

2016

# Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust

Alissa Centivany

Western University, [acentiva@uwo.ca](mailto:acentiva@uwo.ca)

Follow this and additional works at: <https://ir.lib.uwo.ca/fimspub>

 Part of the [Library and Information Science Commons](#)

---

## Citation of this paper:

Alissa Centivany. (2016). Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW'16). ACM, New York, NY, USA, 926-940. DOI: <https://doi.org/10.1145/2818048.2820069>

## **Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust**

**Alissa Centivany**

University of Michigan, School of Information Ann Arbor, USA

University of Toronto, Faculties of Law & Information Toronto, Canada [acentiva@umich.edu](mailto:acentiva@umich.edu)

University of Western Ontario, Faculty of Information and Media Studies, London, Canada  
[acentiva@uwo.ca](mailto:acentiva@uwo.ca)

### **ABSTRACT**

The traditional core of CSCW focuses on the relationships, tensions, and gaps between technical systems and social activity. Policy orbits around this core as a persistent but marginally represented presence. In the last few years, however, CSCW has witnessed an upsurge of interest in (re)integrating policy more explicitly and meaningfully into research and practice. For example, recent scholarship stressed the mutually constitutive and interconnected threads of design, practice, and policy [31]. This paper expands upon those motivations through a qualitative case study of the role of policy in library mass digitization work and the subsequent emergence and evolution of the HathiTrust cooperative partnership. By tracing the origins and impacts of early policy decisions in this context, this research contributes to understandings of how and why policy can both close and open spaces of social practice and technical design, functioning as a source of embedded generativity in complex sociotechnical systems.

### **Author Keywords**

Policy; design; practice; generativity; digitization; digital library; digital scholarship.

### **CITATION**

Alissa Centivany. (2016). Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 926-940. DOI: <https://doi.org/10.1145/2818048.2820069>

## INTRODUCTION

The relationship between technical design and social practice has long been a chief concern of CSCW researchers and practitioners. The (sometimes collision-prone) intersections between technological change and social change have been fertile grounds for cultivating insights and understandings into processes of mutual readjustment and co-evolution in sociotechnical systems [1, 38, 40].

In recent years, these core CSCW concerns have benefitted from research that focuses on and integrates the ethical, legal, and policy dimensions of sociotechnical systems [19, 20, 49, 50, 18, 31]. This paper carries forward those motivations by undertaking an in-depth qualitative interview-based case study into the role of policy in library digitization work and the subsequent emergence and evolution of the HathiTrust cooperative partnership. This research explores collaborative computing technologies at the level of *institutions* and *infrastructures* and, in particular, identifies and traces the origins and impacts of a single, seemingly-unobtrusive clause in the cooperative agreement between the University of Michigan and Google (§4.4.2) [55] to illustrate how policy enabled important reconfigurations in the legal, technical, and organizational ecology of academic research libraries.

This research goes beyond previous related work by arguing that policy can function as a source of embedded generativity in sociotechnical systems. In the context of this research, *generativity* can be understood as “a technology’s overall capacity to produce unprompted change driven by large, varied, and uncoordinated audiences” [70:1980] while *embedded* is meant to signal that generativity may become an inseparable part of a broader, more complex sociotechnical system through non-technical means, such as policies. Policy is often conceptualized as a mechanism for channeling and/or regulating social practice and technological design choices, but 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. Approaching policy as a potential source of embedded generativity can benefit CSCW researchers and practitioners working on a broad range of issues and challenges.

This paper begins with a literature review that surveys relevant research on the relationships among policy, technical design, and social practice, focusing in particular on potential synergies with generativity and value-sensitive design research. The paper then provides a description of the methods used and the case, including the §4.4.2 of the University of Michigan-Google Cooperative Agreement.

Next, a brief historical perspective on large-scale digitization efforts is introduced to signal some of the key challenges and tensions that plagued precursors to the mass digitization project (“MDP”) and which, in turn, may have influenced the drafting of §4.4.2. Focus then turns to the empirical pieces of

Alissa Centivany. (2016). Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 926-940. DOI: <https://doi.org/10.1145/2818048.2820069>

this research. Interview data is used to identify key interpretations, points of resistance and contestation, and trace impacts of §4.4.2 through the subsequent emergence and evolution of HathiTrust. Finally, the paper returns to the notion of policy as embedded generativity, drawing again upon interview data to argue that, in addition to bridging gaps or forming entanglements between technical design and social practice, early policy decisions can also open spaces of social practice and technical design and these decisions are often simultaneously value-driven and function as sources as unanticipated innovation and sociotechnical transformation.

## POLICY KNOTS, GENERATIVITY, AND VALUE- SENSITIVE DESIGN

At its founding, policy was among the chief concerns and interests of CSCW researchers. Scholars associated with the Irvine School, for example, regarded policy as having an integral role in the complex relationships between technical design and social activity [36, 37, 38, 39, 40].

Despite its early showing, in the last few decades policy has more or less faded into the background of much CSCW research and practice. When it does appear, policy tends to be unfairly conceptualized as trailing behind design and practice, tacked on as an addendum to some other research inquiry. Policy, and privacy in particular, has been identified in CSCW research in relation to aspects of user experience, social computing applications, scientific collaboration, data sharing, and infrastructure development [4, 7, 17, 35, 42, 53, 65, 66, 67]. While this is not an unsubstantial showing, this work constitutes a microscopic slice of CSCW work in general and tends neither to regard policy as a first-order concern nor recognize its generativity with respect to technical design and social practices. The implication that policy concerns are inconsequential and/or invisible to the core CSCW constituency is somewhat ironic since CSCW, perhaps more so than other information science venues, is deeply aware of and sensitive to the subtle, nuanced, and often invisible forces that shape and are shaped by sociotechnical systems.

In the last few years, however, a proto-constituency of policy-oriented scholars has emerged to take up the mantle of explicitly reintegrating policy back into CSCW research and work. For purposes of clarity, Policy includes public laws that regulate technology design and use as well as private rules, agreements, and express philosophies (such as corporate policy statements). Policy may also channel and constrain technology use by, for, and within firms, organizations, institutions and among their various customers, members, and constituencies. These scholars have argued that the analytic range of CSCW must expand to take on the complex interplay between policy, technical systems development, and social practice if it is to maintain credibility and relevance, and continue making meaningful contributions to social computing and computer-supported collaborative work [31]. For example, researchers recently made the case that the relationships among design, practice, *and policy* are deeply intertwined, mutually constitutive, co- productive, and dynamically bound and therefore understanding the processes of change and innovation in social computing and work assemblages, will require CSCW (and information science research more generally) to recognize and experiment with the policy threads that are invariably interwoven in social computing practice [31].

Despite the growing promise of policy as a research site and modality of inquiry within CSCW, barriers to its integration remain. For example, the lack of domain expertise on issues of ethics, policy and law within the CSCW community may stifle deep engagement with policy-based issues. Despite a growing recognition that law, technology, and social practice are inextricably intertwined, relatively few scholars are able to engage on a rigorous and nuanced level with the legal and policy complexities of technical change and social transformation. A recent CSCW paper that successfully integrated comprehensive discussions of copyright law and social norms in web-based content

Alissa Centivany. (2016). Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 926-940. DOI: <https://doi.org/10.1145/2818048.2820069>

creation [18] provides a glimmer of hope that this obstacle may also be slowly dissolving as information science, computing, and social science programs enhance law, policy, and ethics training, and collaborations across disciplines find institutional support.

Beyond cultivating new areas of expertise within CSCW, there are a number of still relatively unexplored areas of potential synergy between existing CSCW research and technology law and policy literature around questions and theories of generativity and values. In recent years, generativity has been discussed in the context of healthcare infrastructure [6], and scientific collaboration and metadata platforms [41]. The concept has also been developed in technology law and policy literature including most notably Jonathan Zittrain's work on open innovation and internet governance [70, 71, 72] and provided a framework that was adopted, expanded, and/or modified by researchers studying remixing [28], fair use [47] and computer ethics [29]. Zittrain defined generativity as a "technology's overall capacity to produce unprompted change driven by large, varied, and uncoordinated audiences" [70:1980]. Generativity itself may be understood as a value promoted (or a tradeoff considered [28]) through deliberative technical design.

The work of Batya Friedman and colleagues have made enormous strides toward recognizing and capitalizing on the potential of technical systems design to account for and promote particular human values. Rather than ignoring or pushing aside the political or ethical aspects of technologies and information systems, these researchers promote a responsible approach to design and implementation that not only avoids the creation of social harms but actively promotes important shared human values. Public deliberation [8], user autonomy [19], freedom from bias [19], informed consent in online interactions [20], and anticipatory ethics research [48, 49, 50] are just a few of the areas value-sensitive design researchers and anticipatory design ethicists have sought to promote through technical interventions and implementation.

While generativity, value-sensitive design, and anticipatory ethics approaches tend to be rooted in the more design-oriented strands of CSCW research, they offers lessons that could easily be imputed to policy-oriented approaches. Engaging with technological design in a systematic, principled, and deliberative way with full consideration of human values, morals, and ethics is a point of kinship and convergence with policy-oriented approaches to sociotechnical transformation. Even where technical design elements are not the primary focus of the research, as is the case in this paper, value-sensitive design sensibilities and methodologies can offer important clues about broader processes of sociotechnical readjustment and change and meaningfully inform prescriptive, value-driven policies like §4.4.2.

In combination with the recognition that policy, technical design, and social practice are deeply intertwined, mutually constitutive, co-productive, and dynamically bound, I argue that value-sensitive design and anticipatory ethics approaches, and research on generativity can extend the analytic range of policy-oriented CSCW work to help us better understand and experiment with the processes of change, innovation, and transformation in sociotechnical systems. Policies like §4.4.2 discussed herein can function as a locale and modality of value-sensitive organizational design and embedded generativity.

## RESEARCH METHODS

This research seeks to contribute to understandings of the role of policy in reconfiguring the ecology of technical design and social practice in the context library mass digitization and the subsequent emergence and evolution of HathiTrust. The research referred to in this paper reflect one part of a Alissa Centivany. (2016). Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 926-940. DOI: <https://doi.org/10.1145/2818048.2820069>

much larger and more comprehensive study of the interplay between law, emerging technologies, and processes of social transformation.

For purposes of this paper, my central research question was:

*How did policy (specifically §4.4.2 of the UM- Google Cooperative Agreement) impact social practices and technical design choices within the academic research community and, more specifically, in the development of HathiTrust?*

To answer this question, I adopted a qualitative case study approach drawing upon documentary evidence and data generated from in-depth semi-structured interviews. Documentary evidence included contracts, press releases, news articles, published interviews, court filings and opinions, and interpersonal communications. Notwithstanding this trove of documentary data, interviews were the primary source of data for this study. Interviewee recruitment used a purposive sampling approach that targeted current and former architects, employees, and advisors of HathiTrust as well as persons engaged in parallel and competing digitization projects. During data collection, several participants emerged as gatekeepers, facilitating access to additional participants and suggesting trajectories for the supplemental snowball recruitment strategy that was used.

Thirty-one participants from thirteen different institutions were interviewed for this study. Interviews generally lasted one hour and took place face-to-face or over the phone. Many participants were interviewed on multiple occasions. All interviews were recorded with the informed consent of participants and recordings were transcribed by the author and/or Scribie, a professional transcription service.

Institutional capacities and roles varied across the pool of participants. Current and former provosts, university librarians, chief information officers, HathiTrust advisors, employees, and members constituted the bulk of participants. In addition, founders, architects, and employees of other (i.e. non-HathiTrust and non-Google) parallel or competing digitization projects were interviewed. Finally, an external expert with specialized knowledge of the legal and technical aspects of library mass digitization was also consulted.

Several gaps existed in my interview data collection. While I had informal, off-the-record conversations with individuals from Google and university general counsels' offices, those individuals declined to participate in this study citing the active litigations pending in the fall of 2014, when most data collection took place. To the extent the (reasonable) reluctance of these individuals left holes in the research, I attempted to compensate by triangulating the accounts of participants with relevant documentary evidence to bridge the gaps. Furthermore, the caution, risk aversion, and/or secrecy of these parties may contribute to a research finding in a future project.

Designing the study around qualitative interview data enabled me to develop detailed and richly holistic descriptions, integrate multiple (sometimes conflicting) perspectives, learn how events and policies are interpreted, describe and explain process and enactment, and bridge inter-subjectivities. Working within a constructivist paradigm, my goal was *verstehen* -- understanding the meaning and role of §4.4.2, and policy more generally, in the context of the MDP and HathiTrust from my participants' points of view [68]. Especially important when tracing policymaking and its impacts, this approach remains sensitive to the fact that individuals and organizations do not operate in relation to an external, static policy environment but rather continuously create, sustain, and revise images of policy through a process of rationalizing and reorganizing evolving interpretations over

Alissa Centivany. (2016). Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 926-940. DOI: <https://doi.org/10.1145/2818048.2820069>

time [69]. Therefore, although policy may be represented by a fairly static expression (i.e. §4.4.2), the lived experience of policy suggests that a far more dynamic, evolving, and relational understanding may be required.

Data analysis and coding began with the first data collected and continued throughout the analytic process. An iterative, inductive, open coding approach was used. Themes were extracted from the data by memoing and coding passages relevant to my research questions, and the themes were then organized into categories. Early analysis enabled me to refine and reformulate subsequent interview questions and test theories and narrative. In addition, a process of member checking was used to reorganize and refine themes, patterns, and findings as they emerged. Coding and cross-coding of interview transcripts and other artifacts were done by hand.

Drawing upon diverse and rich sources of data enabled me to explore the role of policy in the emergence and development of HathiTrust using a lens with adjustable analytic aperture. I examined policy threads at a granular level and studied their entanglement with complex organizational processes operating within fluctuating real- world contexts.

#### Description of the Case and §4.4.2

HathiTrust was officially launched in 2008 by thirteen research institutions: the members of the Committee for Institutional Cooperation (CIC) and University of California system [25]. Originally conceived of as a shared digital repository, it has since evolved in depth, breadth, and significance. Today HathiTrust is a partnership of over one hundred research institutions working cooperatively to contribute “to the common good by collecting, organizing, preserving, communicating, and sharing the record of human knowledge” [27]. It has also evolved into a digital library comprised of over thirteen million digital volumes (the majority of which resulted from the MDP) and a research center facilitating non-consumptive computational research across the corpus [26]. HathiTrust is a collectively governed organization with an executive director, employees, committees, and a voting membership [26]. It is also part of an institution: as a legal entity HathiTrust is owned by its progenitor, the University of Michigan (UM) [26]. Finally, until early 2015, HathiTrust was a defendant in a major copyright infringement lawsuit; its victory in court has successfully expanded the bounds of fair use to include mass digitization for facilitating full-text search and expanding access for print disabled patrons [14].

This image reflects the way we see and understand HathiTrust today but it does not necessarily reveal much about *how* or *why* it came to be, and came to become *this* HathiTrust. Answering these questions requires untangling the technical, social, and policy threads of HathiTrust’s story. This process of untangling necessarily requires that we trace the origins of HathiTrust back into its dark history, the formative stage prior to its public reveal. In so doing, we discover that HathiTrust’s origins are closely linked to a single provision, § 4.4.2, in the 2004 University of Michigan- Google cooperative agreement:

***4.4.2 Use of U of M Digital Copy in Cooperative Web Services.*** *U of M shall have to right to use the UofMDigitalCopy.inwholeorinpartatUofM’s sole discretion, as part of services offered in cooperation with partner research libraries such as the institutions in the Digital Library Federation.*

Subsequent sections of this work will describe the foundational role this policy played in the MDP and the emergence and evolution of HathiTrust. Before shifting to a more detailed description of its significance, however, I will contextualize the MDP within the broader historical context of large-  
Alissa Centivany. (2016). Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 926-940. DOI: <https://doi.org/10.1145/2818048.2820069>

scale digitization efforts.

## LARGE-SCALE DIGITIZATION IN HISTORICAL PERSPECTIVE

This section argues that, while the MDP was a technically impressive feat of engineering throughput, and the sheer scale and speed of the project provided enough momentum to surpass some of the obstacles that had plagued previous digitization efforts, the technical design elements of the MDP and the existing organizational infrastructure of the academic research library community alone could not account for the transformative impact the MDP had on the practices of the academic research library world. This section illustrates that a fundamental contribution of §4.4.2 was that it enabled the mass digitization technology to be leveraged for collective action around the digitized corpus. Binding together technical feasibility and organizational feasibility through policy was an essential and determining factor in the ultimate success of the MDP for its partners in higher education. This brief account of the MDP precursors and the dark history of HathiTrust illustrates the central role policy §4.4.2 played in the emergence of this important instance of computer-supported cooperative work.

### Early Efforts

Large-scale retrospective digitization efforts were underway long before Google embarked on its MDP in 2004. Beginning in 1978, when the scanner first entered the commercial market, early adopters of digitization tended to come from paper-laden industries and agencies with fairly predictable workflows, viewing digitization as a way to simplify and streamline print distribution channels [15]. By the early 1980s, the Library of Congress and the National Library of Medicine began experimenting with digital document delivery, envisioned at the time as a sort of streamlined interlibrary loan program [16]. In addition, by the mid-1980s, the National Archives became interested in digital preservation standards and provided two million dollars in funding to explore technical standards for digitization through the Optical Digital Imaging Storage System (ODISS) project.

The technology ODISS used at the time was state-of-the-art but, like so many of these early efforts, was nevertheless prone to rapid obsolescence. Near the end of the ODISS project, for example, after the bulk of the project's two million dollars of funding had been spent, 14 inch optical plates were no longer a viable storage medium. The world was moving on to CD-ROMs and CD writable disks. A lead archivist working on ODISS reflected:

*"We no longer had equipment to make use of the large optical scans. I discovered a bridging technology at the Bush Presidential Library in Texas that we could use to format shift the content to CD-ROMs. The whole thing was going to cost an additional \$125,000. But then we figured, if we did this what would we end up with except for a bunch of 200 dpi scanned images? We ultimately concluded it wasn't worth the extra money. All of the scans and indexing from that \$2 million project were thrown away."*

A few key observations can be made about these early forays into digitization such as ODISS. First, they tended to be project-based one-off efforts due to technical limitations (retrospective conversion of print materials was time-consuming and produced low quality results) and financial limitations (projects tended to be costly and resource intensive). Second, there was a tremendous amount of innovation and change underway in digitization technologies and information technologies more generally. The practical implication for many of these early efforts was that rapid innovation forced

Alissa Centivany. (2016). Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 926-940. DOI: <https://doi.org/10.1145/2818048.2820069>



rapid obsolescence and technologists had little incentive for providing interoperability or continued technical support. Third, policy considerations were conspicuously absent from many of these early efforts. The focus was squarely on bridging the sociotechnical gap, figuring out how to design technical solutions for social challenges and problems. This is not to suggest that the absence of policy had a determinative effect on the outcome of these early large-scale digitization efforts. Rather the observation may suggest that, as digitization technologies and efforts developed and evolved over time, so too did the role of policy become more visible, significant, and in the case of HathiTrust, more generative.

### Networked Precursors

The emergence of the World Wide Web in the mid-1990s was a major turning point in the development of large-scale, collaborative digitization projects. One of the first significant efforts was the Making of America project, launched in 1995 as a joint effort of Cornell University and the UM. Its primary goal was to digitize and make accessible primary source materials related to the development of United States infrastructure, from the antebellum period through reconstruction. By 2007, when UM's Making of America website was last updated (it had been largely subsumed by the MDP), its collection consisted of approximately 10,000 books and 50,000 journal articles [58]. Cornell University continued to grow its collection through at least 2012 when its collection included 267 monograph volumes and over 100,000 journal articles with 19<sup>th</sup> century imprints [11].

A secondary aim of the Making of America Project was to engage the broader research and institutional community on standards-setting for digital preservation. Cornell University Library, which now employed a former ODISS project lead, worked toward developing common protocols and consensus on the selection, conversion, storage, retrieval, and use of digitized materials on a large distributed scale. The focus on standards was, in part, a response to the challenges and obstacles that had plagued earlier digitization efforts. Here again we see the sociotechnical gap in effect: the interests and needs of technology developers and the interests and needs of research and institutional community were divided. The architects of the Making of America Project viewed the establishment of norms, technical standards and protocols, and techniques around digitization as opportunities to introduce much-needed sources of stability into the rapidly changing digital environment, safeguarding against obsolescence and the risk that the scans might be thrown away.

There were a number of other efforts undertaken alongside the Making of America Project. For example, the Million Books Project, led by Raj Reddy and colleagues at Carnegie Mellon University, sought to create a free-to-read, searchable collection of one million books, available to everyone over the Internet [45]. In addition, the University of California system had established the California Digital Library to link its community of users to digital information culled from its various campus libraries [57].

Perhaps the most successful, persistent, and concerted digitization work carried out during this period was led by the Internet Archive. Founded in 1996, the Internet Archive started out as a web-archiving service, expanding in 1999 to digitizing and making accessible the ephemeral films contained in the Prelinger Archive, before moving on to digitizing print texts in the mid-2000s [30]. The Internet Archive's print digitization efforts were motivated, in part, by the "secret murmurings" between Google and some members of the research library community. At this time there was a growing sense, and for the Internet Archive a growing concern, that a Google-led MDP would produce a

closed system, potentially locking up or corporatizing our shared cultural record. In an attempt to call attention to these risks, Internet Archive's founder made a bold move to head off the suspected Google MDP:

*"I was receiving an award at a library conference, the Coalition for Networked Information, so I got to make a speech and I used it as an opportunity to basically ask: 'If we build it will you come? If we make it so you can digitize a book for ten cents per page will you support this?'"*

This speech was essentially a call-to-arms (or a call-to- books) for the research library community. Tapping into principles, norms, and practices that libraries had thrived upon for centuries, the Internet Archive proposed a collaborative print digitization project made freely and openly accessible for all. Several key library and funding partners joined the Internet Archive including the Microsoft Corporation, which provided approximately \$10 million in funding between 2006 and 2008, and the University of Toronto [62]. Through these partnerships, the Internet Archive digitized and made publicly available virtually all of the University of Toronto's public domain materials via its website, [www.internetarchive.org](http://www.internetarchive.org). The Internet Archive continues to supply technical and human support to run dozens of scanning centers in five different countries.

The extensive network of partners affiliated with the Internet Archive's print digitization efforts ultimately adopted a unique identifier: the Open Content Alliance. This was done, in part, out of respect for the contributions of the network's various members. While the Internet Archive played a major role in the project, there was a desire to shift ownership, responsibility, and governance of the endeavor to a more representative body. Ultimately, however, some members noted the resistance or difficulty in making the transition:

*"The Open Content Alliance was defined as an alliance of organizations but it never really developed any organizational infrastructure and I think that was somewhat frustrating to many of the libraries that were part of the initiative. Many of us 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; it was not founded as a true collaborative organization. Efforts to turn it into a more collaborative organization were not successful (and as a result) 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 Corporation) and by the Internet Archive as a galvanizing force."*

The examples in this section lead to a couple of key observations. First, technical advances not only improved the outputs of digitization and reduced costs (or made the costs easier to justify) but, even more importantly, made cooperation more feasible and productive. The potential benefits of digitization expanded beyond streamlining print reproduction and improving interlibrary loan processes. Participants now began to recognize the potential of digitization to transform access in terms of readability, discoverability, and reduced information fragmentation. And these benefits were transferrable and shareable amongst all participants.

Second, through these examples we begin to see policy appear in more obvious ways. The policies tend to reflect particular principles or norms: technical standards are important for sustainability so we should develop a policy to on standards-setting; preservation of cultural heritage is important so we should develop a preservation policy; open access is important so we should develop a policy of cooperative engagement toward this goal. As the Open Content Alliance example illustrated, however, policies around collective organizational or institutional governance were still lacking full

buy-in.

Third, the sociotechnical gap still persisted although it was seeming less and less oriented around the technical design aspects of digitization and more and more oriented around the social or organizational aspects. While the Internet Archive and its partners were successful in overcoming or managing some of the technical and financial challenges of large-scale digitization, problems persisted at the social or organizational level. As the Open Content Alliance participant observed, technical infrastructure *and* organizational infrastructure are both requirements for success. The Open Content Alliance, in the view of some of its members, lacked institutional and infrastructural support for the value of collectivism expected by its partners.

This paper now shifts to the MDP which emerged roughly contemporaneously with efforts of the Open Content Alliance. Through the MDP, we begin to see how an overarching policy approach provides some of the organizational infrastructure that had been lacking in those earlier efforts. In addition, I discuss how a particular provision, § 4.4.2 of the UM-Google contract, functioned as a binding force for the technical and social aspects of digitization and primed the way for the emergence of HathiTrust.

## GOOGLE'S MASS DIGITIZATION PROJECT

### Overview of the Google MDP

Google's Library Project, part of Google Book Search, got underway in 2004 when Google entered into digitization agreements with the five institutions now referenced as the "Google 5." Since it first began roughly a decade ago, the MDP has digitized upwards of twenty million volumes. The precise scope of the MDP, however, remains somewhat uncertain due to strict non-disclosure agreements. Google's website, for example, indicates that it has over 40 partner libraries from around the world but it does not provide an exhaustive list of its partners [22]. A reasonably thorough investigation managed to uncover roughly half of its partners (21 out of 40). (Table 1) Moreover, of the 21 known partnerships, only 4 of the contracts are publically available. (Table 1) The result is that information about the precise terms and scope of the MDP is fairly scant. We are still, to some extent, forced to rely upon "secret murmurings" rather than hard facts.

That said, there are several additional sources of information about the MDP. The first is Google's website which provides some clues about the basic features of the project: Google scans and indexes some portion of the partner libraries' print collection; Google makes the full-text of public domain works available for browsing and reading through Google Books Search; Google makes "snippets" (a few sentences of text around the search term) of in-copyright works available through Google Book Search; Google provides basic bibliographic information about the work, as well as relevant information regarding bookstore and library holdings; and, perhaps most significantly, Google covers virtually all of the costs associated with the digitization.

| Institution           | Date     | Scope                |
|-----------------------|----------|----------------------|
| U. Mich. [59]*        | 12/2004  | PD & ©               |
| Harvard [24]          | 12/2004  | PD                   |
| NYPL [44]             | 12/2004  | PD                   |
| Oxford [60]           | 12/2004  | PD                   |
| Stanford [51]         | 12/2004  | PD (pre-1964) then © |
| U. Cal. [56]*         | 8/2006   | PD                   |
| U. Madrid [%]         | 9/2006   | PD                   |
| U. Wisc. [64]         | 10/2006  | PD & univ.-owned     |
| U. Virginia [63]      | 11/2006  | PD & ©               |
| U. Texas-Austin [61]* | 1/2007   | PD & ©               |
| Catalonia [54]        | 1/2007** | PD                   |
| Princeton [43]        | 2/2007   | PD                   |
| Bavarian State [5]    | 3/2007   | PD                   |
| Lausanne [22]         | 5/2007   | PD                   |
| Ghent [22]            | 5/2007   | PD                   |
| CIC [10]*             | 6/2007   | PD & ©               |
| Keio [34]             | 7/2007   | PD                   |
| Cornell [12]          | 8/2007   | PD & ©               |
| Columbia [%]          | 12/2007  | PD                   |
| Austrian Nat'l [3]    | **       | PD                   |
| Lyon [2]              | 7/2008** | PD*                  |

Table 1. Overview of MDP partners and scope based on press releases and publicly available contracts keyed to join date.  
 \* Agreement is publicly available. \*\* Some degree of uncertainty in data.

Additional facts about the MDP have trickled out over time as a result of the copyright infringement lawsuits. (The lawsuit against Google is still pending [13, 46] while the lawsuit against HathiTrust has reached final resolution [14].) For example, during oral arguments before the Second Circuit Court of Appeals in December, 2014, Google’s counsel stated the company’s total expenditures on the project were approximately one hundred and twenty million dollars to date. In addition, Google acknowledged that Stanford University, whose agreement is still confidential, had not only permitted the digitization of in-copyright works but was also making the full-text of those works available to its patrons; that Stanford had been granting this level of access was a shocking revelation for many in the copyright and library worlds [46].

Finally, with respect to the institutions whose agreements have been made public under state freedom of information laws, (Table 1) we can glean information about the MDP, and sensemaking around the MDP, from those contracts and, in the case of this research, from interviews with those involved in the project. Through these agreements and interviews, we begin to see the important and central role policy played in the MDP.

### The Role of Policy in the MDP

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 [9]. While large- scale library digitization efforts had been ongoing for at least a decade prior to the MDP, none of those precursors touched the scale and magnitude of the Google Library Project. To put it in perspective, a participant from Cornell University – one of the early leaders in library digitization – noted: *“In the first week, Google scanned an amount of material equal to what we had scanned in the previous decade.”* An endeavor like this had never before been undertaken; it was full of potential and rife with uncertainty.

By any account, the MDP was an impressive feat of engineering throughput and led to some specific technological innovations (Google redesigned the library book cart, for example, to facilitate the transport of books via truck), it may be somewhat misleading to view the MDP as a primarily technological story. The social and organizational aspects of digitization were equally as important as the technical ones.

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 MDP 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. Copyright, for example, did not rise to predominate the policy landscape because libraries tended to digitize the relatively uncontroversial, low-risk, public domain materials first and, due to resource restriction, never made much headway concerning in-copyright materials.

Policy was significant in the MDP for at least 4 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 MDP. Policymaking, and contracting in particular, operates as a behavioral channel [21]. It forces parties to slow down, deliberate about, and generate prospective justifications for their decisions and proposed activities. It also signals to the outside world that a meaningful agreement exists and, as will be discussed in the next section, can become self-affirming as it is repeated and imitated in future sensemaking and decision making. In particular, an overarching policy schema can create useful pathways capable of resolving some of tensions previously discussed with respect to the sociotechnical pitfalls of large-scale collaborative digitization.

Third, as the next section describes, policy can become a powerful value lever for transformative social change. The next section describes how a specific clause in the UM- Google agreement, § 4.4.2, empowered the UM to leverage the MDP for collective action in the research library world. Finally, as will be discussed later, policy can open and close spaces of social practice and technical design. In this way, we can think of policy as providing a source of embedded generativity to enable new, innovative, often unanticipated, potentially transformative changes to occur (or to block the same) at the intersections of social practice and technical design.

### UM-Google Cooperative Agreement

The UM was one of the first to join the MDP. In addition, the UM-Google agreement is one of the few that have been made public through the invocation of freedom of information laws. Using the agreement itself and interview data generated from architects of that agreement, this section explains the role of policy in the localized and granular context of the MDP and in the subsequent emergence of HathiTrust.

The UM-Google agreement 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 MDP partners refused ownership (presumably to reduce exposure to copyright

Alissa Centivany. (2016). Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 926-940. DOI: <https://doi.org/10.1145/2818048.2820069>

liability). In isolation, these two details are significant but in combination their significance grows.

UM assumed a higher level of risk through its participation relative to the other Google partners but it also assumed a greater willingness capitalize on the potential of Google's offer. Through the MDP, UM created what was essentially a back-up copy of its entire print collection, approximately six million volumes.

Although its significance emerged in the years that followed, arguably the single most important provision in the UM- Google agreement was the one permitting UM to share its copies with other institutions and use them in the provision of cooperative web-based services, §4.4.2.

The importance of this provision is perhaps best described by the individual 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 consistent with the understood meaning of §4.4.2.

It can be tempting, in retrospect, to normalize the evolution of new sociotechnical forms. In hindsight, the MDP, its policies, 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 development were. With respect to §4.4.2 in particular, this study reveals its inclusion in the UM agreement was largely value-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 was largely 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.

Alissa Centivany. (2016). Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 926-940. DOI: <https://doi.org/10.1145/2818048.2820069>

§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 [13], 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 17<sup>th</sup> 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. 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 function and did not seem to pose a serious threat to Google's interests or market dominance.

## POLICY AS EMBEDDED GENERATIVITY

Hindsight bias can pose a significant obstacle to accurately tracing of 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, 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. 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. In fact, when HathiTrust was initially conceived, years after the UM-Google Agreement and the creation of §4.4.2, the controlling vision of HathiTrust was not as a research library collaboration around the provision of web-based services. HathiTrust was a quick, rather unglamorous fix for a simple but potentially serious technical problem:

*“HathiTrust started as a back-up problem, a file back-up problem. We've known for decades that we need to have offsite backup so that problems that hit you aren't likely to hit them. But the Provost made a mistake when we decided to go forward with the MDP. He didn't set aside the money to back this thing up.”*

Moreover, when I interviewed the manager of HathiTrust's day-to-day operations and asked him to identify key milestones or turning points in HathiTrust's development and evolution, one of the first



ones mentioned was the creation of the graphical user interface because, simply put, in its initial conception as a back-up for UM's dark archive, there was no obvious need for a user interface:

*“Reaching a critical mass of institutions necessary to create and launch HathiTrust was a huge milestone. Coming together and deciding that creating HathiTrust was what we wanted to do and laying out objectives which included having an interface to look at the volumes was a big turning point. We were deciding to tie preservation to access, that there's no value to preservation without access. That was a really big thing because the digital corpus was no longer seen as this big block of dark stuff that we can't do anything with. It's malleable. We can flip the switch from closed to open.”*

Breaking apart this observation a little bit more, one way of understanding the significance of the creation of the graphical user interface was that it began to realize the values of collectivism and cooperation central to §4.4.2. Having lived through some of the earlier large-scale digitization efforts (and failures), the architect of this provision recognized that primary obstacle to large-scale or mass 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.”*

Reaching the critical mass required to actually launch HathiTrust, design the graphical interface to access its contents, thinking of ways to creatively pursue non-infringing and fair uses of the corpus and, in some cases, flip the “switch from closed to open” were all steps in HathiTrust's evolution toward a collective. Today, HathiTrust is now comprised of over one hundred research institution partners and, aside from legal issues (which are still filtered through the University of Michigan), is largely self-governed and autonomous. And each of the steps toward collectivism relied, in fundamental ways, upon the values and generativity embedded in that single policy, §4.4.2.

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 expanding beyond text-based services to include audiovisual nodes. This work implicates complex issues ranging from copyright to captioning which libraries are now seeking ways to resolve and manage.

While HathiTrust's memberships consists primarily of research libraries located in the United States, there are also active discussions about how and whether membership might be extended to include other kinds of institutions, organizations, and even individuals. Possibilities of expanding into other geographical locations abound and these too raise new and complex issues around copyright law and

Alissa Centivany. (2016). Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 926-940. DOI: <https://doi.org/10.1145/2818048.2820069>

questions about how the costs and requirements of membership might change to reflect its changing constituency. These are ultimately questions that require HathiTrust to ask: *Who are we? Who is our collective?*

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 this research has focused on the ways in which §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 consortia. The Internet Archive is not part of a library consortia.”*

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 is likely more to the Internet Archive's exclusion than what appears on the surface.

Personal histories, tensions among differing philosophies held by some of the key players at Google, the Internet Archive, and possibly within the academic research library community may be contributing factors. While these are largely outside the scope of this work, it is interesting to note that Google's restrictions on bulk downloading the public domain did not deter some devotees of a free and open public domain. As described by the Internet Archive's founder:

*“Aaron Swartz, who worked at the Internet Archive for 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 copies attributed them to Google and to the library that they came from originally, and we did our best to OCR them, but they are public domain so they are posted on the Internet Archive site. Google, to its credit, did not assert copyright on these digitized materials. We had gotten unsupportive comments from one of the Google libraries*

Alissa Centivany. (2016). Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 926-940. DOI: <https://doi.org/10.1145/2818048.2820069>

*but very supportive comments from another of the Google libraries for having these materials up on the Internet Archive. I found it a very interesting Rorschach test of the librarians and the libraries.”* 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.

## CONCLUSION

The MDP and HathiTrust story 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 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.

This research contributes to the growing recognition within CSCW that policy matters in several ways. First, this research illustrates how policy played an important role in ameliorating some of the risks and challenges that had plagued other large-scale digitization efforts. 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 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. In the context of new and emerging sociotechnical forms, this research suggests that leaving policy analysis and consideration to other fields and other scholars risk losing out on some of the important lessons the CSCW community might glean from the interplay between these interconnected and entangled threads.

Second, this research illustrates how policy does not always trail behind design and practice in the emergence, development, and evolution of innovative social computing forms but rather can function as an important source of embedded generativity. There was a tremendous sense of potential around the MDP in terms of how it might transform research and scholarship and alter the trajectory of library and archive practices around print management, preservation, access and so forth. But no one could anticipate or predict precisely *how* the MDP might change the practices of academic research libraries or institutions of higher education. This work illustrates not only that policy is important, but that policy can precede and enable subsequent design choices and social practices in socio- technical systems.

Third, 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 much CSCW researchers [1]. 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. Borrowing from value-sensitive design research, we can see that the value of collectivism or *collective action around the collective collection* was a driving force in the design of §4.4.2.

Alissa Centivany. (2016). Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 926-940. DOI: <https://doi.org/10.1145/2818048.2820069>

Fourth, we saw how policy can stabilize and coalesce future sensemaking and decision-making around emergent sociotechnical forms, ultimately tightening the design- practice-policy knot [31]. 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.

Finally, this paper contributes to the growing body of CSCW research and practice oriented around policy considerations by illustrating the powerful role of policy plays in the complex interaction between design and practice at a local level and as a source of generativity for future decision making at the intersection of policy, design, and social activity. In this study, we saw how §4.4.2 helped solved some of the persistent challenges that had plagued prior large-scale digitization attempts and technical and social challenges around specific issues like print collection management. We also saw how even a single contract clause can have tremendous impact on the development, evolution, and generativity of subsequent sociotechnical systems.

Without §4.4.2, it is doubtful that the academic research library community would have had HathiTrust. Now, a decade since that clause was drafted, HathiTrust has grown into a cooperative partnership of over 100 partner institutions that collectively govern a shared trusted digital repository, a digital library, and a research center that facilitates non- consumptive computational research across its vast thirteen million volume corpus. And as we've seen, new forms and services are still emerging. More importantly, this research illustrates how CSCW researchers and practitioners interested in the relationships, tensions, gaps, and knots between technical design and social practice might seek ways to use policy to instill values and embed generativity into CSCW work in the future, safeguarding spaces for new, innovative, unanticipated, collaborative, and potentially transformative changes to emerge.

## REFERENCES

1. Mark Ackerman. 2000. The Intellectual Challenge of CSCW: the gap between social requirements and technical feasibility. In *Human-Computer Interaction*, 15(2-3), 179-203.
  2. S. Clément. 2008. Google partenaire numérique officiel de la bibliothèque de Lyon. ActuaLitté – les univers du livre, July 12, 2008. Retrieved May 22, 2015 from <https://www.actualitte.com/bibliotheques/google-partenaire-numerique-officiel-de-la-bibliotheque-de-lyon-3396.htm>
- Alissa Centivany. (2016). Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (CSCW '16). ACM, New York, NY, USA, 926-940. DOI: <https://doi.org/10.1145/2818048.2820069>

3. Austrian National Library. 2015. Austrian Books Online. Retrieved May 22, 2015 from <http://www.onb.ac.at/ev/about/austrianbooksonline.htm>
4. Louise Barkhuus. 2012. The mismeasurement of privacy: using contextual integrity to reconsider privacy in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.
5. Bavarian State Library. 2015. Mass Digitization within the Framework of a Public-Private Partnership between the Bayerische Staatsbibliothek and Google. Retrieved May 22, 2015 from <http://www.bsb-muenchen.de/en/about-us/projects/digitisation/mass-digitisation-within-the-framework-of-a-public-private-partnership-between-the-bayerische-staatsbibliothek-and-google/>
6. Kirsti E. Berntsen. 2014. Generativity for Infrastructuring eHealth.
7. Matthew J. Bietz, Drew Paine, and Charlotte P. Lee. 2013. The work of developing cyberinfrastructure middleware projects. In *Proceedings of the 2013 conference on Computer supported cooperative work*.
8. Alan Borning, Batya Friedman, Janet Davis, and Peyina Lin. 2005. Informing public deliberation: Value sensitive design of indicators for a large-scale urban simulation. *ECSCW*, 449-468. Springer Netherlands.
9. Alissa Centivany. 2015. Innovative Deviance: A Theoretical Framework Emerging at the Intersection of Copyright Law & Technological Change. In *Proceedings of the iConference, 2015*, 1-9. <http://hdl.handle.net/2142/73651>
10. Committee on Institutional Cooperation. 2015. Google Book Search Project – Introduction. Retrieved May 22, 2015 from <http://www.cic.net/projects/library/book-search/introduction>
11. Cornell University. 2015. Making of America. Retrieved May 22, 2015 from <http://digital.library.cornell.edu/m/moa/>
12. Cornell University Library. 2015. Google Book Search Library Project. Retrieved May 22, 2015 from <https://www.library.cornell.edu/google-book-search-library-project>
13. Court of Appeals for the Second Circuit. Authors Guild, Inc. v. Google, Inc., 721 F.3d 132. 2013.
14. Court of Appeals for the Second Circuit. Authors Guild, Inc. v. HathiTrust et al., 755 F.3d 87. 2014.
15. Digital Pioneers. 2010. Interview with Paul Conway, May 25, 2010. Retrieved from <http://digitalpioneers.library.du.edu/paulConway/transcript.html>
16. Digital Pioneers. 2010. Interview with Peter Hirtle, May 25, 2010. Retrieved from <http://digitalpioneers.library.du.edu/peterHirtle/transcript.html>
17. Paul N. Edwards, Steven J. Jackson, Melissa K. Chalmers, Geoffrey C. Bowker, Christine L. Borgman, David Ribes, Matt Burton, and Scout Calvert. 2013. Knowledge infrastructures: Intellectual frameworks and research challenges.
18. Casey Fiesler. 2014. Copyright and social norms in communities of content creation. In *Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing*. 49-52.
19. Batya Friedman. 1996. Value-sensitive design. *interactions* 3(6), 16-23.
20. Batya Friedman, Daniel C. Howe, and Edward Felten. 2002. Informed consent in the Mozilla browser: implementing Value-Sensitive Design. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences (HICSS '02)*.
21. Lon L. Fuller. 1941. Consideration and Form. *41 Col. Law Rev.* 5, 799-824. 1941.
22. Google Books. 2015. Library Partners. Retrieved May 22, 2015 from <http://www.google.ca/googlebooks/library/partners.html>
23. Google Library Project. 2015. Retrieved May 22, 2015 from <https://support.google.com/books/partner/faq/3396243?hl=en&rd=1>
24. Harvard University Library. 2008. Harvard-Google Project. Retrieved May 22, 2015 from <http://hul.harvard.edu/hgproject/index.html>
25. HathiTrust. 2008. Launch of HathiTrust – October 13, 2008. Retrieved May 22, 2015 from Alissa Centivany. (2016). Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 926-940. DOI: <https://doi.org/10.1145/2818048.2820069>

[http://www.hathitrust.org/press\\_10-13-2008](http://www.hathitrust.org/press_10-13-2008)

26. HathiTrust. About – Welcome to the Shared Digital Future. Retrieved May 22, 2015 from <http://www.hathitrust.org/about>

27. HathiTrust. Mission and Goals. Retrieved May 22, 2015 from [http://www.hathitrust.org/mission\\_goals](http://www.hathitrust.org/mission_goals)

28. Benjamin Mako Hill, and Andrés Monroy-Hernández. 2012. The remixing dilemma: The trade-off between generativity and originality. *American Behavioral Scientist* (2012): 0002764212469359.

29. David Sanford Horner. 2010. Metaphors in orbit: revolution, logical malleability, generativity and the future of the internet. *ETHICOMP*. 201-208.

30. Internet Archive. 2015. Retrieved May 22, 2015 from <https://archive.org/index.php>

31. Steven J. Jackson, Tarleton Gillespie, & Sandy Payette. 2014. The policy knot: Re-integrating policy, practice and design in CSCW studies of social computing. In *Proceedings of the 17<sup>th</sup> ACM conference on Computer Supported Cooperative Work (CSCW '14)*, 1-15.

32. Steven J. Jackson, Stephanie B. Steinhardt, Ayse Buyuktur. 2013. Why CSCW needs science policy (and vice versa). In *Proceedings of the 16<sup>th</sup> ACM conference on Computer Supported Cooperative Work (CSCW '13)*.

33. Ricardo Acebes Jiménez. 2008. The partnership project between the Complutense University of Madrid Library and Google Book Search. *Fourth UNICA Scholarly Communication Seminar*, May 15-16, 2008. Charles University, Prague.

34. Keio University. 2013. Keio University's collection digitized by Google Books Library Project, October, 2013.

35. Karina Kervin and Margaret Hedstrom. 2012. How research funding affects data sharing. In *Proceedings of the ACM 2012 conference on computer supported cooperative work companion (CSCW '12)*.

36. John Leslie King, Rob Kling, and the Irvine School. 2004. *The Information Society* 20(2), 97-99.

37. Rob Kling and W. Scacchi. 1982. The Web of Computing: Computer Technology as Social Organization. *Advances in Computers* 21.

38. Rob Kling. 1991. Computerization and social transformation. *Science, Technology & Human Values*, 16(3), 342-367.

39. Rob Kling, Andy Hopper, and James Katz. 1992. Controversies about privacy and open information in CSCW. In *Proceedings of the 1992 ACM conference on Computer-supported cooperative work (CSCW '92)*.

40. Rob Kling. 1996. Computerization and controversy: value conflicts and social choices. Morgan Kaufmann Publishers: San Francisco.

41. Bing Li. 2014. Sustainable value and generativity in the Ecological Metadata Language (EML) platform: toward new knowledge and investigations. *System Sciences (HICSS), 2014 47th Hawaii International Conference on*. IEEE, 2014.

42. Heather Richter Lipford, Pamela J. Wisniewski, Cliff Lampe, Lorraine Kisselburgh, and Kelly Caine. 2012. Reconciling privacy with social media. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work Companion (CSCW '12)*.

43. Princeton University. 2007. Library Joins Google project to make books available online, February 5, 2007. Retrieved May 22, 2015 from

<http://www.princeton.edu/main/news/archive/S16/84/71S02/index.xml?section=topstories>

44. Barbara Quint. 2004. Google and Research Libraries Launch Massive Digitization Project, December 20, 2004. *Information Today, Inc*. Retrieved May 22, 2015 from

<http://newsbreaks.infotoday.com/NewsBreaks/Google-and-Research-Libraries-Launch-Massive-Digitization-Project-16307.asp>

Alissa Centivany. (2016). Policy as Embedded Generativity: A Case Study of the Emergence and Evolution of HathiTrust. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW '16)*. ACM, New York, NY, USA, 926-940. DOI: <https://doi.org/10.1145/2818048.2820069>

45. Raj Reddy & Gloriana StClair. 2001. The Million Book Digital Library Project. Retrieved May 22, 2015 from <http://www.rr.cs.cmu.edu/mbdl.htm>
46. robertjbernsteinblog. 2014. Authors Guild v. Google Oral Argument Before Second Circuit on December 3, 2014, December 6, 2014. Retrieved May 22, 2015 from <https://robertjbernsteinblog.wordpress.com/2014/12/06/authors-guild-v-google-oral-argument-before-second-circuit-on-december-3-2014/>
47. Pamela Samuelson. 2005. The Generativity of Sony v. Universal: The Intellectual Property Legacy of Justice Stevens. *Fordham L. Rev.* 74, 1831.
48. Katie Shilton, Jes A. Koepfler, and Kenneth R. Fleischmann. 2014. How to see values in social computing: methods for studying values dimensions. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing (CSCW '14)*, 426-435.
49. Katie Shilton. 2015. Anticipatory Ethics for a Future Internet: Analyzing Values During the Design of an Internet Infrastructure. *Science and engineering ethics* 21.1 (2015): 1-18.
50. Katie Shilton. 2015. "That's Not An Architecture Problem!": Techniques and Challenges for Practicing Anticipatory Technology Ethics. In *Proceedings of the '15 iConference*.
51. Stanford University Libraries. 2005. Statement of Support and Participation. Retrieved May 22, 2015 from <https://lib.stanford.edu/google-books/statement-support-participation-2005>
52. Stephanie B. Steinhardt & Steven J. Jackson. 2015. Anticipation Work: Cultivating Vision in Collective Practice. In *Proceedings of the 18th ACM conference on Computer Supported Cooperative Work (CSCW '15)*, 1-11.
53. Frederic Stutzman and Woodrow Hartzog. 2012. Boundary regulation in social media. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)*.
54. Tecnología. 2009. Google digitalize 35 mil libros de la Biblioteca de Catalunya libres de derechos de autor, July 27, 2009. Retrieved May 22, 2015 from <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>
55. UM-Google Cooperative Agreement. 2004. Available at <http://www.lib.umich.edu>, last visited May 22, 2015.
56. University of California. 2006. UC Libraries Partner with Google to Digitize Books, August 10, 2006. Retrieved May 22, 2015 from <http://www.cdlib.org/cdlinfo/2006/08/10/uc-libraries-partner-with-google-to-digitize-books/>
57. University of California. 2015. California Digital Library. Retrieved May 22, 2015 from <http://www.cdlib.org/>
58. University of Michigan. 2007. Making of America. 2007. Retrieved May 22, 2015 from <http://quod.lib.umich.edu/m/moagrp/>
59. University of Michigan Library. 2004. Google Digitization Project: A Brief Overview, Dec. 13, 2004. Retrieved May 22, 2015 from <http://www.lib.umich.edu/michigan-digitization-project/michigan-digitization-project-complete-list-resources>
60. University of Oxford Bodleian Libraries. 2012. Oxford Google Books Project. Retrieved May 22, 2015 from <http://www.bodleian.ox.ac.uk/dbooks>
61. University of Texas at Austin. 2007. The University of Texas Libraries Partner with Google to Digitize Books, January 19, 2007. Retrieved May 22, 2015 from <http://www.lib.utexas.edu/about/news/google/>
67. Na Wang, Jens Grossklags, and Heng Xu. 2013. An online experiment of privacy authorization dialogues for social applications. In *Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13)*.
68. Max Weber. 1949. Objectivity in social science and social policy. *The Methodology of the Social Sciences*, 78: Free Press New York.

69. Karl E. Weick. 1995. *Sensemaking in organizations*. Vol. 3. Sage.
70. Jonathan Zittrain, 2006. "The generative internet." *Harvard Law Review*, 1974-2040.
71. Jonathan Zittrain. 2007. Saving the internet. *Harvard Business Review* 85(6), 49.
72. Jonathan Zittrain. 2008. The future of the Internet. *London: Allen Lane*.
62. UniversityofToronto.2015.UniversityofToronto Internet Archive. Retrieved May 22, 2015 from [https://archive.org/details/university\\_of\\_toronto](https://archive.org/details/university_of_toronto)
63. University of Virginia. 2006. U.Va. Library Joins the Google Books Library Project, November 14, 2006. Retrieved May 22, 2015 from <http://news.virginia.edu/content/uva-library-joins-google-books-library-project>
64. University of Wisconsin. 2012. Overview of the Google Library Initiative. Retrieved May 22, 2015 from <http://proxy.library.wisc.edu/digitization/faq.html#primary>
65. Theresa Velden. 2013. Explaining field differences in openness and sharing in scientific communities. In *Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13)*.
66. Janet Vertesi and Paul Dourish. 2011. The value of data: considering the context of production in data economies. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work (CSCW '11)*.