings and speculates as to its unfortunate downfall:

> "The Orphan Works Project design was reasonable. We planned to undertake a search with due diligence and once we made a determination we would publicize the hell out of that determination and give rightsholders an opportunity to come forward. I think the design was right, but it is very hard to find true orphans and we made a lot of mistakes. I'm not sure what the cause of the mistakes was. I think our answer was to spend more time on the determinations and I'm not sure that more time is necessarily the right answer.
>
> If we had not made so many glaring mistakes in the beginning — mistakes resulting from carelessness on the part of those conducting and reporting the search — it might've been a good experiment to test the publicity piece of things. If we had been able to secure cooperation from publishers and authors, the same thing I think would be true."

On paper, the process was reasonable. A diligent search would be undertaken; if no rightsholder could be identified, the library would publish the work on a title of potential orphan works and provide a window of time (perhaps a couple months) for a rightsholder to come forward; if a rightsholder came forward, the work would no longer be considered an orphan; if no rightsholder came forward, an electronic copy of the work would be made accessible to an authorized UM user. The number of print copies in the collection would determine the number of simultaneous users of the electronic copy so that if the library held two print copies of a work, two authorized UM users could access the electronic copy at any given time. According to the lead project manager, the failure of the Project was that, in its initial stages, several works were included on the orphan works list which were obviously not orphan works. Even though those works were quickly removed from the list (demonstrating that the notice and takedown process was indeed effective), these initial "careless" mistakes cast doubt over the entire Project and raised the ire over some publishers and authors' advocacy groups.

Other participants in the project took expressed a somewhat more dismal perspective on the orphan works challenge:

> "For me, part of the challenge with orphan works is I don't think that there's a way to define due diligence. I like to think that we would know it when we see it, but it's always possible to second-guess somebody. I don't think that we can ever come to a common understanding of what due diligence is, in a measurable way that would be satisfactory to everybody. I think the Orphan Works Project failed because we never grew comfortable with our process. It never engendered the kind of confidence that we felt we needed to have a go forward. It wasn't because of the lawsuit, it was because we had a loss of confidence in the process. I'm not sure what a reasonable level of risk would be."

Others argued that the focus on orphans was entirely wrong because, as the earlier-described anecdotal experiences demonstrate, in reality, that while there may be a lot of absentee rightsholders, there are few-to-zero true orphans. One participant suggested that the focus should have been on rights determinacy rather than orphans:

> "Ultimately, I think that the issue is not about orphans, it's about documentation. I spent a lot of time trying to convince the legal advisory group and then HathiTrust leadership that we needed to build the Book Rights Registry, taking the information that we had gathered in the copyright determination process as the starting place and then, for the works that we had determined to be in-copyright, work to bear down and those and get to the point where we could say 'this is the rightsholder and it is a publisher and it has been identified' or 'she is an author and she has been identified' and 'this is the contact information for the rightsholder' or 'no one has been able to determine whether this person or their heirs exist.' In other words, to document those things and to search for that information. And I think at the end of that process you end up with orphans but you don't start out trying to find orphans.
>
> This is the conclusion that I came to after going through the orphan works project and I think that this is still very important and worthy work that we ought to undertake. I still believe and advocate for the HathiTrust Book Rights Registry, that we should try to document rightsholders. It's not about finding the public domain, it's about fleshing out the picture of copyright. It's not about finding orphan works, it's about finding rightsholders and sometimes not finding rightsholders and sharing that information with others. That's the goal. And by doing that, we have a much better picture of what the publishing and copyright terrain is. Popular representations about works created between 1923 and 1963 is that only 15% of that is still in copyright. Over the course of the CRMS project 65% of the materials were found to be in the public domain and something like 30% or 40% were found certainly to be in copyright. And those facts I think ought to shape the conversation. But you can still find stuff on the web that says that, people who make money have a business around that kind of thing. The conversation about orphan works, out-of-print works, and the public domain could be a more productive conversation if we have those facts."

In essence, the Book Rights Registry proposed by this participant is an attempt to retrospectively fill in the gaps that are left by the United States decision to forgo formalities requirements. What the whole orphan works discussion and Book Rights Registry project would do is provide documentation were documentation doesn't exist. As one of the participants said: "I think the formalities could and should provide that kind of documentation." However, given that the United States is obligated under a multi-national agreement called the Berne Convention to abandon formalities requirements, it is doubtful that reforms to the Act will include formalities.

Another participant suggested that the focus should not have been on so-called orphan works, but on out-of-print works.

> "I think that is the fundamental issue is that I'd rather not have this be about orphans. I'd rather have it be about the original framers' understanding of copyright. If there is no market harm because the works are out-of-print then how does keeping a body of materials locked away advance what is phrased as the science and the useful arts? We should focus our attention instead on developing a way to be smart about determining what is in-print and on the market. And if some publisher says: 'I don't like that, I'm going to put this 1973 work back into print, or make available for sale' there should be mechanisms that recognize that and turn off access. I think that would shift the focus to whether these books are meaningfully on the market, but I'd rather have that problem the current one.
>
> I feel like I found a pessimistic space about this but I'm not at all pessimistic. I think that the mistake that we made is to make it about orphans. And I think that the right conversation is about facts, is about determining facts, and about reasonable uses of the material and if we knew that there were 3 million books that were in copyright, and we knew that 95% of them were not on the market, that you could have a very good conversation about that regardless of whether their orphans or not."

Ultimately, the Orphan Works Project never got off the ground because when the copyright infringement lawsuit was filed in the fall of 2011, Michigan made the decision to indefinitely suspend the Project. Even now that the lawsuit against HathiTrust has been resolved, there is little hope that the Project will be resumed. For many in the library community, the brief experiment that was the Orphan Works Project was perhaps HathiTrust's greatest tragedy to date: "I think that the Orphan Works Project was great. But I think it was rushed and I think that it set everybody back. And I'm in this for the long haul. It's a war not a battle." As the next chapter explains, even though HathiTrust successfully defended its lawsuit,

because the court dismissed the claims based on the Orphan Works Project and thus the Project was never adjudicated, we might never know how the orphan works story might have gone differently.

Constitutional Convention

The Constitutional Convention was one of the major milestones in HathiTrust's evolution. It marked the transition of control from the initial progenitors of HathiTrust to the broader collective membership. HathiTrust had been chartered by the founding partners for an initial five-year term: 2008 – 2012. As previously mentioned, the Constitutional Convention had been planned from the outset of HathiTrust as an event to take place in the third year. It was intended that members would come together to conduct a formal review of governance and sustainability. In the words of one of HathiTrust's lead architects:

> "We knew that HathiTrust would grow and that in a few years we would need to reevaluate everything — see how everything was going and see what other structures were needed for governing it and guiding its development into the future."

The Constitutional Convention was the turning point where HathiTrust went from a very small, relatively homogenous organization in terms of its membership, to a much broader membership in terms of both numbers of members and the range of institutions involved. Institutions who joined HathiTrust by October 31, 2010, were permitted to participate in the Constitutional Convention. The result of this timeline was that HathiTrust experienced a tremendous upsurge in membership leading up to the fall of 2010 as institutions sought to secure their place as participants in the Convention:

> The manager of HathiTrust's day-to-day operations commented:
>
> "The convention was a very big thing for us. There are a lot of institutions that joined in order to be part of the convention. They realize that in order to have a voice in this huge initiative that was going on, they needed to be a member, and they wanted to be at the table. So we went from 23 partners to 52 partners at the constitutional convention. It was a critical mass of institutional partners that were interested in making a long-term commitment to this effort."

The meeting took place the following year in Washington, D.C., October 8-9, 2011. The meeting notes of the Convention, as well as the data generated from in-depth interviews with

participants, offer a richly detailed account of the process and outcomes during this pivotal period in HathiTrust's transition.

The overall tenor of the meeting, if one can be distilled, was that the membership communicated a tremendous amount of reverence and gratitude for HathiTrust founders and took on the mantle of directing its future course with enthusiasm and lively optimism. For example, comments from participants included the following:

> "When we had the constitutional convention, it was not a rebellion of the membership, it was praise for the original group and acknowledgment that we needed to find a way to go forward. And elections took place the next year."

And:

> "It was just alive. It was really exciting. I'm sure some people didn't like it, but I thought, but how often do you get a chance to just kind of feel like you are at the beginning of something?"

And:

> "What was exciting about the constitutional convention is that it was so democratic. We started talking about projects that were important — digitizing government documents and those kinds of things and coalitions were being formed around to support it. It reminded me of the presidential conventions on television in the 1950s when I was a child. They weren't all orchestrated. You didn't know who the winner was at the beginning of the convention. And so you had people running around, prophesizing here, pulling these people together — we are really interested in government docs, if you vote for this will vote for that. It was very lively and a lot of good ideas came out of it. And we did come away with agreements as to what the governance structure would be, how much time before we would form the board, and that sort of thing. And John Wilkin, the Executive Director, he really was the mainstay. When you've got one person who lives, eats, and breathes it, which he did, and had just a core of wonderful people around him. We could all sit back and say, 'Oh, thank you for doing that!' I think the fact that we were able to see that transition in leadership happen speaks very well of the strength of the HathiTrust."

And:

> "The people who came up with HathiTrust and were involved in its founding deserve tremendous credit. They were the ones who took what might have been just another Google project and turned it into something that has a different focus and a different vision."

And:

> "The Constitutional Convention was an excellent success. There were ballots, and motions from the floor, and people rose to object. There were 150 people in the room and that was really when HathiTrust moved into collectivist ownership."

And:

> "This is also where the vision changed of HathiTrust as a place to put your content and preserve it, as a place that serves research and scholarship in other ways as well, a shift from a preservation agency to a content repository. And then into other things as well."

The "other things" alluded to in that last quotation were, in large measure, represented in the ballot proposals submitted by and for HathiTrust's membership. The Convention consisted of a series of sessions attended by delegates from each partner institution and consortium. The sessions were designed to consider the ballot proposals submitted by partner institutions, allowing for amendments to proposals, and culminating in a final vote by partners.

In terms of how votes were allocated, each partnering institution was allocated a certain number of votes that it could cast for the proposals based on a formulation that considered the member's financial contribution and contribution of digital content. HathiTrust published a paper and an accompanying table outlining the vote weighing process.[406] In anticipation of the Constitutional Convention, the HathiTrust Executive Committee requested the University of Michigan to conduct research into effective multiparty weighting models, paying special attention to factors including the level of investment in the enterprise, such as duration of participation, role in establishing HathiTrust, the number of volumes contributed, and financial contributions. As a result of this research and subsequent Executive Committee discussions, a voting model emerged that based on financial contributions (excluding resources-in-kind) and volumes in the repository.[407] The allocations have been normalized to represent a percentage of 100 total votes, and have been rounded to the nearest whole number. Allocations that came to less than 1% of the total were rounded up to ensure that the smaller contributors have a single vote.

Partners were permitted to submit proposals on any topics, in any areas, prior to the convention. As a result, the following seven ballot proposals were considered, amended, and

[406] HathiTrust.org.
[407] Institutions' voting allocation is based on the sum of the square roots of the institutions monetary contributions and deposited volumes.

voted on at the convention. Five proposals were passed, one was rejected, and one was referred to the HathiTrust Board of Governors:

1. HathiTrust Distributed Print Monographs Archive Proposal
    a. Proposes establishment of a distributed print archive of monographic holdings corresponding to volumes represented within HathiTrust that is collectively supported by the HathiTrust membership.
    b. Submittor(s): HathiTrust Executive Committee, California Digital Library, University of California – Los Angeles, University of Michigan, Northwestern University, University of Illinois – Urbana-Champaign, New York Public Library, Columbia University.
    c. Vote: PASSED
    d. "Libraries are running out of space in a lot of places so it would be very nice to say, 'Okay, I don't need to have this material from 1920-1940, people aren't using it, I think I'm going to store that and I will count on the HathiTrust digital copy rather than the print copy.'"
2. Approval process for development initiatives by HathiTrust partner organizations
    a. A proposal for HathiTrust to formalize a transparent process for inviting, evaluating, ranking, and launching development initiatives from HathiTrust partner institutions.
    b. Submittor(s): Cornell University, Columbia University.
    c. Vote: PASSED
    d. "I just asked whether HathiTrust is in the business of archiving and curating data. He said, 'No. I wish they were and they should get into this, but this is not something that most of the members want to do.' I was like, 'Okay, if they don't wanna do it, they don't wanna do it.' I'm no longer making policy."
3. Governance
    a. A proposal to establish an effective governance structure.
    b. Submittor(s): CIC University Librarians.
    c. Vote: PASSED
    d. "At the Convention, one of the proposals was about establishing a new governing structure for HathiTrust that would be inclusive and representative of the larger

membership that had formed. And that governing structure — it's been a couple of years in the forming and solidifying and everything — but that has been a huge thing. To be able to say that, with this large constituency, we have a governing structure that allows institutions to participate and have a say, that's been a really key moment for us going forward."

4. Expanded coverage and enhanced access to U.S. government documents
    a. U.S. federal documents represent a significant resource for research and education. The CIC's initiative to digitize a comprehensive corpus of print documents is making significant headway, but an expanded effort is proposed. Further, problems with cataloging records and basic metadata do not provide sufficiently robust discovery to these complex resources.
    b. Submittor(s): CIC University Librarians.
    c. Vote: PASSSED
    d. "One plan is to create a complete digital copy of the federal legacy government documents with complete location information to facilitate getting a printed copy. One of the reasons we proposed this was Google took a very, very conservative view of rights around government publications. We thought we, as an academic group, would be fully within our rights to make those available. And we were willing to take that chance."
5. Mission and goals
    a. A proposal to broaden the stated mission of HathiTrust
    b. Submittor(s): CIC member universities as represented by their University Librarians.
    c. Vote: Referred to Board of Governors
    d. "I remember at the Constitutional Convention, I raised my hand and said 'I'm kind of puzzled. Is HathiTrust a preservation tool or an access tool?' And now it's clear: access — legal access with full respect of legal rights — is a main part of the mission."
6. HathiTrust implementation review committee

a. A proposal to create a HathiTrust committee to review the implications and applicability of services, policies, and procedures developed by its members before they are adopted by HathiTrust as a community service.
b. Submittor(s): Cornell University, Columbia University, University of California.
c. Vote: NOT PASSED

7. HathiTrust fee-for-service content deposit
    a. A proposal for HathiTrust to have a fee-for-service model to allow contribution of content from non-partner entities.
    b. Submittor(s): University of Michigan
    c. Vote: PASSED

Through the Constitutional Convention, HathiTrust's progenitors gave the organization over to its collective membership. One of the ballot proposals in particular — proposal one which deals with the creation of a distributed print monographs digital archive — represented an important culmination and sort of implicit send-off from HathiTrust's lead architect to the collective.

> "We have a shared responsibility for the print record and we don't have a very strong sense of how our collections align with each other and how that print record is represented in, say, the ARL's, the big research libraries. If we do this inductively, we start with a major body of material like the Michigan collection or what HathiTrust became, you can turn to the overlap analysis and see how our collections are embodied in it, or not embodied in it, and both of those things are helpful in addressing the shared responsibility. So right now were at a median rate of overlap of somewhere around 55%. That is to say any ARL library can find 55% of its collection online in HathiTrust. From there we can begin to look at our print storage responsibilities in a way that is about the whole of our action, about our action spread across institutions. There is a shared print storage facility called ReCAP, the Princeton Columbia New York Public Library storage facility in New Jersey, and it was a very high-quality storage facility under one leadership, with three disparate, uncoordinated, storage strategies. What each institution put in was addressing the needs of each institution. And so maybe the same volume for each of the three institutions was put in. Once we started having conversations about HathiTrust with those institutions, and they committed to coming in, they changed it, and started thinking about their collection as a whole. Driving down costs by having one copy, or two copies, or whatever they thought was the most sensible thing, rather than having the uncoordinated strategy. They started moving in a coordinated direction. We know that this is a possibility for all of our institutions. We do it now with West for journals. The CIC Shared Repository is another example. We now think much more collectively about our collections. And the ballot initiative at the Constitutional Convention around shared prints monographs

> is going forward with the same assumption, that we can coordinate our activities more effectively, save money, and do a better job of preserving the print record."

In interviewing HathiTrust's lead architect — the person responsible for inserting §4.4.2 into the UM-Google, for spearheading the innovative deviance that got the CIC and UC to back the shared digital repository, the person who served as HathiTrust's Executive Director during its first several years — it became clear that the digital corpus was, in his mind, primarily useful as a way to leverage collective action around the collective collection. Passage of the distributed print management system proposal was the clearest reflection of his own vision of success.

Of course, collective action was reflected also in Constitutional Convention more generally. HathiTrust's partners were able to identify what the main priorities of the organization would be moving forward, and make important decisions about its vision, trajectory, and governance. But, as one of the participant said, "We knew we wouldn't get it all done right away because we needed to get the governing structure together." One of the key pieces of governing structure that emerged was the Program Steering Committee which empowered a small subset of representatives to move forward with some of the initiatives that were voted upon: "We needed a group to apportion this work and so forth, and now that there's a working group things have moved along really quickly."

Conclusion

In addition to being a major milestone in the evolution of HathiTrust, the Constitutional Convention also proved to be a deeply empowering and positive experience for HathiTrust's members and reflected a process of reification whereby HathiTrust continued to become real.

Tying this back to the sensemaking literature, the evolution of HathiTrust from its launch to its transition to a collective organization reflects patterns that are common to organizational sensemaking and reification. As Weick argues, one of the ways sensemaking (an essentially social process) becomes visible is through the justifications offered for behavioral commitments. These justifications often "reify social structure" by linking micro-behavioral commitments to macro social consequences. In other words, while people commit to and coordinate instrumental acts before they worry about shared goals, shared goals emerge as people search for reasons that justify the earlier behavioral commitments. This occurs through a process of shared committed interpretations. Applied to HathiTrust, we can see how HathiTrust went from being a technical

solution to an instrumental problem to being a complex values-oriented and goal-oriented collectivist organization through emerging, continuously refined, shared interpretations about what HathiTrust "means."

Reifications that justify social commitment tend to set up expectations that operate like self-fulfilling prophecies and efforts to validate the social justifications tend to spread them to other actors. When this occurs, the social relational aspects of sensemaking, through the process of reification, can generate collectivity. "That action initially explained by reification soon generates the reality that replaces the reification with substance."[408] Reification of a collectivity can become a mechanism for justifying commitment, essentially flipping the earlier sensemaking process on its head. "Having become bound to interdependent action, a person might invoke macro sources of micro-constraints as in, for example, 'that's the way we do things in this culture.'"[409] We see signs of this in the reflections of HathiTrust's members as they take collective ownership of the organization through the Constitutional Convention.

The Constitutional Convention was a milestone and a new starting point in HathiTrust's continuing evolution. Its function as a rite of passage echoes Weick's observations that "reification is an *initial* move in an extended chain of validating actions, many of which lend substance to what originally was a mere presumption of social structure."[410] Presumptions taken seriously often become self-validating.[411] The Convention marked one more big step in the academic research library community taking ownership of HathiTrust as the thing that they made for themselves.

The meeting was really a rite of passage that formally transitioned power from HathiTrust's progenitors and founders to its broader, increasingly numerous and diverse, membership. The Convention sparked specific initiatives and solidified shared expectations and visions for what HathiTrust might become and strengthened its social and organizational infrastructure. One participant summarized the overall sentiment generated by the Convention quite nicely when she said:

---

[408] Weick (1995:23).
[409] Weick (1995:19).
[410] Weick, K. E. (1993). Sensemaking in organizations: Small structures with large consequences. *Social psychology in organizations: Advances in theory and research*, 10-37, p. 25.
[411] Weick, K. E., Gilfillan, D. P., & Keith, T. A. (1973). The effect of composer credibility on orchestra performance. *Sociometry*, 435-462.

> "I'm sure some people thought the Constitutional Convention was imperfect. It's kind of messy trying to be a democracy. There is a tremendous amount of balancing you have to do between what people want to do, what make sense technologically, what make sense in terms of resources, and so forth. But within the library community, I mean, boy, I think it makes people smile. I think it makes us feel very proud that we did something like this, and it came from the library community, and it's a pretty big deal."

Although the Convention marked HathiTrust's transition into a collectivist organization, its legal status as a property and service of the University of Michigan persisted. At this point, this institutional arrangement was not contentious or problematic. If anything, it was the opposite. As one of the HathiTrust's members noted:

> "One of the advantages of having Michigan is that their legal counsel was on board with us and their President was on board with us. Copyright was always in the discussions, and Michigan stepped up there and took a really brave stance on fair use."

One of the reasons that HathiTrust's members felt comfortable proposing innovative, potentially risky, initiatives, is that they felt secure under Michigan's protective wing. As future sections demonstrate, tensions soon began to surface between Michigan's legal control and HathiTrust's emerging semi-autonomous collective form.

Chapter VIII: Adjudication of HathiTrust

In September, 2011, one month before the Constitutional Convention took place, twenty authors and authors' associations ("Authors Guild") sued HathiTrust, Cornell University, and the presidents of the Universities of Michigan, California, Wisconsin, and Indiana University ("HathiTrust") for copyright infringement, asserting that the systematic digitization of copyrighted materials without authorization violates authors' exclusive rights under §106 of the Copyright Act. In its response, HathiTrust conceded that the Authors Guild had established a *prima facie* case of infringement with respect to certain works but defended its activities on the basis of fair use.

Before describing and analyzing the case — including the District Court and Circuit Court decisions — I will briefly return to some of the legal doctrines discussed in Chapter II. In particular, I will highlight potentially significant changes that occurred between late 2004 (when Michigan joined the Google Library Project and the wheels were set into motion that eventually led to the emergence of HathiTrust) and late 2011, when the copyright infringement lawsuit was filed. In particular, I will focus on the doctrine of sovereign immunity, the Digital Millennium Copyright Act, and fair use jurisprudence. This is primarily an effort to fill the gap where the discussion in Chapter II's doctrinal analysis of mass digitization left off. This chapter then delves into the sensemaking of HathiTrust and its partner institutions in response to the lawsuit before briefing the judicial opinions. This chapter concludes by returning to the innovative deviance framework applied throughout this work and suggests how this analytic approach can help us describe and explain judicial sensemaking in *Authors Guild v. HathiTrust.*

Pertinent Evolutions in the Law, 2004-2011

This section provides a brief update to the legal doctrines discussed in Chapter II of this thesis. In particular, this section focuses on potentially relevant evolutions in the application of sovereign immunity principles and discusses fair use and transformative use case law. This provides an updated context for discussions of sensemaking that follow.

In 2004, the doctrine of state sovereign immunity as applied to the intellectual property infringement seemed strongly to favor public universities. Under the existing jurisprudence such as *Florida Prepaid,* the University of Michigan (and its subsidiary, HathiTrust) and many HathiTrust member organizations seemed insulated from liability for monetary damages stemming from tortious acts, such as copyright infringement. That said, a 2006 U.S. Supreme Court ruling potentially re-opens some central issues debated in the *Florida Prepaid* case and, for that reason, is now briefly discussed.

In 2006, the U.S. Supreme Court ruled in *Central Virginia Community College v. Katz* that Congress's powers under the Bankruptcy Clause of Article I *could* be used to abrogate state sovereign immunity in bankruptcy suits.[412] This is the first, and so far only, decision acknowledging Congress's power under Article I to authorize citizens to sue states. Writing for the majority, Justice Stevens explained that this result was not inconsistent with the precedent because the Bankruptcy Clause was not at issue and not fully debated in *Seminole, City of Boerne, Alden,* or *Florida Prepaid*.[413] In addition, Justice Stevens cited the legislative history of the Bankruptcy Clause, emphasizing the reasons why it was written into the Constitution. In particular, he stressed a justification based upon the overriding goal of ensuring "uniform laws upon the subject of bankruptcies."[414] In addition, he notes that the legislation passed in the wake of the Bankruptcy Clause's ratification demonstrates "that it was intended not just as a grant of legislative authority to Congress, but also to authorize limited subordination of state sovereign immunity in the bankruptcy arena."[415]

While *Central Virginia* is not directly applicable to the copyright context, it may suggest that Congress's perceived authority to abrogate state sovereign immunity under Article I is still evolving. In particular, the legislative impetus for the drafting the Bankruptcy Clause neatly aligns with the impetus for drafting the Patent and Copyright Clause — ensuring uniformity and certainty with respect to federal intellectual property law. In addition, when the Act was amended in 1976, federal copyright law also became fully comprehensive, as well as uniform; the Act articulates in §301 that federal law and courts have exclusive jurisdiction over copyright claims and state copyright and copyright-like privileges are preempted. While *Central Virginia*

[412] *Central Virginia Community College v. Katz* 546 U.S. 356 (2006).
[413] *Central Virginia* at 363.
[414] *Central Virginia* at 368-369.
[415] *Central Virginia* at 363.

does not overrule the earlier rulings related to sovereign immunity in the copyright context, this decision reminds legal doctrine that what appears fairly well-settled and static may in fact be undergoing a process of evolution and transformation.

Another, perhaps unanticipated, consequence of state sovereign immunity jurisprudence is that further complicates rulemaking concerning fair use and the library and archives exemptions. For example, in 2008, a section 108 Study Group convened to report and advise the Copyright Office on rulemaking to update the section in light of evolution technologies and library and archive practices. That committee reported:

> "The study group believes that it would've been able to reach greater consensus on certain proposed changes to section 108 if not for the issue of sovereign immunity. Many of the largest U.S. libraries are state operated, and rightsholders are concerned that they will not be able to obtain effective redress should such libraries exceed the bounds of section 108 and fair use."[416]

In other words, the specter of sovereign immunity compels some rightsholders to seek even tougher restrictions and enhanced enforcement mechanisms (through the use of technical protection measures, for example) to make up for their perceived (or assumed) inability to adequately redress their interests through the courts. This may lead some rightsholders toward extra-legal mechanisms, such as technical protection measures ("TPM"), as a way to adequately protect and enforce their copyrights.

The Digital Millennium Copyright Act, codified in §§1201-1205 of the Act, expressly prohibits the circumvention of technological measures that effectively control access to protected works. These provisions signal a stark departure from the existing copyright law — which had always been oriented around making, modifying, and publicly communicating copies — to now also including *accessing* copies even when those copies are for personal use and have been lawfully obtained through purchase or other means. Because the mass digitization project which is the primary focus here involved the digital conversion of print materials — which have not historically employed TPM — the DMCA may be of limited immediate relevance. However, its provisions may present a serious problem in the future, particularly as libraries and archives collections are increasingly comprised of born digital materials.

[416] Section 108 Study Group Report (2008).

The Act permits libraries to circumvent TPMs if the purpose is to access and review the protected work in good faith for purposes of determining whether or not to purchase it. As one commentator notes:

> "The exemption is narrowly and meticulously constructed, and a library is subject to serious legal penalties if it utilizes the exemption but is later determined to have misapplied the law. One has to seriously question whether the benefits of attempting to use this exemption will outweigh the accompanying risks of possible liability."[417]

Ultimately, this issue may become moot as digital content acquired by libraries and archives is increasingly subject to negotiated licensing agreements with publishers. The terms of those agreements are generally understood to control the conditions of access and use. As we saw in Chapter II, the terms of licensing contracts govern library and archives privileges under §108. While there are growing concerns amongst the library and archives community that license terms may erode core library practices under §108, the rule has not been reformed. An even greater uncertainty exists with respect to the extent to which contract terms might limit or preempt fair use or other exceptions under the Act with respect to digital content. This continues to be an active area of discourse and debate.

In terms of fair use and transformative use jurisprudence, there are a number of key post-2004 decisions that may bear on *Authors Guild v. HathiTrust.* As Chapter II described, fair use is a four-factor balancing test requiring courts to analyze a secondary (allegedly infringing) use in relation to the original work. Since the Supreme Court's ruling in *Campbell v. Acuff Rose*, transformative use (which is tied to the first fair use factor) is generally deemed as the most important determinant. While a use need not be transformative in order to qualify as a fair use, a finding that a use is transformative tends to have a dispositive influence with respect to the remaining fair use factors, as will be demonstrated in the following discussion.

In 2006, the Second Circuit Court of Appeals ruled in *Bill Graham Archives v. Dorling Kindersley, Ltd.* that shrunken reproductions of the Grateful Dead's concert posters and tickets chronologically placed along a timeline for inclusion in a biographical history about the musical group was a transformative use.[418] In particular, the Court noted that the purpose and character of the secondary use — enhancing the biographical information contained in the book with a

[417] Crews, Kenneth D. *Copyright law for librarians and educators: Creative strategies and practical solutions*. American Library Association, 2012, p. 98.
[418] *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605 (2d Cir. 2006).

reduced size of image displayed along with informational commentary — is transformative with respect to the original's primarily expressive nature.[419]

Cases in other circuits have also held that wholesale copying of protected works may be transformative where the use and purpose of the copies was clearly distinguishable from those of the original. For example, in *AV ex rel. Vanderhye, v. iParadigms*, the Fourth Circuit held that the copying and archiving of students papers for use in conjunction with anti-plagiarism software was transformative because the purpose and character of the secondary use was to identify and discourage plagiarism as whereas the originals were primarily expressive in nature.[420] In addition, in *Perfect 10 v. Amazon.com,* the Ninth Circuit held that wholesale copying of Internet images for use in the provision of web-based search services was transformative because the purpose and character of the secondary use was to point or direct a user to a source of information whereas the purpose of the originals were primarily expressive.[421] Furthermore, notwithstanding the test for transformative use articulated in *Campbell,* the Ninth Circuit held that a secondary work need not add anything "new" in order to be transformative. The Court in *Perfect 10* held that "even making an exact copy of a work may be transformative so long as the copy serves a different function than the original."[422]

In a footnote accompanying *Perfect 10,* the Court distinguishes the facts from earlier cases saying in *Texaco* a corporation made photocopies of copyrighted articles for use by its researchers. The court concluded that the majority of the copies served "the same basic purpose that one would normally seek to obtain the original — to have it available on his shelf for ready reference."[423] Likewise, the Court distinguished the ruling in *UMG Recordings, Inc. v. MP3.com, Inc*, where that court had found that conversion of CDs into computer files for use by users over the Internet was not transformative because the use to which the copies were put was not

[419] *Bill Graham*, 448 F. 3d at 609.
[420] AV ex rel. Vanderhye v. iParadigms, LLC, 562 F. 3d 630 (4th Cir. 2009).
[421] Perfect 10 v. Amazon.com, Inc., 508 F.3d 1146, 1165 (9th Cir.2007).
[422] Perfect 10 v. Amazon.com, Inc., 508 F.3d 1146, 1164 (9th Cir.2007). In an accompanying footnote, the Court distinguishes it the facts from earlier cases saying, in *American Geophysical Union v. Texaco, Inc.,* 60 F.3d at 913, a corporation made photocopies of copyrighted articles for use by its researchers. The court concluded that the majority of the copies served "the same basic purpose that one would normally seek to obtain the original — to have it available on his shelf for ready reference." *Id.* at 919. Likewise, the Court distinguished the ruling in *UMG Recordings, Inc. v. MP3.com, Inc.,* 92 F.Supp.2d 349, 351 (S.D.N.Y.2000), where that court had found that conversion of CDs into computer files for use by users over the internet was not transformative because the use to which the copies were put was not different than the use for the originals.
[423] *American Geophysical Union v. Texaco, Inc.,* 60 F.3d at 919.

different than the use for the originals.[424] In other words, the Ninth Court holds the position that wholesale copying for iterative, socially productive, non-expressive purposes may be transformative.

As mentioned earlier, a finding that a use is transformative tends to have a dispositive effect on the other three fair use factors. For example, the Second Circuit noted that "the second factor may be of limited usefulness where the creative work of art is being used for a transformative purpose."[425] Furthermore, if the transformative purpose of the secondary use requires that the entire original work be copied, as was the case with the intermediary copying involved in *AV ex rel, Perfect 10, and UMG Recordings*, then copying a work in its entirely may still be deemed a fair use. Finally, with respect to the fourth fair use factor — impact on the market for or value of the original — under *Texaco*, courts are instructed to only consider "traditional, reasonable or likely to be developed markets"[426] which, by definition, exclude "transformative markets."[427] Thus, a determination that a secondary use is transformative, while not determinative, in practice seems to carry a disproportionate weight in comparison to the remaining fair use factors.

As the above jurisprudential refresher indicates, a number of key decisions involving fair use and transformative use took place in the period between 2004, when the University of Michigan digitized its entire library through the Google Library Project, and the filing of the *Authors Guild v. HathiTrust* lawsuit. Before discussing the courts' sensemaking and decision-making in that case, I will first discuss HathiTrust's and its members' sensemaking and response to the lawsuit.

Sensemaking around the lawsuit

As an initial matter, the sensemaking HathiTrust's members around the lawsuit is, in some ways, contextually bound to the so-called companion lawsuit of *Authors Guild v. Google*, which was filed much earlier (in 2005) and is still active today, over a decade later. In that case, the Authors Guild and publishers sued Google for copyright infringement based on its mass digitization of works in connection with the Google Library Project. Google defended on the

[424] *UMG Recordings, Inc. v. MP3.com, Inc*, 92 F.Supp.2d 349, 351 (S.D.N.Y.2000).
[425] *Bill Graham*, 448 F.3d at 612.
[426] *American Geophysical Union v. Texaco, Inc.*, 60 F.3d at 931.
[427] *Bill Graham*, 448 F.3d at 614.

basis of fair use but that case was not adjudicated on its merits; instead the case largely stalled out over protracted negotiations around a hotly debated settlement agreement between Google and the plaintiffs. In the spring of 2011, several months before the case was filed against HathiTrust, the district court finally rejected the settlement agreement (which by this time had been amended) and the case was effectively sent back to the drawing board.

Given its incredibly protracted (and expensive!) litigation with Google, the Authors Guilds' lawsuit against HathiTrust came as a surprise to many observers. This was especially true given that the two cases had very similar facts (both revolving around the same mass digitization project) but with one very potentially important difference: HathiTrust was not a private, commercial firm like Google but rather a partnership of academic libraries oriented around strong educational and research missions. These characteristics of HathiTrust made it, arguably, a far more sympathetic defendant in a case that would ultimately turn on fair use.

Indeed, for the key administrators at the University of Michigan were surprised when the Authors Guild filed suit:

> "We did not expect to get sued for copyright infringement of this kind. I thought that we might get sued by certain individual rightsholders hoping to make a low probability/high stakes run at particular usage that might be infringing, and would carry statutory damages which, of course, they probably wouldn't under the 11th amendment. But I thought we might get individual suits from authors and authors' families who were unhappy about particular uses. I did not expect a mass class action suit of this kind and indeed, after the Google suit was filed, we weren't defendants at all for years. They sued us many years later.
>
> I didn't expect the suit because I didn't think we were very suitable. I still don't. We weren't suitable because we weren't doing anything that looks like infringement. We were copying, that's for sure, and we were keeping copies that Google gave to us, but our usage didn't look to us, in practice, like their usage."

Other participants viewed the lawsuit as authors and publishers grasping at straws, trying to preserve their preexisting business models despite the changing sociotechnical environment. This perspective was communicated in various ways. Those with a technical bent made comments like: "People on the outside are still threatened by the future and what technology means, and they don't get the open is the future." While those who had already had long careers in academic librarianship characterized the motivations of the plaintiffs slightly differently, said things like:

> "It's very disappointing how the authors have responded to this. It reminds me a lot of publishers, years ago, when we think about interlibrary loan of journal articles and how restrictive everyone was trying to be especially as journals went digital. If you would make these things affordable, we would give up our interlibrary loan operations and just buy all the stuff from you. If it were five dollars for an article, we would have professors that would buy the articles. But it's really hard to get people to think of what their business model might be in a new environment. And so the first reaction is just to close ranks and make it as limited as possible. These are groups that are just really worried that something that they could make money on will be given away for free."

What quotes like these suggest is that there was a sense that the lawsuit was motivated by a fear, shared by authors and publishers, that HathiTrust and the Google Library Project represented not just a loss of control over protected works, but a lost opportunity for further exploiting the works, never mind the fact that, as the history of large-scale digitization makes clear, it was a functional impossibility that authors and publishers would ever be able to carry out a mass digitization project or the creation of an organization like HathiTrust.

Another reason why the lawsuit came as a surprise to those at Michigan was that Michigan and HathiTrust were fairly open and transparent with the Authors Guild, the Association of American Publishers ("AAP"), and the U.S. Copyright Office about HathiTrust, its efforts, and initiatives. Things had even gotten as far as a scheduled meeting at Michigan amongst the some of the interested parties:

> "We had an initial meeting in September or early October to discuss and talk about what we were doing with the AAP and the Authors Guild. The AAP came and the Authors Guild didn't, because they sued us a week or two before the meeting was supposed to take place. And that was just to have a conversation."

A plausible theory advanced by a number of the participants in this study links the primary motivation for the Authors Guild's lawsuit against HathiTrust with the emergence of the Orphan Works Project. One of the participants said: "I've often wondered if our announcement that we were going to participate in the Orphan Works Project was the reason why Cornell was one of the defendants in the HathiTrust suit." In particular, there was a sense that early missteps in, and misconceptions about, the Orphan Works Project may have prompted the litigation. One of HathiTrust's members reflected:

> "I've always believed that it was the Orphan Works Project that prompted the lawsuit against the HathiTrust in the first place — the anger about how poorly it had been implemented by Michigan. And then it was almost like, 'As long as

> we're suing over it, we might as well sue about the general scanning as well, too. We'll open a second front in the war against Google." What didn't make any sense to me was why they would run the risk that a court would find that, when libraries are doing it or are involved, it's a fair use, and so that might impact upon the Google decision. It just seemed like a really stupid move for them to make. But I think it's because they were just so incredibly angered by the poor implementation of the Orphan Works Program. Until we can get into the Authors Guilds' archives and look at the internal memos and discussions and find out what it was thinking, I guess we'll never know."

Another participant who was more directly involved in the Orphan Works Project cited acknowledged that the initial vetting process which led to some non-orphan works slipping onto the list may have played a role in the Authors Guild's decision to file suit:

> "We published the potential orphan works list. We saw that some works had mistakenly gotten through. We pulled them down. We went back and reviewed the process. We refined the process. There was a lot of attention given to process. Documentation on process developed and we shared that with the Authors Guild in those early days after the initial suit was filed trying to stimulate the dialogue there. I think that everybody felt that a good dialogue about these things is important. But we didn't succeed in doing that; there has not been a good dialogue. We sat down with Maria Pallante[428] and Alan Adler[429] to talk about a way that we could collectively — the Copyright Office, the AAP, the libraries — could undertake an Orphan Works Project."

Discussions on how various stakeholders might undertake or design a collaborative Orphan Works Project did not get very far, in part, because of the lawsuit and widespread misconceptions about the nature and operation of the Project at Michigan. HathiTrust's Executive Director at the time recalled:

> "The popular misconceptions about what we were doing made their ways pretty heartily into Congress and the Copyright Office. They believe that the entire corpus of materials, all of it including in-copyright works, was being made available on campuses at partner institutions. They also thought that when we made those orphan determinations, and posted the list of possible orphan candidates, we also opened those works to the world. And as you know, no orphans were made available to anybody."
>
> Maria Pallante invited us to come talk to her in Washington DC in early December, 2011, right in the middle of the lawsuit. And when I described to her what we were actually doing, her reaction was: 'Why are we having this conversation?' Her visual reaction seemed not troubled."

[428] Maria Pallante is the current Register of Copyrights for the U.S. Copyright Office.
[429] Alan Adler is Vice President for Legal and Governmental Affairs for the Association of American Publishers.

In other words, much of the impetus for the lawsuit may have been based on misconceptions about a Project that had not even reached the implementation stage. Even within the Project, there were misconceptions, or at least ambiguities, about its scope:

> "The Orphan Works Project didn't go anywhere. We hadn't even worked out the details. But I think our thinking was that if the work went through the Orphan Works process and was identified as being an orphan work, we would then make that text available to a user of the Cornell system in electronic form. But because this process never really developed, none of those details were worked out."

The Orphan Works Project had only begun a few months earlier and was still very much an experiment — an experiment that, so far, did not involve opening up access to any in-copyright works whether they were suspected to be orphans or not.

Pallante's purported lack of concern once the accurate details of the Project had been communicated also seems perfectly reasonable given the fact that libraries and archives had long-standing traditions of digitizing their special collections which were often comprised of presumptive orphans.

A librarian involved in the Orphan Works Project at Michigan reflected on her previous experiences working at the Library of Congress:

> "Special collections materials, much of which is treated as orphan works, has been digitized and made available for the last 20 years. That's been an uncontroversial practice that's existed since I was at the Library of Congress in the '90s. But trying to actually address it head-on, maybe that was too rigid."

Obviously Michigan's Orphan Works Project would have involved materials outside of special collections, but this observation suggests that practices within libraries and archives co-evolve and build upon each other. As another participant noted, Pallante's response aligns with the general sense that "practice in new areas should inform the legislation rather than the other way around." Although legislators had explored reforms to address the orphan works problem, there was a strong sentiment that a "wait and see" approach is preferable to a proactive approach, at least where Congress and the Copyright Office is concerned.

In the case of the Orphan Works Project, the lawsuit stifled any real opportunity to see what the results of the experiment might be. When the lawsuit was filed, the Project was immediately suspended by Michigan's Office of the General Counsel. We never got a chance to

see how the Project might grow and evolve, and solve or fail to solve the challenges of orphan works. The Project had the potential to remove some of the dead weight generated by absentee copyright rightsholders.

In a particularly tragic twist of fate, as the next section describes, the Authors Guild's claims based on the Orphan Works Project were dismissed as unripe, meaning that this issue remains wholly unresolved and the challenges of absentee rightsholders persist.

Authors Guild v. HathiTrust

The basis of the Authors Guild's lawsuit against HathiTrust has surfaced in a number of other sections of this work. Nevertheless, this section will begin with a brief description of the case before discussing the resulting judicial decisions at both the district and circuit court levels.

Background

According to the district court opinion, the Authors Guild sued HathiTrust and several of its members for copyright infringement stemming from the mass digitization efforts of the Google Library Project. At the time the opinion was written, in October, 2012 (approximately one year after it was filed), HathiTrust contained over 10 million volumes, 73% of which are believed to be in-copyright. Between the initial digitization and the final deposit in HathiTrust, the Authors Guild contends that twelve unauthorized digital copies of each work have been made.[430]

The court effectively splits HathiTrust's digital corpus into two categories. The first category includes works with known authors. With respect to these works, HathiTrust and its members use them in the following three ways: (1) for full-text searches; (2) for preservation; (3) in the provision of access for people with certified print disabilities.

As the court describes, the full-text search capability enables users to search for a particular term across the entire corpus. For works that are not in the public domain or for which the copyright holder has not authorized use, the search results only return information on the number of instances, and the page numbers, on which the term appears. For works in the public domain or for which the copyright holder has given permission, the search term is displayed within a snippet of additional surrounding text and a link to the full page is provided. (Figure 17)

[430] *Authors Guild, Inc. v. HathiTrust*, 902 F. Supp. 2d at 448.

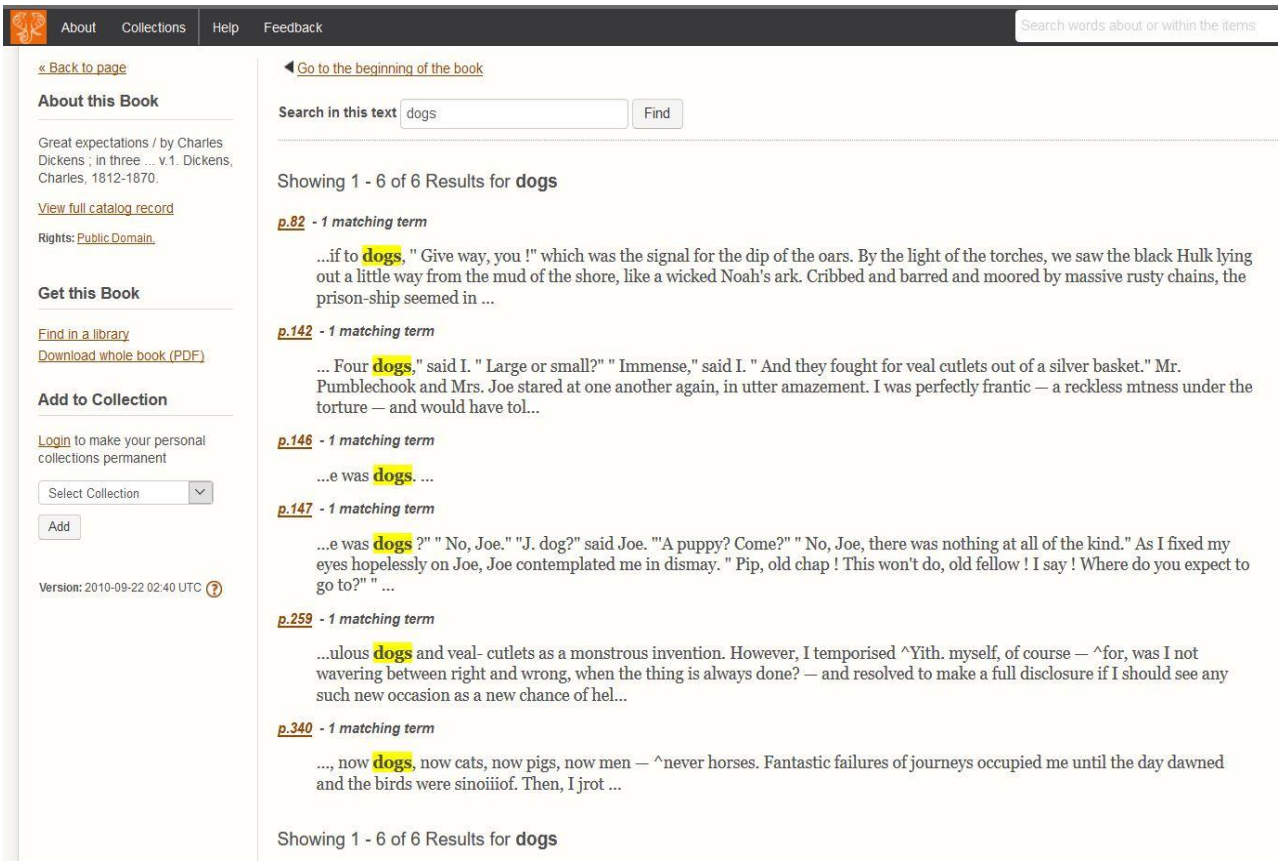


Figure 17. HathiTrust full-text search results for the query "dogs" in "Great Expectations," Charles Dickens.

With respect to access for certified print-disabled persons, the court notes that, as a general matter, digitization has had a tremendous effect on the blinds' ability to access print materials. Prior to digitization, "the blind could access print materials only if the materials were converted to braille or if they were read by a human reader, either live or recorded."[431] Through Michigan's (and by extension, HathiTrust's) service:

> "Print-disabled individuals read digital books independently through screen access software that allows text to be conveyed audibly or tactilely to print-disabled readers, which permits them to access text more quickly, reread passages, annotate, and navigate, just as a sighted reader does with text."

Through the university's secure system, certified print-disabled individuals have full access to materials and, as a result, academic participation by this community has been "revolutionized" by HathiTrust.[432]

[431] *Authors Guild, Inc. v. HathiTrust*, 902 F. Supp. 2d at 448.
[432] *Authors Guild, Inc. v. HathiTrust*, 902 F. Supp. 2d at 449.

The second category of works at issue in the case are those with unknown authors, i.e. orphan works. The court outlines the Orphan Works Project review work-flow process as follows:

1. Is the commercially available for sale? If yes — STOP — if no, continue;
2. Is the copyright holder contactable? If yes — STOP — if no, continue;
3. HathiTrust lists the bibliographical information for the work on the HathiTrust Orphan Candidates webpage for ninety days. Does a rightsholder come forward? If yes — STOP — if no, continue;
4. The full text of the work becomes viewable on HathiTrust to UM students, professors, and other authenticated users and visitors to libraries at UM's campuses. The number of users permitted to view a given work limited at any one time is limited to the number of copies held by the UM library.

The court notes that the Orphan Works Project had been suspended indefinitely after a number of works made their way onto the Candidates list in error.[433] All of the named Plaintiffs in the case, except Indiana University, had agreed to participate in the Orphan Works Project, a fact that lends some credence into the just-discussed speculations of some of the Project's participants. (Perhaps Indiana was included because it played such a pivotal role in making the shared digital repository and HathiTrust possible.)

The court rejected the Authors Guild's claims based on the Orphan Works Project because the cause of action was not ripe. The Project had been suspended without indication of renewal. No access to the so-called orphans had ever been granted. Thus, the court determined that there was no actual case or controversy because the "mere possibility" the one of the Plaintiffs' works might be included on a future Candidates list, or made available, is not enough. Therefore, the court granted summary judgment in favor of HathiTrust with respect to orphan works, leaving the three other uses — full-text search, preservation, and access for the print-disabled — intact.

In terms of the procedure of the case, both parties moved for summary judgment. For a court to resolve a fair use determination at the summary judgment stage it must determine that no genuine issues of material fact exist that could lead a reasonable jury to side with the non-

[433] *Authors Guild, Inc. v. HathiTrust*, 902 F. Supp. 2d at 449.

moving party.[434] Because fair use is a mixed question of law and fact, granting summary judgment in favor of HathiTrust necessarily requires that a high burden of proof and persuasion has been met. Essentially, to side with HathiTrust, the court would have to conclude that, *even if everything the Authors Guild asserts is true*, no jury could reasonably conclude that HathiTrust's activity constituted copyright infringement and therefore the Defendant is entitled to judgment as a matter of law.[435]

With the facts and procedural background of the case laid out, attention now turns to the sensemaking and decision-making of first the district court then, because Authors Guild appealed, the Second Circuit Court of Appeals. A brief section of critical reflection, interpretation, and discussion follows each of the courts' decision, followed by a final conclusion.

DISTRICT COURT DECISION

The district court ruled in HathiTrust's favor, finding that the three uses HathiTrust permitted were transformative uses. This section walks, factor by factor, through the court's fair use analysis. The first factor considered by the court is the "purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes."[436] The court then describes how the mass digitization project was undertaken with several goals in mind including preservation, search and discovery, and enhanced access for print-disabled patrons, and that all of these uses serve an educational or research purpose which tilts the first factor in HathiTrust's favor. In addition, the court notes that with respect to in-copyright works, access to the actual text of those works is strictly limited to persons with certified print disabilities and for purposes of full-text search.

Importantly, the court concludes that each of the three uses at issue in the case are transformative. The key inquiry in transformative use determinations is:

> "Whether the new work merely supersede[s] the objects of the original creation ... or instead adds something new, with a further purpose or different character, altering the first with new expression, meaning, or message."[437]

---

[434] *Castle Rock Entm't, Inc. v. Carol Publ'g Grp, Inc.*, 150 F.3d 132, 137 (2nd Cir. 1998).
[435] *Citibank, N.A. v. Morgan Stanley & Co. Int'l, PLC*, 724 F.Supp.2d 407, 414 (S.D.N.Y.2010).
[436] 17 U.S.C. §107(1).
[437] *Authors Guild, Inc. v. HathiTrust*, 902 F. Supp. 2d at 459-460 citing *Campbell,* 510 U.S. at 575, 114 S.Ct. 1164.

Drawing upon the *Graham, Perfect 10, and AV ex rel* trio of cases, the court determines that, even though the protected works had been copied in their entirety, full-text search, preservation, and access for print-disabled patrons are uses that are functionally unrelated to the expressive intent of the originals.

As previously described, a determination that a use is transformative strongly influences the court's reasoning with respect to the remaining three fair use factors. The court essentially brushed over the second and third factors — nature of the copyrighted work and amount of the work copied — before also concluding that Authors Guild has not suffered any cognizable harm to its traditional market because HathiTrust's uses fall within a "transformative market" that is necessarily outside the purview of the Authors Guild. Thus, the district court concluded that all four fair use factors tilted in favor of HathiTrust and therefore no genuine issues of material fact existed that could prevent a reasonable jury from ruling in HathiTrust's favor.

Given the fairly straightforward application of fair use and transformative use doctrine, it is somewhat surprising how the decision takes a markedly different turn toward its end. It seems, in some ways, to backtrack or perhaps buttress the seemingly clear cut doctrinal support just outlined. It does so first by stressing that its rationale in the case was primarily motivated by "the goal of copyright itself, whether 'promoting the Progress of Science and useful Arts would be better served by allowing the use than by preventing it'"[438] rather than doctrine. Even more striking, however, is the unusual transition in tone at the end of the opinion where it shifts from the dispassionate and impartial rhetoric characteristic of most court opinions to Judge Baer's personal, almost confessional, reflections on the case:

> "Although I recognize that the facts here may on some levels be *without precedent*, I am convinced that they fall safely within the protection of fair use such that there is no genuine issue of material fact. I cannot imagine a definition of fair use that would not encompass the transformative uses made by Defendants' MDP *and would require that I terminate this invaluable contribution to the progress of science and cultivation of the arts* that at the same time effectuates the ideals espoused by the ADA."[439] (Emphasis added).

In other words, he seems to acknowledge that existing doctrine may not stretch so easily and completely to suit the emerging contours of the mass digitization project but, nevertheless, it

[438] *Authors Guild, Inc. v. HathiTrust*, 902 F.Supp.2d 445, 458 (Dist. Court, S.D.N.Y. 2012).
[439] *Authors Guild Inc. v. HathiTrust*, 2012, p. at 464.

so clearly promotes the overriding goals of the copyright law that we simply cannot let the law's means defeat its purpose.

The innovative deviance framework developed in this thesis can help us understand and explain Judge Baer's sensemaking in this case. He seems to acknowledge that, when copyright doctrine and precedent do not adequately carve out safe paths for particular copyright-related behaviors, i.e. prior to this decision legitimate non-infringing means did not exist for copying millions of in-copyright books, the system will find ways to support and protect those behaviors because they accomplish the overriding goals of copyright law. Judge Baer's rationale demonstrates how innovative deviance plays a crucial role in teaching copyright law how to change, fostering not just transformative use in the copyright sense, but transformation more broadly construed.

The district court opinion raises a number of other interesting questions and tensions that are relevant to large-scale collaborative knowledge infrastructure development involving potentially infringing behavior. In particular, the decision raises questions around timing and the extent to which the subjective intent of those doing the digitizing matters in fair use determinations. As earlier chapters of this work demonstrate, the University of Michigan's decision to partner with Google in the digitization of its library was complex and motivated by a series of pragmatic, ideological, and reputational concerns. We know that Michigan desperately wanted to make a back-up copy of the entire library and believed it was justified in doing so under the so-called "dark archive" which tends to support the preservation purpose cited by the court. But as we also learned, many of the reasons it had for undertaking the mass digitization project were far more inchoate and uncertain than the court seems to acknowledges. Instead of being predetermined and deliberate, the uses of the digitized corpus emerged over time through complex co-evolving technical, legal, and organizational interactions. The court's reasoning here suggests that the purpose and character of the use at the time the alleged infringement took place is irrelevant. Rather, we look at the purpose and character of the uses at the time the lawsuit was filed.

In the context of large-scale collaborative projects and knowledge infrastructure development, this creates somewhat of a conundrum. As we saw, HathiTrust emerged several years after the mass digitization project was underway. Moreover, the uses adjudicated by the court emerged over time (although it appears that Michigan had been provided the services prior

to the creation of HathiTrust and its partners). However, the cause of action arose when the digitization occurred, in 2004, before time passed allowing the subsequent (defensible) uses to emerge. Are we to understand that HathiTrust was simply "lucky" that the Authors Guild waited so long to sue — or at least long enough that HathiTrust had time to develop the technical and organizational infrastructure to implement the uses at issue in this case? If Michigan had been sued for copyright infringement after the mass digitization project started, but while the scans were still in a dark archive, how differently things might have turned out! What cautionary lesson might future innovators and investors, and rightsholders, take from this?

CIRCUIT COURT DECISION

The Authors Guild appealed and the Second Circuit affirmed the lower court's holding with respect to full-text search and enhanced access for print-disabled individuals — they were fair uses. The Appeals Court, however, vacated the lower court's holding with respect to preservation because it found that the Authors Guild lacked standing to bring the claim.[440]

Despite reaching the same ultimate result with regard to the full-text search, the court disagreed with the lower court's transformation analyses. In particular, the Court of Appeals took issue with Judge Baer's goal-motivated decision-making just discussed:

> "Contrary to what the district court implied, a use does not become transformative by making an 'invaluable contribution to the progress of science and cultivation of the arts.' Added value or utility is not the test: a transformative work is one that serves a new and different function from the original work and is not a substitute for it."[441]

In other words, innovative deviance cannot be "rescued" by how well it accomplishes the goals and purposes of copyright law. The court then concluded that the creation of a full-text searchable database is a "quintessentially transformative use" because the results of the search are "different in purpose, character, expression, meaning, and message" from the original work from which it is drawn.[442]

Likewise, the court took issue with the lower court's transformation analysis with respect to HathiTrust's improved access for print-disabled patrons, writing: "providing expanded access

[440] *Authors Guild v. HathiTrust*, 755 F. 3d 87 (2nd Cir. Court of Appeals 2014) at 104.
[441] *Authors Guild v. HathiTrust*, 755 F. 3d at 96 quoting Authors Guild v. HathiTrust, 902 F.Supp.2d at 464.
[442] *Authors Guild v. HathiTrust*, 755 F. 3d at 97.

… is not 'transformative'" because it does not "add something new to the copyrighted work"[443] and HathiTrust's purpose in improving access was no different from the purpose of the original works — both are primarily expressive. Transformation, the Court tells us, requires more than simply enabling a new audience to read a book.

While the court declined to find HathiTrust's improved access from print-disabled patrons transformative, it nevertheless found it to be a non-infringing fair use because HathiTrust took no more than was necessary to effectuate its valid purpose. In addition, the Court determined that HathiTrust did not harm the potential market for the original because (sadly) the fact that a market for handicap-accessible books is virtually non-existent necessarily forecloses the possibility of harm.

Both Courts reached essentially the same holding — HathiTrust's uses were found non-infringing, signifying that mass digitization for full-text search and the provision of access to print-disabled persons is non-infringing — but I do not think this means that libraries and others seeking a broad interpretation of fair use should grow complacent. The difference in rationales between the District and Circuit Courts in *Authors Guild v. HathiTrust* signals a growing tension between creative or expressive works, which traditionally form the core of copyright, and an emerging subset of iterative, non-expressive, fundamentally technical and non-expressive class of works. Fair use and transformative use jurisprudence arose out of the traditional, expressive, core of copyright, and its application to those sorts of work makes sense. If we listen to 2Live Crew's song, "Pretty Woman," we can hear how it is a transformation of Roy Orbison's original, "Oh, Pretty Woman."[444] Likewise, we can see how Annie Leibovitz's portrait of Demi Moore was transformed by Paramount in its poster for the film "Naked Gun 33 1/3: The Final Insult"[445] and how Richard Prince's work transformed Patrick Cariou's photograph.[446] (Figures 18 and 19)

[443] *Authors Guild v. HathiTrust*, 755 F. 3d at 101.
[444] These work formed the basis of the copyright infringement claims in *Campbell v. Acuff Rose*, US, 1995.
[445] *Leibovitz v. Paramount Pictures Corp.*, 137 F. 3d 109 (2nd Circuit Court of Appeals, 1998.
[446] *Cariou v. Prince*, 784 F. Supp. 2d 337 - Dist. Court, SD New York 2011.

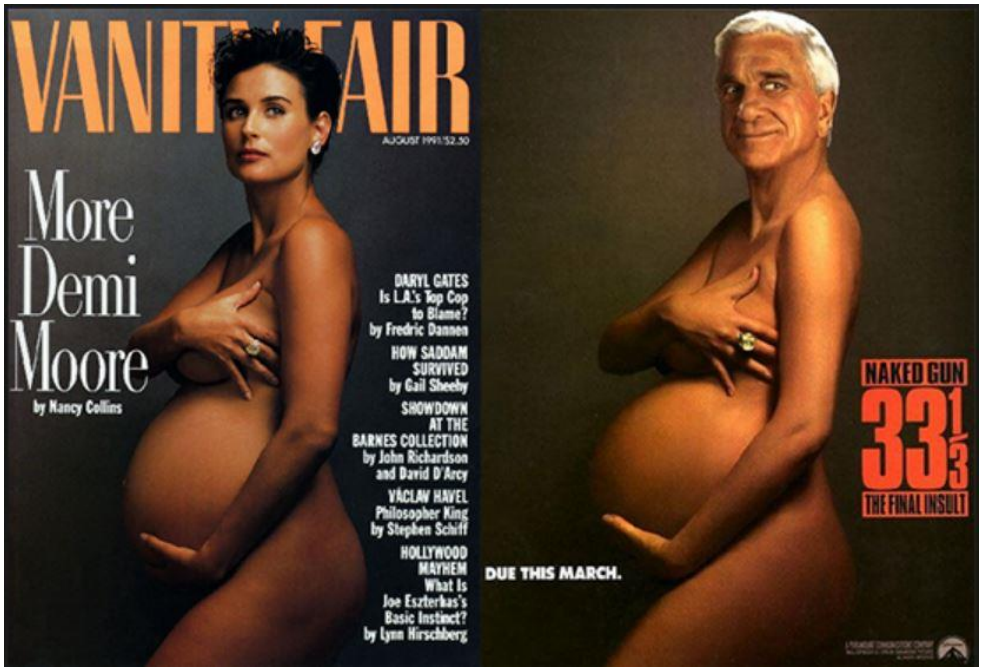


Figure 18. Annie Leibovitz's photograph on the left, Paramount's movie poster on the right.

Figure 19. Patrick Cariou photograph on left, Richard Prince work on the right.

Fair use and transformative use doctrine begins to lose cogency as we move away from the traditional core of copyright. When we begin to consider uses like full-text search, for

example, it does not make as much sense to compare the secondary work — the search results showing the instance of the word "dogs" in Charles Dickens' "Great Expectations" — with Dickens' original.[447] (See Figure 17) The doctrine and precedent does not map particularly well to emerging, iterative, non-expressive, technology-oriented uses because the transformation we are interested in promoting is not captured by the secondary work. The transformation we are seeking to promote through things like HathiTrust's full-text search, and Google's thumbnail image search, and iParadigm's anti-plagiarism software is the broader sense of sociotechnical transformation. Full-text search is transformative because it fundamentally and beneficially changes scholarly communication practices. It makes new meanings, messages, and understandings possible by giving students, faculty, and others new tools for processing information, and in so doing promotes the overriding goals of copyright law. But the search results in and of themselves do not accomplish this.

We might speculate that this sort of understanding may have partially motivated Judge Baer's conclusions in the district court opinion. He seemed to recognize that the legal rules, tests, doctrines, precedents, and algorithms fall short in the case of innovative information technologies' iterative, non-expressive use of protected works. Arguably what HathiTrust, Google, Amazon, iParadigms and other are doing is more akin to "contributory transformative use" — providing tools and technologies so that others may make transformative secondary works — but, as of yet, no such legal doctrine exists. If Baer's decision had remained intact, it might have created space for judges to make determinations based upon the degree to which a use promotes the overriding goals of copyright. Instead, the Circuit Court tells us that the proxy — the transformative use test — is more important, in some ways missing the forest for the trees.

Conclusion

In this chapter we explored sensemaking around the *Authors Guild v. HathiTrust* lawsuit. In particular, we learned how participants interpreted the lawsuit as a consequence of the Orphan Works Project, citing its early missteps and growing misconceptions in the broader community as key factors. Not only was the Project, its process, and its purpose misunderstood, but the entire framing of the underlying issue as an "orphan works problem" may have been in error.

[447] This comparison is only meant as an example. Dickens' work is in the public domain and thus it is perfectly legal to conduct a full-text search or otherwise modify the original.

Participants suggested that the focus should have been on rights determinacy and discovering facts about the collective collection, particularly with regard to out-of-print works, instead of focusing on expanding the public domain.

We also saw how the lawsuit forced the still-nascent Orphan Works Project into a suspended state. In terms of legal strategy, this helped ensure that the court would agree that the Authors Guild's claims with respect to the Project were not ripe for adjudication. It also carried the unfortunate consequence of suspending momentum on solving the challenges of absentee rightsholders.

This chapter also explored judicial sensemaking through the two *Authors Guild v. HathiTrust* decisions. In particular, a comparison of the rationales developed by the courts highlighted the tensions between the goal of copyright and the legitimate means available for accomplishing its goals. The district court's opinion provided a broader support and protection for activities that promote the overriding purposes of copyright but yet do not fit neatly within existing doctrine and precedent. Judge Baer's rationale demonstrates how innovative deviance plays a crucial role in teaching copyright law how to change, fostering not just transformative use in the copyright sense, but transformation more broadly construed. The Circuit Court rejected that approach, however, emphasizing instead the need to adhere to formal rules and procedures instead of broad policy-oriented appeals.

While the outcome resulted in a victory for HathiTrust, I suggested that those seeking a broad interpretation of fair use would be wise to avoid complacency for several reasons. First, while the case against HathiTrust is over, the Authors Guild has filed for certiorari in their case against Google. If the Supreme Court hears the case, there is the potential that an adverse ruling would negatively impact HathiTrust and its partners. Second, the Circuit Court's rejection of Judge Baer's reasoning undermines copyright law's adaptability to technological and social change, and its ability to promote transformation more broadly construed. It suggests that the means are more important than the end. Strong protections for iterative, non-expressive forms of use are necessary to ensure that copyright law remains relevant, functional, and credible as technologies change and new forms of social practice around protected works continue to emerge. I suggested that a contributory transformation use doctrine might be small, but potentially meaningful, step in the right direction. I would also argue that a test which gives judges broad discretion to weigh the potential harm of a use against its potential social benefit, as

Judge Baer seemed to attempt, might also be a worthwhile reform. Even that, however, might not support so-called "consumptive uses" of the digitized content, e.g. reading the books. One might draw an inference that the next fair use battle might involve using the digital copies for reading purposes, within secure networked environments, perhaps limited to the number of print copies an institution holds in an off-site storage facility.

The outcome of the case also raises some potentially troubling consequences for innovative deviance in general. For example, one might wonder whether the *Authors Guild, Perfect 10, iParadigms,* and (so far) Google quartet of cases represent a custom-fit solution only available to very large and/or powerful companies or organizations. It is difficult to imagine how an individual, small-scale, more modest, and/or more distributed exercise of innovative deviance would fare under this precedent. But if we agree that deviance is a natural and unavoidable consequence of a disequilibrium between societies' goals and the institutionalized means to accomplish those goals, then we should seek broader protections and support for those forms of deviance that promote the goals of society through the creative use of illegitimate means. By limiting the protections of fair use to massive scale projects like those of HathiTrust and Google, we risk losing out on the societal benefits resulting from less heroic, less visible, but still significant, endeavors.

The *Authors Guild v. HathiTrust* lawsuit suggests that while fair use and transformative use are incredibly important, powerful, and flexible doctrines, they might not take copyright law far enough in protecting innovative deviance and supporting sociotechnical transformation on a broader scale. In addition, as the next chapter demonstrates, the lawsuit also impacted the organizational character of Michigan and HathiTrust, and their relationship to each other, and dampened the willingness to take chances and engage in innovative deviance.

Chapter IX: The Long Now of HathiTrust

"My dream is that when Star Trek becomes real in the twenty-fourth century, and Mr. Data is looking back into some historical archive and reading something on his screen, that may have come from the HathiTrust."

- CIO, a HathiTrust founding institution

The last two chapters each covered a major milestone in HathiTrust's evolution. This chapter traces the implications and aftermaths of the Constitutional Convention and the Authors Guild's suit against HathiTrust. In combination, these events signal a shift from HathiTrust's pre-and-early history — highly dynamic, reactive, action-oriented, and immediate — toward an organization searching for what we might call its "long now."[448] This chapter traces key aspects of that readjustment, as HathiTrust responds to the lawsuit, begins to implement the initiatives approved at the Constitutional Convention, and casts its gaze toward the future.

Organizational Readjustment

Notwithstanding the victory, the lawsuit had tremendous impact on the evolution of HathiTrust and its relationship to the University of Michigan. Recall that the lawsuit was filed just before the Constitutional Convention took place. The result was a sort of conceptual disjointing of HathiTrust. On the one hand, HathiTrust's members were seizing the opportunity to take control of the organization — making decisions about what projects it would undertake, how its governance would work, what its goals and mission ought to be. On the other hand, the lawsuit foregrounded HathiTrust's status as a service and property of the University of Michigan. Together, the events signaled a period of significant transition for HathiTrust.

A number of key personnel changes shaped and were shaped by the organizational reconfigurations happening. The University Librarian at Michigan resigned from his post,

[448] The Long Now was a term coined by Brian Eno who, together with Stewart Brand and others, have created a foundation that seeks to encourage long-term thinking, serving as a counterpoint to todays "accelerating culture." http://longnow.org/about/

returning to a position he half-jokingly calls "just a humble school teacher." The Executive Director of HathiTrust — who had been its lead architect at the start and continued to live, eat, and breathe HathiTrust through the Convention — resigned from his post as HathiTrust Executive Director and AUL at UM to take a position at another University. HathiTrust was, for a period of time, in somewhat of a holding state as its governance structure emerged and as leadership at UM changed. As one participant recalled, "after the Constitutional Convention, a lot fell dormant." It took time to elect an Executive Director, put together an advisory board and form committees, draft bylaws and so forth.

During this period tensions arose, likely exacerbated by the pending copyright lawsuit, between the University of Michigan's General Counsel's Office and the interim director over implementing some of the initiatives approved by HathiTrust's membership. For example, the government documents project had been approved by HathiTrust's membership but had, for years, been a source of conflict between HathiTrust's leadership and UM's general counsel's office. In fact, disagreements about how to approach the digitization of government documents arose as early as 2006, two years prior to HathiTrust's official launch. One of the lead librarians who spearheaded Michigan's digitization efforts recalled:

> "When I was the co-interim Director of the University of Michigan Library, an attorney who worked in the General Counsel's Office wrote me a scolding letter saying: 'You must not provide access to federal government documents because of the copyright risks.' As Director, you have the right to say this is what I'm going to do. And I wrote a response and was able to cite for him decisions from legal authorities and the government code of federal regulations on the inherent copyright status of things that were included in the public record."

When HathiTrust's membership ultimately approved the government documents initiative at the Constitutional Convention in 2011, these tensions, or seeming cross-purposes, between HathiTrust and the opinions and advice of the University of Michigan's legal counsel became more pronounced. The tensions may have played a role in hiring decisions with respect to both HathiTrust's new Executive Director and the new University Librarian at Michigan. It should be noted that repeated requests for interviews with both the key attorney at the General Counsel's office and the new Executive Director of HathiTrust were denied, and therefore the description and explanation of sensemaking is necessarily limited. Nevertheless, it appears that there was a compelling desire shared by key figures at Michigan to reduce conflict and reestablish a unity of vision and purpose around a new, perhaps more conservative, institutional character:

After moving to a new position at the helm of a different library, the former Executive Director of HathiTrust recalled a conversation he had with the University of Michigan's new Librarian regarding a body of older materials in HathiTrust that had been opened under GATT restoration:

> "I told him, 'This is wrong. These are older materials and you've got the protections in the law. In responding to things (i.e. take down requests), you can say, 'We're sorry.' And turn off access. You can make restorations. But right now for the vast majority of these things, there is no identifiable rightsholder. No one's been in touch. We're not getting claims. And he said that he disagreed. That it was about trust and authority and they needed to back off. And I think those kinds of decisions now are the decisions of the day."

When I interviewed the new University Librarian at Michigan, about HathiTrust's general approach to risk and innovation, he remarked:

> "I don't think of myself as risk-averse, but on a scale of legal risk aversion, the founder and first Executive Director of HathiTrust is at a very different place from where I am."

During this period of transition, key administrators at the University of Michigan became relatively more risk-averse. The institutional perspective shifted. As one participant described it, Michigan was "interested in the war not the battle. It means that when we make decisions, we have to consider the legal analysis of our practices." This was a marked departure from nearly a decade earlier when key decision-makers embarked upon the mass digitization project without much hand-wringing about the copyright issues. "I'm sure we did the legal analysis, but I don't remember doing it" was the response of one of the key decision-makers.

The world looked very different to Michigan's revised leadership:

> "We (Michigan) go out of our way to make HathiTrust operate as though it were independent of the University, except in one important regard and that is legal decisions. The definitive legal answer must remain with the University of Michigan so long is HathiTrust is in fact an extension of the University of Michigan. In the beginning HathiTrust was very idealistic, and Michigan was able to push it that way because it was a land-grant institution and couldn't be sued like Harvard could. They were very bold and idealistic and I think what's happening now is it's becoming much more realistic. It's much more grounded in the law. The lawsuit helped. It gave us perspective. So HathiTrust can have all the democratic governance, but around legal issues Michigan calls the shots."

In terms of the governance that emerged after the Constitutional Convention, the HathiTrust Executive Committee continued to govern HathiTrust until the Board of Governors

was established in April, 2012. The Board of Governors is composed of six members appointed by HathiTrust's founding institutions, six elected from the membership-at-large, and the Chief Executive Officer (who is also the Executive Director). Policy for HathiTrust is set by the Board but, as one participant notes, "many of the decisions do not need real Board action but happen through the Executive Director." With respect to HathiTrust's current Executive Director, a participant commented: "The current executive director of HathiTrust has a very different temperament then the founding Executive Director. It's like a different person for a different moment." Bylaws were passed in 2013, and they stipulated that there should be an annual member meeting; the first annual meeting took place in October, 2014.[449]

In addition to the Board and the Executive Director, a number of groups and committees have formed to research and report first to the Executive Director who then reports to the Board. For example, there have been groups working on the monographic print archive and government documents proposals and there is a working group for HathiTrust User Support. A Rights and Access Committee was formed to think of ways to improve copyright determinacy in support of new lawful uses of the corpus but, as one of the committee members noted: "Basic copyright criteria policies and copyright determinations were set before the Constitutional Convention and have not been called into question within HathiTrust." And a Program Steering Committee was formed to think about strategic directions for HathiTrust. One of the members of that committee remarked:

> "The way that the Executive Director appears to want to use the Program Steering Committee is to make recommendations to him. Because these papers are pretty long and they're written by people who are really, really into that stuff, so he's looking for the Program Steering Committee to sort of parse it out and make recommendations that he can then bring to the Board or the larger general membership."

Many of the day-to-day operations of HathiTrust were largely managed by its Assistant Director and a very small administrative staff. As the Assistant Director described it:

> "I work for the University of Michigan, I'm under that administrative structure, but my salary is 100% paid for out of the HathiTrust budget. I work 100% on HathiTrust. Of course many of the things that are in HathiTrust benefit the University of Michigan and other institutions. So, formally I work for the University of Michigan but I definitely see myself as working for HathiTrust because everything that I do is for HathiTrust."

[449] https://www.hathitrust.org/member_meeting_2014

This suggests that, while HathiTrust's organizational governance structure is forming, functionally HathiTrust seems to operate in a somewhat murky (or illusory) zone of organizational semi-autonomy. While participants noted that there has been "zero conflict" between HathiTrust and UM, I got the impression that there was a fair amount of hesitancy and uncertainty shrouding sensemaking and decision-making in this area. As a librarian at Michigan said:

> "There is a balance and we need to tread carefully. A lot of the legal decisions are ultimately Michigan's responsibility because of how HathiTrust is structured right now. There is not shared liability per se, but there is shared action. Where the decision-making ultimately happens … I don't know to what extent Michigan and our General Counsel's Office make final decisions. These are things that HathiTrust's Executive Director is trying to work through. The key attorney from the General Counsel's Office spoke at the fall meeting of HathiTrust, and I find it really amazing that that's the first time he's been invited to meet with HathiTrust. The Executive Director and the Executive Committee of HathiTrust are very thoughtful people and they're working through a lot of these things about shared decision-making, shared responsibility, how we transition this from a new thing to something that's going to be sustainable."

Over the course of an interview with the University Librarian at Michigan, the locus of decision-making authority and the relationships of power among Michigan and HathiTrust's Executive Director became somewhat more transparent:

> "So HathiTrust has its own board and they make all kinds of decisions. But the Executive Director of HathiTrust is a University of Michigan employee and reports to me. And he and I and the Board of HathiTrust have a very clear understanding that in terms of deciding what HathiTrust's priorities are, and all those other things, they're fine. But when it comes to legal advice and decisions, while we are happy to have the Board engage in opinion-ing … (trails off)
>
> There are only two things that I'm aware of that the HathiTrust Executive Director could do that would cause the University of Michigan to remove them. There's offenses that will always get you fired — fiscal impropriety, something like that. But decision-wise, strategy-wise, if our attorney provided advice for HathiTrust — and General Counsel here is very good about saying: 'We provide advice, the decision authority rests with the administration, not with general counsel' — so, if General Counsel provides advice, and I concurred with that, and HathiTrust's Executive Director went rogue … that would be a problem. Right? But that's not gonna happen. From the very beginning in the hiring, we had this clear understanding.

> And it is my belief that over time, when the dust from the lawsuits settle, the world will be better and HathiTrust will become truly independent of the university. But it ain't time yet."

There were a number of factors interacting during this period that contributed to a shift in institutional character, reconfigurations of power, and affected the evolution of HathiTrust. Chief among these factors were the lawsuit, the emergence and establishment of HathiTrust's governance, and key personnel changes.

The result so far has been that HathiTrust's perspective stretched. It is no longer focused on the battle, but the war. One of the words that kept surfacing during my interviews with HathiTrust members during this period was "sustainability." Participants seemed primarily interested in the long now of HathiTrust. One of the ways sustainability was discussed was in relation to tensions between legal decision-making authority — which pretty clearly rests with the University of Michigan — and fiscal decision-making authority — which rests with HathiTrust's Board of Governors.

The University Librarian at Michigan reflected on this issue, saying:

> When the new Executive Director joined, immediately there were some decisions that were made that were in part influenced by our legal analysis. We made some changes and it caused the Board to suddenly react and say, 'Wait a minute, what do you mean Michigan says "no"?!' This is a very useful thing. We've had conversations. I've met with the Board. We've talked about everything.
>
> But this creates an interesting tension because the Board of HathiTrust has complete fiscal authority, and even though Michigan holds the money, HathiTrust decides how to spend it. I would say that most of the people in the library community see HathiTrust's budget as an extension of their own budget. They focus on the fact that HathiTrust has a big reserve, thinking 'Why are my dues structured the way that they are? Shouldn't we cut the membership dues to make it easier on librarians and help their budgets?' This potentially sets up a dynamic where HathiTrust will be fiscally very conservative, but legally not conservative because they bear none of the risk. My fear that they're going to become fiscally conservative and legally risk tolerant. That would be a mismatch. I talked about that with the board and I think were on a good place about it.
>
> If HathiTrust wants to see more access, more things opened up, the best thing HathiTrust can do is not try to get us to change some of our decisions, but to put serious funding behind the copyright review process; to create more access in a way that we can go defend in court and help the war."

As described in previous chapters, the University of Michigan has done groundbreaking work in the area of copyright rights determination — a potentially useful, and also incredibly expensive and time-consuming form of non-consumptive research that, in view of its principal investigators, would be functionally impossible without access to a resource like HathiTrust. That said, rights determination was not one of the proposals submitted and voted on by HathiTrust's members and so it appears unlikely, for the time being, that funds will be dedicated to that endeavor as the University Librarian at Michigan recommends.

Implementation of Initiatives

As tensions between legal and fiscal authority, between sustainability and growth, between ideologies, approaches, leadership, and governance continue to prompt readjustments and reconfigurations for HathiTrust and Michigan, implementation on a number of initiatives passed at the Constitutional Convention has begun to move forward. In addition to a proposal for the establishment of a governance structure (the implementation of which has already been discussed), recall that the initiatives passed include: (1) revising the HathiTrust's fee model; (2) creating a distributed print monographs archive; (3) creating a government documents archive; (4) creating a transparent process for inviting, evaluating, ranking, and launching new initiatives; and (5) broadening HathiTrust's mission and goals. This section traces the implementation of each initiative in turn.

Fee Model

Currently, membership in HathiTrust is limited to academic and research institutions. When an institution reaches out to HathiTrust about becoming a member, HathiTrust asks for a list of their OCLC numbers pulled from their institutional records which is then compared against HathiTrust's holding. As one of my participants acknowledged:

> "There are millions of small calculations that go into determining how many partners hold a given volume, in order to determine what the final cost to an institution will be. The more partners join, the lower the costs are for everyone."

The current fee model, passed in 2013, uses information about the overlap to apportion costs as follows. One portion of the costs is based on the number of public domain or open access materials and the number of HathiTrust partners. All HathiTrust partners share the cost of those materials equally. HathiTrust has come up with a number that essentially bundles up all of

its organizational operations and overhead costs for staffing, storage, web server infrastructure, etc., into a per volume cost. At last check, this cost amounted to approximately $0.15 per volume per year. HathiTrust currently has just over 100 partners, which means that any given partner would pay about $0.15 divided by 100 per public domain or open access volume per year. That's one portion of the costs.

The other portion of the costs is based on the overlap between an institution's in-copyright holdings and HathiTrust's holdings. As previously described, in-copyright works are not available to read or download in HathiTrust; nevertheless, the volumes add value for all members because they're full-text searchable. That said, the institutions that really gain value from the in-copyright volumes are institutions that hold those volumes in their print collections because HathiTrust is essentially providing a preservation backup of the volume. In addition, these institutions are able, in some cases, to make lawful uses of the volumes and so, under the fee model, they are treated as deriving more full benefit than the general membership. So for those volumes, HathiTrust divides the per volume cost among the institutions that hold a print copy. Thus, if five institutions hold a print copy of a particular in-copyright work, each one would pay three cents per year for that volume (the standard $0.15 per volume cost divided by 5).

In terms of HathiTrust's membership, the overwhelming majority are large, academic research institutions from the U.S. Membership appears to be growing amongst smaller libraries, especially liberal arts colleges, which are drawn to HathiTrust primarily as a preservation strategy but also benefit from full-text search and enhanced access for print-disabled patrons. In terms of copyright law and access restrictions, the path is somewhat simpler for U.S. institutions who do not have to contend with a separate, potentially conflicting, set of regulations from their home country.

HathiTrust's fee model has been a major stumbling block for at least one major ARL, the University of Toronto, whose OCLC records "need to be in better shape." As it currently stands, the University of Toronto is unable to provide HathiTrust with the full record list which, in turn, makes HathiTrust unable to compare holdings and develop a fee estimate. As a result, University of Toronto is unable to join HathiTrust, although the Chief University Librarian has discussed the possibility of University of Toronto becoming a new HathiTrust node, perhaps called HathiTrust North.

With respect to the long now of HathiTrust, there is a strong sense among many participants that HathiTrust needs to consider new forms of membership and new types of members. The Program Steering Committee has been charged with exploring new fee models that might attract international members:

> "Right now we have a single payment model, a single way to buy into the HathiTrust and then there you are. I don't know exactly why more institutions outside of the United States haven't joined but it may be because of our current value calculation."

Another possibility being contemplated is to permit members to join without contributing anything to the corpus:

> Another way of expanding the partnership model is through offering some kind of service model partnership, maybe not in preserving the content, but for receiving services. The community will change over time and there may be different ways that institutions see value in HathiTrust in different services that we can provide.

Relatedly, some participants discussed developing membership models that expand eligibility beyond academic and research institutions:

> "Perhaps an individual might one day be able to join HathiTrust. There may be people who may just want to download content from us; they may want that ability. They may want the ability to deposit content in HathiTrust without necessarily becoming partners. They may have a different fee model that they're interested in. I think there a lot of possibilities right now and it's a matter of just weighing, what's on our core mission? Who is the community that we're serving and how can we serve them most effectively?"

These are all considerations for HathiTrust's governance as it works toward sustainable evolution into the future.

Print Management

The distributed print monographs archive was proposed by HathiTrust's Executive Committee and was, for HathiTrust's lead architect, one of the primary motivations for creating HathiTrust. HathiTrust represented, in many ways, his efforts to leverage the digital corpus generated primarily by the mass digitization project for "collective action around the collective collection." In particular, he believes that HathiTrust has the potential to revolutionize shared print management.

It turns out that a very small percentage, about 3%, of an academic research library's print holdings are actually used. The other 97% is essentially stored in the libraries stacks or in

off-site storage facilities. This reflects a costly, inefficient and, in the view of HathiTrust's progenitor, unnecessary reality. While he agrees that "libraries do need to store this material, we don't need to have twenty or thirty copies stored in big refrigerators all over the country."

In his view, the HathiTrust's print management project is fundamentally about:

> "libraries gaining efficiencies in storage and preservation of services while offering better access to those materials and allowing them to do even better things in the new digital world. HathiTrust really grows out of the recognition of the great power of digital access to materials over great distances and the recognition of digital formats as a valid preservation strategy. HathiTrust has the potential to revolutionize shared print management. We don't have a firm grasp on what the print record is but if we use HathiTrust right, if we get this print book storage effort moving forward, we will begin to understand our print collections in a much more coordinated way and manage it more effectively. We can have fewer copies, have the right number of copies, not hundreds of copies of commonly unused works."

One library director opined that print books were going to become "the albatross around the necks of libraries — all cost, no value."

After planning and reporting phases reached completion, the program shifted into its implementation phase in 2015.[450]

The distributed print monographs archive also has potential implications for the long now of HathiTrust. Although the risk tolerance is not there for Michigan, some members of the library community within and beyond HathiTrust have suggested that fair use might enable digital access to in-copyright works where the institution holds a print copy in storage. Particularly if the number of readers of the digital copy was limited to the number of print copies held in storage, some participants speculate that fair use might support this activity:

> "I'd like to think that format shifting was a fair use, and that you could take analog material and convert it into digital form and then do everything with the digital work that you could do with the analog work. But, I don't know if that's the case or not. I haven't run through the analysis."

Indeed the Internet Archive has already begun doing this with its Lending Library program. Kahle has purchased (or received donations of) over 16,000 contemporary works, digitized them, stored them, and now lends them to the public through its website.[451] Whether

[450] https://www.hathitrust.org/print_monograph_archiving
[451] https://archive.org/details/lendinglibrary

the Lending Library will prompt a copyright infringement lawsuit giving rise to a fair use determination, we can only speculate as to the legality of such an arrangement.

Government Documents

The government documents proposal is well underway. In particular, significant progress has been made on the creation of a metadata registry of U.S. Federal Government Documents "including documents produced at the government's expense, in all formats, at the item level, from 1789 to the present." It may also include grant-funded or contract work, declassified materials, pieces of legislation, administrative publications, and/or data sets.[452] As previously mentioned, disagreements arose over this project with respect to potential copyright implications. According to my participant, the University of Michigan's General Counsel's office advised against the project due to the potential copyright risks involved, but the participant decided to move forward with it after conducting a review of relevant legal rules and literature. Under §105 of the Copyright Act, works prepared by an officer or employee of the federal government within the scope of employment are not entitled to protection under U.S. law. While this seems fairly straightforward, like many things in the copyright universe, boundary lines get blurry pretty fast.

In terms of the long now of HathiTrust, some participants suggest that the government documents project might signal that HathiTrust's future could become, in some ways, integrated with national infrastructure.

Initiatives Review Process

The purpose of this proposal was to initiative the development of a transparent process for inviting, ranking, evaluating, launching and assessing development initiatives from HathiTrust partner institutions. Unlike some of the earlier proposals, this one did not result in a stand-alone project but was rather folded into aspects of governance (with the working groups and committees, for example). As described earlier, the general sense of participants in the study was that the Constitutional Convention provided members with enough to do that there has not been an opportunity or need for the review of additional initiatives, yet.

[452] https://www.hathitrust.org/usgovdocs_registry

MISSION AND GOALS

The proposal regarding broadening HathiTrust's mission and goals was referred to the Board of Governors. HathiTrust's mission is to contribute to the common good by collecting, organizing, preserving, communicating, and sharing the record of human knowledge. Its website lists a number of specific goals including:

- To build a reliable and increasingly comprehensive co-owned and co-managed digital archive of library materials converted from the print collections of the member institutions.
- To dramatically improve access to these materials in ways that, first and foremost, meet the needs of the co-owning institutions, with a particular emphasis on ensuring access for individuals with print disabilities.
- To develop cost-effective and robust infrastructure for digital content of value to scholars and researchers, including a variety of formats and born-digital materials.
- To develop partnerships and services that ensure preservation of the materials in HathiTrust and the entire print and digital scholarly record.
- To reduce long-term capital and operating costs of storage and care of print collections through redoubled efforts to coordinate shared storage strategies among libraries.
- To build infrastructure that facilitates cost-effective and productive collaborations among partnering institutions to reduce the cost of securing campus intellectual assets.
- To define and make available a set of services that supports research using the HathiTrust corpus.
- To create a technical framework that allows for both central and distributed creation of tools and services.
- To sustain the HathiTrust enterprise as a "public good" while at the same time defining a set of services that benefits member institutions.

One of the standard questions I asked of all my participants dealt with whether the mission and goals of HathiTrust was well-understood. I got a somewhat surprising range of responses. There were participants who believe, in the academic library community, that "we all understand HathiTrust and stand behind it and are proud of it." Another said, "HathiTrust has achieved broad acceptance even though it's expensive to join." One even went all the way back to HathiTrust's progenitors saying:

> "Michigan and Indiana IT are mighty. It's just fantastic to store things correctly for long-term digital preservation and create new applications and new kinds of search mechanisms. People respect that and know that it's done just about the best way it could possibly be done anywhere."

On the other hand, as you move beyond the academic research library community, HathiTrust's resonance fades fairly quickly:

> "I don't think that the goals and nature of HathiTrust are probably well understood in the broader community. When I talk about the HathiTrust I look around the room to see how much it resonates with the group. I would say that it resonates much more than it used to but could somebody define it? Probably not. I think where we have a brand with people, it's very strong and I think that it is of values brand. Is it a brand that tells me what it is exactly? Probably not."

Another participant said:

> "Within the larger academic community the most prevalent view is to not have ever heard of it at all. It's a challenge that HathiTrust visions and objectives are not well understood by a broad community. I think what we need to do is mobilize the HathiTrust membership to create more awareness of HathiTrust in more positive ways."

Participants were mostly in agreement that, beyond the library community, HathiTrust does not have a particularly strong brand. In terms of what to do about that, opinions varied. Some, like the participant above, thought that HathiTrust should do more education and outreach, communicating its mission, goals, and services to a broader constituency. Others, like the participant below, worried that additional attention and interest in HathiTrust might overwhelm the fledgling organization (still buried under the weight of the Constitutional Convention initiatives):

> "Normal people don't really know what HathiTrust is. It's kind of opaque. I'd like to do a lot more PR and outreach but I don't think we have the resources to both be out there and respond to the kind of possibilities that I think would come from that."

One possible approach might be to do outreach that is oriented around that expansion of particular services. For example, given the Court's ruling on the issue of access for print-disabled patrons, some HathiTrust members have expressed interest in working to expand that access beyond HathiTrust's members:

> With respect to users who have print disabilities, we're actively working to expand access outside of HathiTrust members for users with print disabilities. I

> don't know what a service model might look like to sustain not work if one would even be needed, but that's something that's on the horizon.

This is another example of the shift in perspective from the highly dynamic, reactive, action-oriented, and immediate early days of HathiTrust, to the more structured, deliberative, long now of HathiTrust.

The Long Now of HathiTrust

This section concludes *this* story of HathiTrust. We began in the dark, somewhat hidden, catacombs of early large-scale digitization efforts. We were catapulted into the digital future by the Google Library Project and the whirlwind of mass digitization's great unknowns. We navigated the terrain of inter-institutional team-building and decision-making which was marked by tremendous acts of courage, trust, and a few landmines. We witnessed the tremendous growth of HathiTrust through its intensive incubation and development period. We observed its rite of passage from a small, nimble, informal organization to a semi-autonomous, collectively governed organization with over 100 partners. We watched with cautious optimism as it fought for fair use and returned home victorious. We surveyed the aftermath, as it rejoices and recoils, shaping and becoming shaped by its broader sociotechnical environment.

Now we turn our gaze toward the future. Using data generated from my interviews, this section distills some of the key themes, concerns, hopes, fear, and dreams of participants as they speculate about the long now of HathiTrust.

When asked to hypothesize and contemplate the future and HathiTrust's role in it, many respondents returned to a focus on technology. Technological change has always been a significant propellant of sociotechnical change. "Technological change changes the context of decisions," one of my participants said. The future will depend, in large part, on how technologies continue to evolve. Another participant reflected, "[t]he mass digitization project was the starting place to whatever the endgame was going to be. We are still nowhere near the endgame."

It may be true that we are nowhere near the endgame, but one thing that the mass digitization project and subsequent emergence and evolution of HathiTrust demonstrated is the academic research communities' ability to engage in effective collective action. As one of my

participants remarked: “One of the breakthroughs of HathiTrust is that we've proven to ourselves that we can aggregate scale to mutual benefit and outcome in the digital era.”

In terms of how technologies might impact HathiTrust’s continued evolution, participants agree that a number of project developments might be in its near future. One project focuses on continuing to push forward and innovate in the area of accessibility technologies. As one participant stressed:

> “This is where the quality of the scan does impact access, because the adaptive technology may not permit the print disabled person from accessing the content if the quality of the scan is poor.”

In addition, building better tools and systems for digitizing, preserving, and organizing non-text items is thought to part of our near future. The HathiTrust Collections Committee has begun exploring the possibilities of pushing forward into image and audio file formats, but the ultimate role HathiTrust will play in that transition is still up for debate.

More broadly speaking, participants talked about how technological change will continue to have important implications for emerging social practices implicating copyright law. As one participant said,

> “I think that it will be decided by society that in order to maintain a balance of rights and access, some things are going to have to be changed and the thing that will drive that, somehow or another, is technology.”

The intersections of copyright, technology, and social practice will continue co-evolve. Whether this process of mutual-readjustment will ultimately lead to greater access and openness or tighter restrictions was a question that weighed heavily on the minds of my participants. One of the key decision-makers involved in the UM-Google project reflected:

> “My hope from the beginning was that, in time, the existence of this corpus of work would cause various actors in the world to want to figure out a way to use it and to use it well. And that sort of pressure, especially from the youth, would eventually lead to arrangements and outcomes that would make the work readable, as well as searchable and metadata rich, and that’s still my hope.”

Another participant said:

> “I truly believe that open is the future. Do we get there in one fell swoop? No. But we're pushing the envelope as much as we can through things like HathiTrust and so I think that the naysayers are the ones that are trying to hold back the ocean.”

Other participants were somewhat more cautious in their assessment of the future:

> Is everything going to be locked up and corporatized? Is it all going to be Apple iPhones all the way down with closed apps closed operating systems? Or are we actually going to have a World Wide Web-style future? I guess I'm a little bipolar on that. Sometimes I think we're going to win and sometimes I think we are going to lose. Looking back, I think it's going to be the seeds of things to come. We'll either be locked up in this sort of: 'Open was a nice idea by a bunch of idealists but they don't understand the rigor of what we're talking about in MBA school.' That's one version of the future. Or there's the open competition and open system and method that allowed newcomers to come in and innovate. That's the winning system and we knew that all along. And I can't tell you which of those histories is going to be the one that's written. I can tell you which one I'm throwing everything I've got into trying to make happen."

What does it mean to throw everything we've going into trying to make the "Open" version of the future the one that succeeds? For many participants it means taking on more of an advocacy position and working collaboratively with other stakeholders and constituencies to advance the interest, values, and goals of HathiTrust. Advocacy happens on multiple fronts. One front is social. HathiTrust is collaborating with other digital libraries and preservation networks:

> "The landscape is evolving right now and we are beginning to form a community that is thinking about the different functions of the library and how to fulfill them and specializing within the community, recognizing that some people or institutions have certain aptitudes for doing certain things really well. For example, Europeana and the Digital Public Library of America are very much into aggregation and the Digital Preservation Network is an underlying preservation network. Right now HathiTrust is in the middle of those. We have a preservation mission and we also have a deeply ingrained access mission. If we do well as a community, and we succeed, it will be because we're performing these functions collaboratively."

There is a recognition that, for HathiTrust to succeed in its mission, collaboration and advocacy will need to expand beyond the library universe. "As long as it's seen as a library thing, a library responsibility, it's not really going to get off the ground." Another participant said:

> "I just don't think it's realistic that HathiTrust can continue maintaining the status quo, doing exactly what it's doing now, and remain viable and valuable for users. If our users continue to just be the librarians, were okay. But I think we would want, in the future, for any citizen in the United States or abroad to be able to leverage this information. And right now the only audience is librarians talking to librarians."

Another participant said:

> "I hope that in the future more people other than librarians are going to care about this. A part of me thinks that HathiTrust and libraries are not important enough to advance these problems. I think that if Google or Amazon said, 'You know what? Public domain books — that's where it's at! We have to invest money in that. That's the thing!" I mean, we wouldn't have HathiTrust if it weren't for Google so part of me is a little cynical about whether or not the research library community has enough moxie and power to really articulate these issues to the average citizen, or even the average faculty member, to make a difference and get people to agree that this is something worth investing money in, that people need to stop giving up their rights, that we need to make these things open."

There was a shared sense amongst participants that so long as issues around open access and open systems and fair use were treated as "library issues," initiatives were going to be of limited success. Figuring out ways to make these issues resonate with a wider audience is going to be a key part of advocacy moving forward.

Another important, and somewhat more contentious, front for advocacy around these issues is on the law side. Some participants recommend engaging with the Copyright Office and legislators in an effort to reform the Copyright Act:

> "Responsible government should revise the Copyright Act in the next 20 years because of the technological changes that are continuing to happen. What reforms should there be? Any reforms that make it perfectly possible for us to do exactly what we want without any extra money being paid! Right now, based on the model of library practice that we've been pursuing for the last century, we have what we need. But if changes can be made that maintain protections for the copyright holder — especially authors — and permit libraries to provide full, legitimate, lawful access for educational purposes … that is my dream."

Others are not so convinced that the Copyright Act will be, or should be, revised.

> It is a worry that if they open up the Copyright Act in this day and age, the results will be more limiting. Most copyright experts don't seem to have a lot of hope that the copyright regime is going to change for the better. For the copyright regime to change, Congress has to make a change. And none of them would say that that's going to happen in their lifetime."

As it arguably did in *Authors Guild v. HathiTrust*, additional acts of innovative deviance may be the best approach to teaching Copyright law to change along with the changing sociotechnical environment. Programs such as providing digital access to out-of-print works that

are held in storage, in hard copy, at the institution may be a reasonable next step forward in the fight to advance fair use. As one participant noted:

> "The problem is always, of course, with copyright and the real big change will come if and when we can actually read these books. That's why I was so disappointed that the Google settlement eventually collapsed, because that was the only avenue that was gonna be able to allow us to read most of the books published in the twentieth century."

The Internet Archive's Lending Library is already taking an aggressive approach to making contemporary works accessible by offering digital access to in-print works without tying digital access to physical access. Kahle recalled a conversation he had with Michael Lesk:

> "Michael Lesk, who I think of as the father digital libraries, said I'm worried about the twentieth century — no problem with the twenty-first century it looks like we're in good shape — but it looks like the twentieth century might be locked up."

The Lending Library takes a step toward unlocking the twentieth century. However, few participants were convinced with Lesk's assessment that the twenty-first is in good shape particularly because of trends of publishing moving increasingly toward a licensing regime.

> "Another legal regime, based on licensing or something else, is going to pop up dominate the landscape in terms of the way access to these materials is handled. We are just at the beginning of this era of license and that is going to continue to move forward."

> "Initial first publications are increasingly being published in a digital format and we're basically going along operating as if e-books have the permanence of a printed book on our shelves. But there is a big difference. If something bad happens to the computing infrastructure of our nation there would be a blank on our shelves of everything starting about 2007 moving forward. Many of the typical academic library publications that we buy from vendors will not be there."

Another participant notes:

> "I think we're under tremendous risk to not have long-term access to works that are being created now and in the future that are licensed rather than sold to libraries. I think it's overly optimistic to view the twenty-first century as problem-free. Works that are currently being commercially exploited, and published in electronic form, they will be okay for so long as their commercially exploited, and then they're likely to turn into digital dust without anybody noticing. So I think the twenty-first century is deeply fraught."

So far, HathiTrust has not addressed the challenges inherent in the shift from a copyright to a licensing regime. HathiTrust is a fundamentally about retrospective conversion and the challenges of continuing to bring in new materials, particular born digital or licensed materials, and grow the collection is not happening. The challenges posed by the transition to licensing may prevent HathiTrust from continuing to meet the needs of the community. As one participant said:

> "It's really the community that's driving things and our success right now completely depends on being aligned with the needs and goals of our members. To the extent that we were able to stay on target with that, stay in touch with our members and leverage the abilities of our members, we'll do well. If we don't do that, I think we'll shrivel pretty quickly. I think there is the ability for it to do even more. I think as long as University of Michigan exists, HathiTrust will exist. Michigan will not let HathiTrust fail."

We can notice an interesting tension hidden in the above quote. On the one hand, the participant recognizes the continuing relationship between Michigan and HathiTrust as a source of strength and security. On the other hand, they recognize that the community is driving things and it is HathiTrust's ability to serve the goals of its members that will determine its success or failure. Finding space to experiment, grow, and innovate in the current, relatively more conservative, sustainability-focused environment at Michigan will be one of HathiTrust's main challenges moving forward.

One of my participants made an interesting observation related to this point:

> "One of the revolutionary aspects of HathiTrust is that it models the future. The revolution here is in the community coming together around a very important value and concept for libraries and then figuring out how we leverage this use it to the best and greatest extent for our user community for our scholars and really push the boundaries of fair use as much as we can and be willing to do that. Technology is going to force us to really think differently and the more we can do that as a community, the better."

There's almost an inherent tradeoff between sustainability and generativity. We can focus on finding ways to ensure that HathiTrust remains relevant, functional, credible, and sustainable in the future, treating the future as some relatively stable uncertainty. Or we can focus on finding ways to create the future we want to inhabit; this approach necessarily involves more risks and more opportunities for failure and success. There is not one correct approach, but it is very

difficult to attempt both approaches simultaneously. One of my participants who runs the IT department at a major library had the following reflection:

> "Librarians come from a fairly long and well-established book culture, where there's all kinds of structure and format around how you encode your knowledge and how you describe your knowledge, and how you accredit your knowledge, and how you distribute your knowledge. And now we're in this environment that is fundamentally different in ways that we do not yet understand. We're still figuring it out. But we have this mindset that's very, very deeply ingrained in academia and in libraries. And so the place where I tend to fret a little bit in thinking about the future, is preserving enough 'figuring it out space' while we're still in a place of very rapid transition. I want us to resist the urge to encode too carefully what we're doing and how we do it because we also need to keep figuring out what it is that we're playing with. And really, this is young. It's 20 years old right now. 25 years old maybe. And I worry that you expect everything to walk like a duck and quack like a duck because ducks is what you deal with. But if an ostrich wanders onto the scene, you're going to make it look like a duck because you deal in ducks. We need to retain enough space for the technology team to think creatively about what this means so that they don't simply make us more ducks."

Finding that balance between safety and innovation is going to a key challenge moving forward. The participant just quoted is a technologist and brings that perspective to the issue, but as this work demonstrates, technology is one strand of a much larger, incredibly complicated, sociotechnical knot. If HathiTrust's past is any indication, it will be the organizational and institutional relationships that play the most central role in its continuing evolution.

Perhaps it makes sense to conclude this story of HathiTrust with an expression of its earliest progenitor's hopes and fears about the long now of HathiTrust:

> "If HathiTrust succeeds, I think it will succeed because we recognize in it the potential to do the things libraries needs to do and that libraries will leverage the scale of HathiTrust to be better at stewarding the cultural record and devote more resources to engagement on their campuses.
>
> If HathiTrust fails, it will fail because our collectivity limited the opportunity for expansive interpretations, for brave actions, the kinds of things that happened with the Authors Guild lawsuit that got us to her we are now. Now that there's this collectivity there is a danger that good and bold things will not be done because we would not be able to act with the same kind of courage as a collective and that would drive us to narrower and narrower positions. And I don't think that shift in thinking is an inevitability. These things are choices. A successful enterprise is going to make brave decisions no matter how big it gets and how much shared governance you have. You can make decisions to push the organization in that direction. You can push them to timidity or you can push

them to bravery and I think that this is the big challenge that HathiTrust's new Executive Director is going to have. He's got to encourage them to be brave."

Chapter X: Conclusion

This research traced the emergence and evolution of HathiTrust. The story of HathiTrust provides a window through which we can describe and explain some of the complex, often murky, intersections among technological change, emerging social practices, and copyright law as part of a larger reflection of processes of sociotechnical transformation. This research asked questions including: How and why does a relatively risk-averse academic institution decide to undertake an incredibly risky new initiative? How and why does technological change influence decision-making around risks and opportunities? How do decision-makers make sense of, explain, and justify those decisions, actions, and tradeoffs? How do new technical, organizational, and legal forms emerge and evolve over time to support the social transformations? How do emerging forms of practice and design reflect and respond to internal and external factors, tensions, relationships, and stakeholders?

This research on HathiTrust attempts to provide answers to those questions and others by offering a richly detailed account of the entangled and mutually constitutive relationships among law, technology, and social practice. The pre-history of HathiTrust, which explored large-scale digitization efforts up to and including the mass digitization project, offers clues about how changes in technology change the context of decisions. Innovations in technology alter the possibility space of our activity as well as the ways in which we identify, evaluate, and make sense of opportunities and risks, and the ways in which other stakeholders in our environment interpret, react to, and in some cases adjudicate our choices. Technology may emerge and evolve in a relatively more fast-paced and dynamic way, but it is neither determinative nor always the primary force or source of sociotechnical transformation. New spaces for social practice and engagement are made possible through changes in technology and, through processes of invention, use, modification, and regulation we come to recognize technologies are also socially constructed. Transformation in the context of this research is shown to be characteristically dynamic, non-linear, multi-directional, and guided by careful deliberation and planning as well as unanticipated, accidental, and/or sometimes controversial choices and actions.

Technological change is a useful jumping off point, but very quickly we begin to recognize the ways that technologies are only a small piece of a much larger and more complicated system that also includes, in this instance, legal institutions like copyright law, and academic institutions that have deeply encoded practices and shared traditions. As technology, social practice, and law interact and constantly readjust to each over time new sociotechnical forms emerge, like HathiTrust, that provide signposts to broader still processes of sociotechnical transformation.

In addition to providing a telling of HathiTrust's origination and early evolution, this research makes a number of other contributions which will be briefly summarized in this final chapter. Its first, and perhaps most potentially significant, contribution is the development of a new analytic framework for describing, explaining, and understanding processes of sociotechnical transformation based on the concept of innovative deviance.

Innovative Deviance and Sociotechnical Transformation

The innovative deviance framework provides important clues into some of the ways that social and legal structures may influence, prompt, and channel emerging behavior, particularly in environments destabilized by rapid change. A mismatch between culturally accepted goals and the institutionalized means to accomplish the goals can generate a sense of disequilibrium or strain within a society or community that pushes people toward deviance. Deviance can follow a number of different paths but one path in particular, innovative deviance, merits special attention. The reason that innovative deviance differs from the other paths of deviance is because it is the only one that promotes the accepted goal, although it rejects the institutionalized means to accomplish the goal and/or legitimate means are unavailable for accomplishing the goal. Innovative deviance is the creative use of illegitimate means to accomplish a desired legitimate end. Innovative deviance therefore represents a tension between our desire to accomplish the goals to which we all aspire, and the sense that we should all play by the same rules, and that the rules are a fixed point. Innovative deviance is a method by which institutions, social structures, law, and other slow-moving forms learn to change.

This research expands upon the innovative deviance framework articulated by Robert Merton and others working in the structuralist vein in two ways. The first is by applying it to a new context, namely copyright law and emerging research and scholarly communication

practices. The innovative deviance framework helps us understand how and why transformations occur at a grand scale, influencing collaborative computing and knowledge sharing at the level of institutions and infrastructures.

Secondly, and perhaps more importantly, this research expands on the basic framework advanced by Merton and others by supplementing the structuralist/functionalist approach with a process-oriented approach informed most meaningfully by Weick and colleagues' research on sensemaking. Sensemaking and decision-making theories provide insights into how change and transformation is not simply a reaction to structured imbalances in society, but also coalesces and percolates up from individual and organizational relationships and interactions. When we shift our perspective from the bird's eye to the worm's eye, we can see that the on-the-ground realities of personal relationships and small-group dynamics play a big part in how the structural tensions are channeled and interpreted. Dispersed, emerging, mutually-constitutive sensemaking processes are also crucial; sensemaking shapes and is shaped by the entanglements of law, technology, and emerging social practice. An understanding of sociotechnical transformation calls for a combined top-down and bottom-up analytic approach.

There were three key instances in the story of HathiTrust where the innovative deviance framework provided a useful analytic lens. The first instance was Michigan's decision to join the Google Library Project. As described, lots of institutions including Michigan had been actively pursuing digitization on a large scale. There was a strong sense within academia, echoed in many other sectors as well, that "digital is the future." There was a drive to digitize but various technical, organizational, financial and (to a lesser extent) legal impediments hampered the implementation and follow-through on those early endeavors. Google's offer to digitize Michigan's collection in a matter of years and cover virtually all of the associated costs seemed, to those at Michigan, like a way to overcome many of the obstacles that had stood in the way of earlier digitization efforts. Unfortunately, copyright law, as it existed at the time, did not provide an obvious safe passage from the analog to the digital world. There were no institutionalized means within copyright law for digitizing millions of in-copyright books. Therefore, when Michigan decided to engage in the digitization and assume the copyright risks associated with this new behavior, it was essentially making creative use of illegitimate means to accomplish a desired legitimate end. There was the sense that Michigan's choice would force theories of fair use, theories that had not yet been borne out by practice or adjudicated by the courts. The mass

digitization project was a way of recalibrating the balance in copyright between its overriding goals and the institutionalized means articulated in the Copyright Act; it was a way of teaching copyright law to change and adjust to changes in technology and social practices.

The second key instance of innovative deviance discussed in this research was Michigan (and Indiana's) decision to move outside the CIC governance structure and forge a bi-lateral agreement to form the digital repository. All of the CIC institutions recognized the need for a shared digital repository, i.e. there was a shared goal, but the institutionalized means for accomplishing the goal were, in Michigan's view, far inadequate and restrictive. The governance structure and shared culture of many academic libraries favored slow-moving, fine grained, detail-oriented deliberation via egalitarian voting processes. Michigan and Indiana attempted to "reason" with the group but ultimately rejected that decision-making process in favor of their own bi-lateral agreement. Michigan and Indiana made creative use of illegitimate means to accomplish a desired legitimate end. As with the decision around the mass digitization project, the decision to create the shared digital repository was a way of teaching the CIC to change. The innovative deviance "bomb" was the thing that unstuck the shared digital repository and got the process moving forward again. As one of my study participants reflected: "HathiTrust is not going to change libraries. What it's going to do, it's going to allow libraries to change. That's the big one."

The third key instance of innovative deviance was reflected in the district court's sensemaking around the mass digitization project. I suggested that while fair use and transformative use doctrine and precedent provide an incredible source of protection and flexibility for new and emerging forms of practice and engagement with protected works using digital technologies, the legal mechanisms and tests employed by courts may not go far enough to protect and support innovative deviance. The district court judge seemed to recognize this when he concluded that the current uses of the digital corpus so clearly promote the overriding goals of copyright that we should not be particularly concerned about whether or not legal precedent exists to support a finding of fair use. In other words, he seemed to accept innovative deviance — creative use of illegitimate means to accomplish a desired legitimate end — as a legally valid exercise of one's rights under the Act. The Circuit Court disagreed, however, explaining that the behaviors are not analyzed on the basis of how well they accomplish the goals of copyright, but rather on whether or not they satisfy the test for fair use, i.e. the

institutionalized means. While the ultimate outcomes of the decisions were functionally the same — HathiTrust won — the difference in reasoning may have important, potentially negative, implications for future adjudications of innovative deviance.

The innovative deviance framework may help us also make sense of HathiTrust's continued evolution. For example, where the story left off, there were growing tensions emerging from the power dynamics between HathiTrust's membership and the University of Michigan. In particular, there were concerns over future initiatives and directions from HathiTrust, particularly where they raise potential copyright implications. The institutional character of Michigan has changed in significant ways in the last decade, becoming increasingly conservative and risk-averse as a consequence of the lawsuit and key changes in personnel and leadership. As HathiTrust's members are still working through the implementation of several of its first-round initiatives, these tensions have not come to a head. One could imagine, at some future date, that the dynamic between HathiTrust and Michigan will be in need of recalibration either through some act(s) of innovative deviance and/or through severance, i.e. HathiTrust becomes an independent legal entity.

In addition, the innovative deviance framework can help us make sense of other controversies emerging within and outside of the copyright world. For example, it could provide some helpful insights into describing, explaining, and understanding Aaron Swartz's copying of JSTOR's archive and the ensuing reaction from the affected institutions, the government, and the public media around the events. Similarly, the current controversies involving Sci-Hub, a website that hosts 40 million mostly unauthorized copies of copyrighted scientific works, might be partially explained through an application of the innovative deviance framework. Beyond copyright, the framework may help us understand how structured inequalities produce socially beneficial transgressions. We might look to controversies involving Edward Snowden and Wikileaks as possible departure points for future explorations.

Transformation and Transformative Use

In addition to contributing to understanding of the emergence and evolution of HathiTrust and processes of sociotechnical transformation through the innovative deviance framework, this research also contributes understandings of the relationship between transformation, broadly construed, and transformative use. Transformation in the copyright context is related to, but far

more restrictive than, the broader notions of sociotechnical transformation mentioned above. In the context of research and scholarly communication, transformation provides a particularly interesting point of comparison between the overlap, and lack of overlap, in meanings and understandings.

Transformation is critical to the scholarly endeavor. Researchers, teachers, and students routinely read, use, adapt, reconfigure, modify, synthesize between and among various resources as part of the learning, teaching, and communicating process. Copyright recognizes the importance of transformation in educational contexts, in large part through the fair use doctrine. In copyright law, transformation is made quantifiable through the application of elements, factors, comparisons, analogies, and tests. The *Authors Guild v. HathiTrust* case revealed some interesting insights into judicial sensemaking about transformation. But transformation in the scholarly communications context is far broader than its legal definitions. It refers to how *emerging technologies* are fundamentally altering the ways that scholarship is created, disseminated, used, preserved, and controlled. And transformation also refers to how *social practices* in the scholarly environment are changing at the institutional, organizational, disciplinary, and individual levels.

To the extent that fair use and transformative use in the copyright law context are overly focused on tests and precedent (as is the custom in the practice of law) it can sometimes lose sight of the forest for the trees. In other words, sensemaking around copyright law can lose sense of the broader purposes of fair use, transformative use, and the goals of copyright law more generally. This can become problematic when the courts are asked to rule on innovative deviance which involves behaviors that, by their very definition, promote important social goals but do so via illegitimate, unorthodox, and/or infringing means. If we are restricted to tests that privilege the test itself over the ultimate purpose of the determination, then we risk condemning emerging socially productive uses of protected works.

As I described in Chapter VIII, iterative, non-consumptive, technology-enabled uses of protected works do not fit neatly within the core traditions of fair use and transformative use. With things like search and data mining, the transformation we're seeking to protect through fair use isn't captured by the secondary work — displays of the search results or the instances of the word "dogs" in a given work. Rather, the transformation we are seeking to protect is this broader sense of transformation that I described earlier. Full-text search is transformative because it

fundamentally and beneficially changes scholarly communication practices; it makes new meaning, message, and understanding possible and in so doing promotes the overriding goals of copyright law. But it is not HathiTrust that is doing that work, it is the person using HathiTrust. To promote broad socially beneficial sociotechnical transformations, copyright law will need to offer robust protection for technology intermediaries and service providers. I suggested that the development of a contributory transformative use doctrine might be called for. In addition, a balancing test that provides judges with broad discretion to weigh the potential social benefits of a use (i.e. how well it promotes the overriding goals of copyright) against potential harms to copyright holders might also be a way of building greater flexibility into the copyright law system and ensuring more channels of support for innovative deviance. Finding ways to bridge sensemaking around transformation in the context of copyright law will be an important locale for future research.

Policy as Embedded Generativity

This work also contributes to understandings of the role of policy in processes of sociotechnical transformation. Section 4.4.2 of the UM-Google agreement teaches how, in addition to opening and closing spaces of technical design and social practice, policy can also function as a source of embedded generativity in sociotechnical systems, preserving open spaces in support of future transformations.

The MDP was groundbreaking in terms of its scale, speed, and innovative deviance -- its willingness to creatively push against the boundaries of existing copyright laws to accomplish the overriding goal of copyright. While large-scale library digitization efforts had been ongoing for at least a decade prior to the MDP, an endeavor like this had never before been undertaken; it was full of potential and rife with uncertainty. This research showed how policy and policymaking played a central role in defining and navigating the emerging technical and social environments of mass digitization. Perhaps by virtue of its scale, speed, and concomitant copyright risk, the mass digitization project generated an overarching web of policies around digitization that had been largely absent or uncoordinated in previous large-scale digitization efforts. As earlier described, those precursors tended to be one-off, relatively short-lived, resource-restricted endeavors with limited organizational infrastructure and support.

Policy played a significant role in the story of HathiTrust for at least four reasons. First, policy was functional. The agreements were designed and intended to mitigate uncertainty and provide a safety net against some of the potential risks (copyright and other) associated with digitization. Second, policy was important in terms of internal and external sensemaking around the mass digitization project and the co-construction of meaning. Third, policy, specifically clause §4.4.2 in the UM-Google agreement, can become a powerful value lever for sociotechnical transformation, empowering the University of Michigan to leverage the mass digitized corpus for collective action in the research library world around challenges and problems common to all of its members. Finally, this research shows how policy can function as a source of embedded generativity, enabling new, innovative, often unanticipated, potentially transformative changes to occur at the intersections of social practice and technical design.

Absent §4.4.2, it was doubtful that research libraries would have been legally permitted to share or collaborate around the digital corpus. We would have ended up with Google's range of services and up to forty siloed back-up copies (one for each participating library). While such digital siloes are not without value, it is the capacity of the libraries to cooperate and collaborate around this material, and build something by, of, and for the academic research world that was so potentially transformational. Without §4.4.2, there would have been no HathiTrust, and likely far fewer opportunities for collective action around the collective collection.

It can be tempting, in retrospect, to normalize the evolution of new sociotechnical forms. Hindsight bias can pose a significant obstacle to accurately tracing the creation and subsequent impacts of a policy such as §4.4.2. There is a strong tendency for people to understand HathiTrust as it currently exists and retrospectively make sense of its emergence and evolution as a rational, planned, fairly linear, perhaps even unavoidable progression from point A to point B. In reality, however, that could not be further from the truth.

When the UM-Google Agreement was entered into, and section §4.4.2 was drafted, Michigan was operating on the belief that the digitized corpus would be a dark archive – a digital back-up copy of the library with little to no provision for access, an understandable position to take given the state of copyright law as it was understood to exist at the time. This is not to suggest, however, that there was no inkling that the digitized corpus might someday, somehow, under a modified set of legal and organizational facts become something more than a dark archive. There was certainly a utopian vision of creating a universal digital library underlying

§4.4.2, but it was deeply buried and very much inchoate when the agreement was made. No one, not even the architect of the provision, knew the extent of its significance and meaning in advance. It was, at that point, a matter of embedded generativity, safeguarding the possibility of future transformation within the academic research library world by preserving open spaces for collaboration and new, innovative, and unanticipated uses of what was assuredly an unparalleled digital resource.

It took time for that generative potential to take shape and evolve. Today, the values and generativity embedded in §4.4.2 are finding new outlets. Projects like the HathiTrust Research Center are beginning to providing academic researchers with support and assistance in conducting non-consumptive computational research including data mining and other forms of digital humanities work. Libraries are also using the corpus to engage in cooperative non-consumptive research for their own ends. For example, under the leadership of the University of Michigan Library and with funding support in the form of a National Leadership Grant from the Institute of Museum and Library Services, libraries have used the HathiTrust corpus to research key dates in the lifecycle of copyrighted works (publication date and location, author information and death dates if applicable, GATT restoration, and other complex considerations) to build evidence for opening works to the public domain or setting trigger dates for their future release. This copyright review management system would have been functionally impossible without the MDP and §4.4.2. Additional nodes of expansion are being discussed to support new and innovative uses of the corpus. Some HathiTrust members are actively seeking ways to expand beyond text-based services to include audiovisual nodes. These are just a few of the changes currently being discussed and developed by HathiTrust's membership and each of these new potentials reflects the values and generativity embedded in §4.4.2.

While §4.4.2 opened space for new, often unanticipated, forms of innovation and generativity, it is important to note that the clause also foreclosed some opportunities. The Internet Archive, for example, has been largely excluded from participation in HathiTrust because of §4.4.2 and other aspects of the UM-Google Agreement. In addition, the Agreement has fairly rigid restrictions about sharing and collaboration even with respect to the public domain materials. Recognition that policies like §4.4.2 can open and close spaces of social practice and technical design must also accept that even these barriers are permeable and, with sufficient technical skill and social support, may be prone to dissolution.

The story of HathiTrust thus provides a localized, granular example of how policy, design, and practice are entangled, dynamic, and mutually constitutive. While policy, particularly §4.4.2, had a tremendous impact on library mass digitization and the subsequent emergence and evolution of HathiTrust, it was value-driven and generative rather than determinative. §4.4.2 did not cause, create, or produce the HathiTrust we know and understand today. Rather, it embedded within the Project a source of generativity capable of opening (and closing) spaces of social practice and technical design in new, innovative, often unanticipated, and potentially transformative ways.

This research also suggests that an increased focus on policy can help bridge "the gap between what we need to do socially and what we can do technically" that has been a primary concern and challenge for researchers studying sociotechnical systems.[453] Policy can provide a flexible-yet-binding link between relatively rigid technical forms and relatively fluid and unpredictable social practices. Policy can simultaneously enable rigorous action by reducing uncertainty around risk and also imbue relationships with a degree of nuance and functional ambiguity that supports adaptability and change.

In addition, this research demonstrates how policy can stabilize and coalesce future sensemaking and decision-making around emergent sociotechnical forms, ultimately tightening the design-practice-policy knot.[454] The UM-Google Agreement provided important clues and signals about individual and organizational sensemaking and decision-making around the mass digitization project. Policy provides a window into what the parties believed was important, what concerns (or lack of concerns) they might have had, and what sorts of justifications might underlie those beliefs and decisions. In addition, when the UM-Google Agreement became public, its power became demonstrable; the policy altered the ways in which other institutions negotiated, enacted, and engaged in digitization. The irreversible and visible commitment between UM and Google channeled future sensemaking and decision-making around the mass digitization project, tightening these processes and making them more predictable, orderly, organized, and self-reaffirming. The policy became more than a simple artifact of a binding

[453] Ackerman, M. S. (2000). The intellectual challenge of CSCW: the gap between social requirements and technical feasibility. *Human-computer interaction*, *15*(2), 179-203.
[454] See e.g. Jackson, S. J., Gillespie, T., & Payette, S. (2014, February). The policy knot: Re-integrating policy, practice and design in Cscw studies of social computing. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 588-602). ACM.

commitment between UM and Google. It became a reserve for leveraging and generating collective action.

Thus, although policy is often conceptualized as a mechanism for channeling and/or regulating social practice and technological design choices, this research suggests that policy can also function as an important safeguard for the emergence of new, often anticipated innovations and transformations in technical and social spheres. This is particularly important where large-scale collaborations at the level of institutions and infrastructure are concerned. It takes time to explore and co-construct meanings around a massive new digital resource like the digital corpus resulting from the mass digitization project. This understanding contributes to future decision-making in emerging sociotechnical environments.

Conclusion

This research into the emergence and evolution of HathiTrust facilitates deeper understanding of how institutions and organizations shape and are shaped by the interactions among copyright, technological change, and emerging social practices. In addition, the analytic framework that expands and extends innovative deviance in combination with sensemaking approaches provides meaning and insights into the complex, often murky, processes of sociotechnical transformation.

References

Ackerman, M. S. (2000). The intellectual challenge of CSCW: the gap between social requirements and technical feasibility. *Human-computer interaction*, 15(2), 179-203.

Atwood, M. (2011). *The Publishing Pie: An Author's View,* presentation at O'Reilly Media Tools of Change for Publishing conference, Fed. 15, 2011, https://www.youtube.com/watch?v=-6iMBf6Ddjk

Authors Guild (The) (2015). "The Authors Guild's Top Legislative Priorities – 2015," available at https://www.authorsguild.org/wp-content/uploads/2015/05/2015-AG-Top-Legislative-Priorities.3.23.15.pdf

Bartunek, J. M., & Moch, M. K. (1987). First-order, second-order, and third-order change and organization development interventions: A cognitive approach. *The Journal of Applied Behavioral Science*, *23*(4), 483-500.

Baxter, P., & Jack, S. (2008). Qualitative case study methodology: Study design and implementation for novice researchers. *The qualitative report*, *13*(4), 544-559.

Becker, G. S. (2009). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press.

Benkler, Y. (2002). Coase's Penguin, or, Linux and" The Nature of the Firm". *Yale Law Journal*, 369-446.

Bianco, M. W. (1926). *The Velveteen Rabbit*. Pioneer Drama Service, Inc..

Bijker, W. E., Hughes, T. P., Pinch, T., & Douglas, D. G. (2012). *The social construction of technological systems: New directions in the sociology and history of technology*. MIT press.

Borgman, C. L., Wallis, J. C., Mayernik, M. S., & Pepe, A. (2007, June). Drowning in data: digital library architecture to support scientific use of embedded sensor networks. In *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries* (pp. 269-277). ACM.

Bowker, G. C., Edwards, P. N., Jackson, S. J., & Knobel, C. P. (2010). 1.1 The Long Now of Cyberinfrastructure. *World Wide Research: Reshaping the Sciences and Humanities*, 40.

Brand, S. (1999). The Clock of the Long Now. *London: VVeidenfeld and Nicolson*

Brickman, P., Abbey, A., & Halman, J. L. (1987). *Commitment, conflict, and caring*.

Castells, M. (2011). *The rise of the network society: The information age: Economy, society, and culture* (Vol. 1): John Wiley & Sons

Castells, M. (2011). *The power of identity: The information age: Economy, society, and culture* (Vol. 2): John Wiley & Sons.

Castells, M. (2010). *End of Millennium: The Information Age: Economy, Society, and Culture* (Vol. 3): John Wiley & Sons.

Centivany, A. (2015). Innovative Deviance: A Theoretical Framework Emerging at the Intersection of Copyright Law and Technological Change.

Centivany, A., & Jackson, S. (2012, February). Concentration, incumbency, and interconnection: broadband development and the lessons of history. In *Proceedings of the 2012 iConference* (pp. 317-324). ACM.

Christenson, C. (1997). The innovator's dilemma. *Harvard Business School Press, Cambridge, Mass*.

Cloward, R. A. (1959). Illegitimate means, anomie, and deviant behavior. *American sociological review*, 164-176.

Cloward, R. A., & Ohlin, L. E. (2013). *Delinquency and opportunity: A study of delinquent gangs* (Vol. 6). Routledge.

Coase, R. H. (1937). The nature of the firm. *Economica*, 4(16), 386-405.

Coase, R. H. (2012). *The firm, the market, and the law*. University of Chicago press.

Cohen, A. K. (1955). DELINQUENT BOYS; THE CULTURE OF THE GANG.

Coopey, J., Keegan, O., & Emler, N. (1997). Managers' Innovations as 'Sense-making'. *British Journal of Management*, *8*(4), 301-315.

Cornell University Library, *Making of America, website available at* *http://digital.library.cornell.edu/m/moa/about.html*

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Crews, K. D. (2012). *Copyright law for librarians and educators: Creative strategies and practical solutions*. American Library Association.

Daft, R. L., & Lengel, R. H. (1986). Organizational information requirements, media richness and structural design. *Management science*, *32*(5), 554-571.

Daft, R. L., & Macintosh, N. B. (1981). A tentative exploration into the amount and equivocality of information processing in organizational work units. *Administrative science quarterly*, *207-224*.

David, P. A. (1990). The dynamo and the computer: an historical perspective on the modern productivity paradox. *The American Economic Review*, *80*(2), 355-361.

Digital Pioneers interview with Paul Conway, dated May 25th, 2010. Available at http://digitalpioneers.library.du.edu/paulConway/transcript.html.

Edwards, P. N., Jackson, S. J., Chalmers, M. K., Bowker, G. C., Borgman, C. L., Ribes, D., ... & Calvert, S. (2013). Knowledge infrastructures: Intellectual frameworks and research challenges.

Eisenstein, E. L. (2005). *The printing revolution in early modern Europe*. Cambridge University Press.

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human perception and performance*, *1*(3), 288.

Fischhoff, B., & Beyth, R. (1975). I knew it would happen: Remembered probabilities of once—future things. *Organizational Behavior and Human Performance*, *13*(1), 1-16.

Fister, B. (2011) "Selfless Audacity" Means Creating a Sustainable Not-a-Business Model," Peer to Peer Review, March 24, 2011.

Foray, D. (2004). *Economics of knowledge*. MIT press.

Forchuk, C., & Roberts, J. (1993). How to critique qualitative research articles. *Canadian Journal of Nursing Research*, *25*, 47-47.

Franke, N., Von Hippel, E., & Schreier, M. (2006). Finding Commercially Attractive User Innovations: A Test of Lead-User Theory*. *Journal of product innovation management*, *23*(4), 301-315.

Gallini, N., & Scotchmer, S. (2002). Intellectual property: when is it the best incentive system?.

In *Innovation Policy and the Economy, Volume 2* (pp. 51-78). MIT Press.

Geertz, C. (1994). Thick description: Toward an interpretive theory of culture. *Readings in the philosophy of social science*, 213-231.

Geffert, B. (2011). Libraries, Publishers, and a Plea for Shotgun Weddings. *The Chronicle of Higher Education*, *20*.

Gergen, K. J. (1982). From self to science: What is there to know. *Psychological perspectives on the self*, *1*, 129-149.

Google (2016). "Google Books Library Project: What does a Google Books Library Project book look like?" available at http://books.google.com/googlebooks/library/index.html

Goold, P. R. (2015). Fair Use: Why Copyright Infringement Is Not a Strict Liability Tort. *Berkeley Technology Law Journal*.

Gordon, W. J. (1982). Fair use as market failure: a structural and economic analysis of the" Betamax" case and its predecessors. *Columbia Law Review*, 1600-1657.

Gordon, W. J. (1991). Asymmetric Market Failure and Prisoner's Dilemma in Intellectual Property. *U. Dayton L. Rev.*, *17*, 853.

Gordon, W. J. (1997). Intellectual property as price discrimination: Implications for contract. *Chi.-Kent L. Rev.*, *73*, 1367.

Gordon, W. J. (2008). Harmless Use: Gleaning from Fields of Copyrighted Works. *Fordham L. Rev.*, *77*, 2411.

Gould, M. (1987). *Revolution in the Development of Capitalism: the Coming of the English Revolution*. Univ of California Press.

Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *ECTJ*, *29*(2), 75-91.

HathiTrust (2008). "Launch of HathiTrust – October 13, 2008," Press Release, available at http://www.hathitrust.org/press_10-13-2008.

HathiTrust (2011). "HathiTrust Constitutional Convention Meeting Notes," available at http://www.hathitrust.org/constitutional_convention2011

HathiTrust (2016). "Mission and Goals," available at https://www.hathitrust.org/mission_goals

HathiTrust Digital Library (2016). "Statistics and Visualizations," available at http://www.hathitrust.org/statistics_visualizations

Heller, M. A. (1998). The tragedy of the anticommons: property in the transition from Marx to markets. *Harvard law review*, 621-688.

Heller, M. A., & Eisenberg, R. S. (1998). Can patents deter innovation? The anticommons in biomedical research. *Science*, *280*(5364), 698-701.

Hippel, E. V. (1982). Appropriability of innovation benefit as a predictor of the source of innovation. *Research policy*, *11*(2), 95-115.

Hippel, E. V. (1986). Lead users: a source of novel product concepts. *Management science*, 32(7), 791-805.

Hippel, E. V., & Krogh, G. V. (2003). Open source software and the "private-collective" innovation model: Issues for organization science. *Organization science*, *14*(2), 209-223.

Hippel, E. V. (2009). Democratizing innovation: the evolving phenomenon of user innovation. *International Journal of Innovation Science*, 1(1), 29-40.

Hirschi, T. (2002). *Causes of delinquency*. Transaction publishers.

Hirschman, A. O. (1959). The Strategy of Economic Development.

Hirschman, A. O. (1970). *Exit, voice, and loyalty: Responses to decline in firms, organizations, and states* (Vol. 25). Harvard University Press.

Hobbes, T. (1969). *Leviathan*, 1651. Scolar Press.

Hughes, T. P. (1993). *Networks of Power: Electrification in Western Society*, 1880-1930. JHU Press.

Isabella, L. A. (1990). Evolving interpretations as a change unfolds: How managers construe key organizational events. *Academy of Management journal*, *33*(1), 7-41.

Jackson, S. J., Gillespie, T., & Payette, S. (2014, February). The policy knot: Re-integrating policy, practice and design in Cscw studies of social computing. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 588-602). ACM.

James, W. (1890). *Principles of Psychology, Vol 2. 1950 Reprint*.

Jenkins, H. (2012). *Textual poachers: Television fans and participatory culture*. Routledge.

Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. NYU press.

Joyce, C., & Patterson, L. (2003). Copyright in 1791: An essay concerning the founders' view of the copyright power granted to Congress in Article I, Section 8, Clause 8 of the US Constitution. *Emory Law Journal*, *52*(909).

Kling, R. (1991). Computerization and social transformations. *Science, Technology & Human Values*, *16*(3), 342-367.

Kling, R. (Ed.). (1996). *Computerization and controversy: value conflicts and social choices*. Morgan Kaufmann.

Kimberly, J. R., & Quinn, R. E. (1984). *Managing organizational transitions*. McGraw-Hill/Irwin.

King, J. L., & Schrems, E. L. (1978). Cost-benefit analysis in information systems development and operation. *ACM Computing Surveys (CSUR)*, *10*(1), 19-34.

Kipling, R. (1982). *The Portable Kipling*. I. Howe (Ed.). Penguin Books.

Landauer, T. K. (1995). *The trouble with computers: Usefulness, usability, and productivity* (Vol. 21). Cambridge, MA: MIT press.

Landes, W. M., & Posner, R. A. (1989). An economic analysis of copyright law. *The Journal of Legal Studies*, 325-363.

Landes, W. M., Posner, R. A., & Landes, W. M. (2009). *The economic structure of intellectual property law*: Harvard University Press.

Lehmann, L., Feldman, M. W., & Kaeuffer, R. (2010). Cumulative cultural dynamics and the coevolution of cultural innovation and transmission: an ESS model for panmictic and structured populations. *Journal of evolutionary biology*, *23*(11), 2356-2369.

Lessig, L. (2004). How I Lost the Big One. *Legal Affairs. Retrieved January*, *15*, 2007.

Leval, P. N. (1990). Toward a fair use standard. *Harvard Law Review*, *103*(5), 1105-1136.

Lincoln, A., & Miller, M. M. (1908). *The wisdom of Abraham Lincoln*. New York: A. Wessels company. “Lecture before the Springfield, IL. Library Association: Discoveries, Inventions and Improvements,” (Feb. 22, 1860).

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry* (Vol. 75). Sage.

Lindlof, T. R., & Taylor, B. C. (2002). Asking, listening, and telling. *Qualitative communication research methods*, 170-208.

Litman, J. (2007). Lawful personal use. *Texas law review*, *85*, 1871.

Litman, J. (2010). Real copyright reform. *Iowa Law Review*, *96*(1), 09-018.

Madison, J. (1788). THE FEDERALIST NO. 43.

Mays, N., & Pope, C. (2000). Assessing quality in qualitative research. *British medical journal*, *320*(7226), 50.

Meltzer, D. J. (1996). The Seminole Decision and State Sovereign Immunity. *The Supreme Court Review*, *1996*, 1-65.

Merton, R. K. (1938). Social structure and anomie. *American sociological review*, *3*(5), 672-682.

Merton, R. K. (1957). Social theory and social structure (rev.

Merton, R. K. (1968). *Social theory and social structure*. Simon and Schuster.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.

Milton, J. (1976). 1644.“. *Areopagitica: A Speech of Mr. John Milton for the Liberty of Unlicensed Printing, to the Parliament of England*.

Nielson, Howard C., Jr., (2005) “Whether The President May Sign a Bill by Directing That His Signature Be Affixed To It,” Report of the United States Department of Justice, available at http://search.justice.gov/search?query=whether+the+president+may+sign+a+bill+directing&op=Search&affiliate=justice.

Pallante, M. A. (2012). Next Great Copyright Act, The. *Colum. JL & Arts*, *36*, 315.

Park and Burgess, *Human Communities*, 1952.

Patry, W. (2000). *Copyright Law and Practice*. The Bureau of National Affairs, Inc., available at http://digital-law-online.info/patry/patry1.html,

Patry, W. (2005). “Productive Use, Transformative Use, Complementary Use: Who’s Right?,” The Patry Copyright Blog, Oct. 28, 2005, available at http://williampatry.blogspot.com/2005/10/productive-use-transformative-use.html

Patry, W. (2009). *Moral panics and the copyright wars*: Oxford University Press.

Penner, D. D., Fitch, G., & Weick, K. E. (1966). Dissonance and the revision of choice criteria. *Journal of personality and social psychology*, *3*(6), 701.

Pfeffer, J. (1981). *Management as symbolic action: The creation and maintenance of organizational paradigms*. Graduate School of Business, Stanford University.

Pfohl, S. J. (1994). *Images of deviance and social control: A sociological history* (p. 2). New York: McGraw-Hill.

Reckless, W. C. (1961). New Theory of Deliquency and Crime, A. *Fed. Probation*, *25*, 42.

Reddy, R. and StClair, G. (2001). “The Million Book Digital Library Project,” Dec. 1, 2001, available at http://www.rr.cs.cmu.edu/mbdl.htm

Ribes, D., & Finholt, T. A. (2009). The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Information Systems*, *10*(5), 375.

Samuelson, P. (2010). Copyright Principles Project: Directions for Reform, The. *25 Berkeley Tech. LJ 1175*.

Schumpeter, J. (1942). Creative destruction. *Capitalism, socialism and democracy*, 82-5.

Seltzer, L. E. (1976). Exemptions and Fair Use in Copyright: The Exclusive Rights Tensions in the New Copyright Act. *Bull. Copyright Soc'y USA*, *24*, 215.

Shapin, S. (1981). Bruno Latour and Steve Woolgar, Laboratory life. The social construction of scientific facts, Beverly Hills, Calif., and London, Sage Publications, 1979. 8vo, pp. 272, illus.,£ 1 1.25 (£ 5.50 paperback). *Medical history*, *25*(03), 341-342.

Shear, Michael D., "Making Legislative History, With Nod from Obama and Stroke of an Autopen," The New York Times, May 28, 2011, available at http://www.nytimes.com/2011/05/28/us/politics/28sign.html?_r=0

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual review of psychology*, *28*(1), 1-39.

Stake, R. E. (1995). *The art of case study research*. Sage.

Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information systems research*, *7*(1), 111-134.

Starbuck, W. H. (1982). Congealing oil: Inventing ideologies to justify acting ideologies out. *Journal of Management Studies*, *19*(1), 3-27.

Staw, B. M. (1980). Rationality and justification in organizational life. *Research in organizational behavior*, *2*, 45-80.

Strauss, C. L. (1962). *Savage mind*. University of Chicago.

Thomas, D. S. (1928). The Child in America: Behavior Problems and Programs.

Thomas, W. I., & Znaniecki, F. (1918). *The Polish peasant in Europe and America: Monograph of an immigrant group* (Vol. 2). University of Chicago Press.

University of Michigan Library, *Making of America*, website available at http://quod.lib.umich.edu/m/moagrp/about.html

Vaidhyanathan, S. (2005). A Risky Gamble with Google. *Chronicle of Higher Education*, *52*(15).

Vaidhyanathan, S. (2006). Googlization of Everything and the Future of Copyright, The. *UC DAViS l. reV.*, *40*, 1207.

Van de Ven, A. H., & Ferry, D. L. (1980). *Measuring and assessing organizations*. John Wiley & Sons.

Vleuten, E. V. D. (2006). Patterns and principles in infrastructure development. Position paper, September 2006.

Walterscheid, E. C. (2002). *The nature of the intellectual property clause: A study in historical perspective*. William S. Hein & Co., Inc..

Weick, K. E., & Prestholdt, P. (1968). Realignment of discrepant reinforcement value. *Journal of personality and social psychology*, *8*(2p1), 180.

Weick, K. E., Gilfillan, D. P., & Keith, T. A. (1973). The effect of composer credibility on orchestra performance. *Sociometry*, 435-462.

Weick, K. E. (1979). Cognitive processes in organizations. *Research in organizational behavior*, *1*(1), 41-74.

Weick, K. E. (1993). Sensemaking in organizations: Small structures with large consequences. *Social psychology in organizations: Advances in theory and research*, 10-37.

Weick, K. E. (1995). *Sensemaking in organizations* (Vol. 3). Sage.

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization science*, *16*(4), 409-421.

Weick, K. E. (2012). *Making sense of the organization, Volume 2: The impermanent organization* (Vol. 2). John Wiley & Sons.

Weiss, R. S. (1995). *Learning from strangers: The art and method of qualitative interview studies*. Simon and Schuster.

Wu, T. (2010). The master switch: The rise and fall of digital empires.

Yates, J. (1993). *Control through communication: The rise of system in American management* (Vol. 6). JHU Press.

Yin, R. (1994). Case study research: Design and methods. Beverly Hills.

Yin, R. (2003). K. (2003). Case study research: Design and methods. *Sage Publications, Inc*, *5*, 11.

Yin, R. K. (2009). Case study research: Design and methods, 4th. *Thousand Oaks*.

Legal Authorities:

8 Anne, c. 19 (1710). "The Statute of Anne," April 10, 1710. Available at http://avalon.law.yale.edu/18th_century/anne_1710.asp

U.S. Const. Art. I, § 8, cl. 8. "The U.S. Constitution Intellectual Property Clause." Available at http://www.archives.gov/exhibits/charters/constitution_transcript.html

17 U.S.C. (1976). "The U.S. Copyright Act." Available at https://www.law.cornell.edu/uscode/text/17

Reports:

Conf. Rep. No. 94-1733, at 71-74 (1976).

House Report No. 83, 90th Cong., 1 Sess. 33 (1968).

House Report No. 94-1476 (1990, 1995, 1999, 2002).

Pub.L. 105-298. Sections amended: 17 U.S.C. §§108, 203(a)(2), 301(c), 302, 303, 304(c)(2).

Section 108 Study Group Report (2008). An Independent Report sponsored by The United States Copyright Office and the National Digital Information Infrastructure and Preservation Program of the Library of Congress, March 2008. Available at http://www.section108.gov/docs/Sec108StudyGroupReport.pdf

Section 108 Study Group (2005), "Overview of the Libraries and Archives Exception in the Copyright Act: Background, History, and Meaning," April 14, 2005. Available at http://www.section108.gov/docs/108BACKGROUNDPAPER%28final%29.pdf

Senate Report No. 983, 93d Cong., 2d Sess. 122 (1974)

Senate Report No. 473, 94th cong., 1st Sess. 70-71 (1975)

Senate Report No. 105-190, at 62 (1998).

Testimony of Mary Bono, "H9952" *Congressional Record*. Government Printing Office. October 7, 1998.

Testimony of Register Kaminstein, *Copyright Law Revision: Hearings Before the Subcomm. on Patents, Trademarks, and Copyrights of the Comm. on the Judiciary on S. 1006*, 89th Cong., 1st Sess.65 (1965).

Testimony of Register Kaminstein, *Copyright Law Revision Part 6: Supplementary Report on the Register of Copyrights on the General Revision of the U.S. Copyright Law: 1965, Revision Bill*, 89th Cong., 1st Sess. xiv-xv (House Comm. Print 1965).

Testimony of Register of Copyrights, LIBRARY REPRODUCTION OF COPYRIGHTED WORKS (17 U.S.C. 108) 14 (1983), at 139.

Case law:

Alden v. Maine, 527 U.S. 706, 119 S. Ct. 2240, 144 L. Ed. 2d 636 (1999).

American Geophysical Union v. Texaco, Inc. 882 F. Supp. 1, 27 (S.D.N.Y. 1992) , *aff'd,* 60 F.3d 913 (2d Cir. 1994).

Authors Guild, Inc. v. Google Inc., 954 F. Supp. 2d 282 (S.D.N.Y. 2013).

Authors Guild, Inc. v. HathiTrust, 755 F.3d 87 (2d Cir. 2014).

AV ex rel. Vanderhye v. iParadigms, LLc, 562 F.3d 630 (4th Cir. 2009).

Bill Graham Archives v. Dorling Kindersley Ltd., 448 F.3d 605 (2d Cir. 2006).

Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569, 114 S. Ct. 1164, 127 L. Ed. 2d 500 (1994).

Cariou v. Prince, 784 F. Supp. 2d 337 (S.D.N.Y. 2011).

Castle Rock Entertainment v. Carol Publishing Group, 150 F.3d 132 (2d Cir. 1998).

Chavez v. Arte Publico Press, 204 F.3d 601 (5th Cir. 2000).

CITIBANK, NA v. MORGAN STANLEY & CO. INTERN., 724 F. Supp. 2d 407 (S.D.N.Y. 2010).

Central Va. Community College v. Katz, 546 U.S. 356, 126 S. Ct. 990, 163 L. Ed. 2d 945 (2006).

City of Boerne v. Flores, 521 U.S. 507, 117 S. Ct. 2157, 138 L. Ed. 2d 624 (1997).

Columbia Broadcasting System, Inc. v. Democratic National Committee, 412 U.S. 94, 93 S. Ct. 2080, 36 L. Ed. 2d 772 (1973).

Edelman v. Jordan, 415 U.S. 651, 94 S. Ct. 1347, 39 L. Ed. 2d 662 (1974).

Eldred v. Ashcroft, 537 U.S. 186, 242-243 (U.S. 2003).

Florida Prepaid Postsecondary Ed. Expense Bd. v. College Savings Bank, 527 U.S. 627, 119 S. Ct. 2199, 144 L. Ed. 2d 575 (1999).

Folsom v. Marsh, 9 F Cas. 342, 344 (C.C.D. Mass 1841).

Ford Co. v. Dept. of Treasury, 323 U.S. 459, 65 S. Ct. 347, 89 L. Ed. 389 (1945).

Fox Film Corp. v. Doyal, 286 U.S. 123, 52 S. Ct. 546, 76 L. Ed. 1010 (1932).

Grant v. Raymond, 6 Pet. 218.

Halo Electronics, Inc. v. Pulse Electronics, Inc., 769 F.3d 1371 (Fed. Cir. 2014).

Hans v. Louisiana, 134 U.S. 1, 10 S. Ct. 504, 33 L. Ed. 842 (1890).

Harper & Row Publishers, Inc. v. National Enters., 471 U.S. 539, 558 (1985).

Hudson v. Palmer, 468 U.S. 517, 104 S. Ct. 3194, 82 L. Ed. 2d 393 (1984).

Kelly v. Arriba Soft Corp., 336 F.3d 811, 818-822 (9th Cir. 2002).

Kendall v. Winsor, 21 How. 322.

Lapides v. Board of Regents of Univ. System of Ga., 535 U.S. 613 (2002).

Leibovitz v. Paramount Pictures Corp., 137 F.3d 109 (2d Cir. 1998).

Lincoln County v. Luning, 133 U.S. 529, 10 S. Ct. 363, 33 L. Ed. 766 (1890).

Pacific and Southern Co., Inc. v. Duncan, 744 F.2d 1490 (11th Cir. 1984).

Parratt v. Taylor, 451 U.S. 527, 101 S. Ct. 1908, 68 L. Ed. 2d 420 (1981).

Pennhurst State School and Hospital v. Halderman, 465 U.S. 89, 104 S. Ct. 900, 79 L. Ed. 2d 67 (1984).

Perfect 10, Inc. v. Amazon. com, Inc., 508 F.3d 1146 (9th Cir. 2007).

Seminole Tribe of Fla. v. Florida, 517 U.S. 44, 116 S. Ct. 1114, 134 L. Ed. 2d 252 (1996).

Sony Corp. of Am. v. Universal City Studios, Inc., 464 U.S. 417, 447-450 (1984).

Sony Comp. Entertainment, Inc. v. Connectix Corp., 203 F.3d 596, 599-601 (2000).

Twentieth Century Music Corp. v. Aiken, 422 U.S. 151, 95 S. Ct. 2040, 45 L. Ed. 2d 84 (1975).

Twin Peaks Productions v. Publications Intern., 996 F.2d 1366 (2d Cir. 1993).

Ty, Inc. v. Publications Int'l Ltd., 292 F. 3d 512, 2002.

UMG Recordings, Inc. v. MP3. Com, Inc., 92 F. Supp. 2d 349 (S.D.N.Y. 2000).

Williams & Wilkins Company v. United States, 487 F.2d 1345 (Ct. Cl. 1973).

Williams & Wilkins Co. v. United States, 420 U.S. 376, 95 S. Ct. 1344, 43 L. Ed. 2d 264 (1975).

Worldwide Church of God v. Philadelphia Church of God, Inc., 227 F.3d 1110 (9th Cir. 2000).

Wright v. Warner Books, Inc., 953 F.2d 731 (2d Cir. 1991).