

1 Understanding Mass Digitization

Introduction

Mass digitization is first and foremost a professional concept. While it has become a disciplinary buzzword used to describe large-scale digitization projects of varying scope, it enjoys little circulation beyond the confines of information science and such projects themselves. Yet, as this book argues, it has also become a defining concept of our time. Indeed, it has even attained the status of a cultural and moral imperative and obligation.¹ Today, anyone with an Internet connection can access hundreds of millions of digitized cultural artifacts from the comfort of their desk—or many other locations—and cultural institutions and private bodies add thousands of new cultural works to the digital sphere every day. The practice of mass digitization is forming new nexuses of knowledge, and new ways of engaging with that knowledge. What at first glance appears to be a simple act of digitization (the transformation of singular books from boundary objects to open sets of data), reveals, on closer examination, a complex process teeming with diverse political, legal, and cultural investments and controversies.

This volume asks why mass digitization has become such a “matter of concern,”² and explores its implications for the politics of cultural memory. In practical terms, mass digitization is digitization on an industrial scale. But in cultural terms, mass digitization is much more than this. It is the promise of heightened access to—and better preservation of—the past, and of more original scholarship and better funding opportunities. It also promises entirely new ways of reading, viewing, and structuring archives, new forms of value and their extraction, and new infrastructures of control. This volume argues that the shape-shifting quality of mass digitization, and its

social dynamics, alters the politics of cultural memory institutions. Two movements simultaneously drive mass digitization programs: the relatively new phenomenon of big data gold rushes, and the historically more familiar archival accumulative imperative. Yet despite these prospects, mass digitization projects are also uphill battles. They are costly and speculative processes, with no guaranteed rate of return, and they are constantly faced by numerous limitations and contestations on legal, social, and cultural levels. Nevertheless, both public and private institutions adamantly emphasize the need to digitize on a massive scale, motivating initiatives around the globe—from China to Russia, Africa to Europe, South America to North America. Some of these initiatives are bottom-up projects driven by highly motivated individuals, while others are top-down and governed by complex bureaucratic apparatuses. Some are backed by private money, others publically funded. Some exist as actual archives, while others figure only as projections in policy papers. As the ideal of mass digitization filters into different global empirical situations, the concept of mass digitization attains nuanced political hues. While all projects formally seek to serve the public interest, they are in fact infused with much more diverse, and often conflicting, political and commercial motives and dynamics. The same mass digitization project can even be imbued with different and/or contradictory investments, and can change purpose and function over time, sometimes rapidly.

Mass digitization projects are, then, highly political. But they are not political in the sense that they transfer the politics of analog cultural memory institutions into the digital sphere 1:1, or even liberate cultural memory artifacts from the cultural politics of analog cultural memory institutions. Rather, mass digitization presents a new political cultural memory paradigm, one in which we see strands of technical and ideological continuities combine with new ideals and opportunities; a political cultural memory paradigm that is arguably even more complex—or at least appears more messy to us now—than that of analog institutions, whose politics we have had time to get used to. In order to grasp the political stakes of mass digitization, therefore, we need to approach mass digitization projects not as a continuation of the existing politics of cultural memory, or as purely technical endeavors, but rather as emerging sociopolitical and sociotechnical phenomena that introduce new forms of cultural memory politics.

Framing, Mapping, and Diagnosing Mass Digitization

Interrogating the phenomenon of mass digitization, this book asks the question of how mass digitization affects the politics of cultural memory institutions. As a matter of practice, something is clearly changing in the conversion of bounded—and scarce—historical material into ubiquitous ephemeral data. In addition to the technical aspects of digitization, mass digitization is also changing the political territory of cultural memory objects. Global commercial platforms are increasingly administering and operating their scanning activities in favor of the digital content they reap from the national “data tombs” of museums and libraries and the feedback loops these generate. This integration of commercial platforms into the otherwise primarily public institutional set-up of cultural memory has produced a reconfiguration of the political landscape of cultural memory from the traditional symbolic politics of scarcity, sovereignty, and cultural capital to the late-sovereign infrapolitics of standardization and subversion.

The empirical outlook of the present book is predominantly Western. Yet, the overarching dynamics that have been pursued are far from limited to any one region or continent, nor limited solely to the field of cultural memory. Digitization is a global phenomenon and its reliance on late-sovereign politics and subpolitical governance forms are shared across the globe.

The central argument of this book is that mass digitization heralds a new kind of politics in the regime of cultural memory. Mass digitization of cultural memory is neither a neutral technical process nor a transposition of the politics of analog cultural heritage to the digital realm on a 1:1 scale. The limitations of using conventional cultural-political frameworks for understanding mass digitization projects become clear when working through the concepts and regimes of mass digitization. Mass digitization brings together so many disparate interests and elements that any mono-theoretical lens would fail to account for the numerous political issues arising within the framework of mass digitization. Rather, mass digitization should be approached as an *infrapolitical* process that brings together a multiplicity of interests hitherto foreign to the realm of cultural memory.

The first part of the book, “framing,” outlines the theoretical arguments in the book—that the political dynamics of mass digitization organize themselves around the development of the technical infrastructures of mass digitization in late-sovereign frameworks. Fusing infrastructure theory and theories on the political dynamics of late sovereignty allows us to understand mass digitization projects as cultural phenomena that are highly dependent on standardization and globalization processes, while also recognizing that their resultant infrapolitics can operate as forms of both control and subversion.

The second part of the book, “mapping,” offers an analysis of three different mass digitization phenomena and how they relate to the late-sovereign politics that gave rise to them. The part thus examines the historical foundation, technical infrastructures, and (il)licit status and ideological underpinnings of three variations of mass digitization projects: primarily corporate, primarily public, and primarily private. While these variations may come across as reproductions of more conventional societal structures, the chapters in part two nevertheless also present us with a paradox: while the different mass digitization projects that appear in this book—from Google’s privatized endeavor to Europeana’s supranational politics to the unofficial initiatives of shadow libraries—have different historical and cultural-political trajectories and conventional regimes of governance, they also undermine these conventional categories as they morph and merge into new infrastructures and produce a new form of infrapolitics. The case studies featured in this book are not to be taken as exhaustive examples, but rather as distinct, yet nevertheless entangled, examples of how analog cultural memory is taken online on a digital scale. They have been chosen with the aim of showing the diversity of mass digitization, but also how it, as a phenomenon, ultimately places the user in the dilemma of digital capitalism with its ethos of access, speed, and participation (in varying degrees). The choices also have their limitations, however. In their Western bias, which is partly rooted in this author’s lack of language skills (specifically in Russian and Chinese), for instance, they fail to capture the breadth and particularities of the infrapolitics of mass digitization in other parts of the world. Much more research is needed in this area.

The final part of the book, “diagnosing,” zooms in on the pathologies of mass digitization in relation to affective questions of desire and uncertainty.

This part argues that instead of approaching mass digitization projects as rationalized and instrumental projects, we should rather acknowledge them as ambivalent spatio-temporal projects of desire and uncertainty. Indeed, as the third part concludes, it is exactly uncertainty and desire that organizes the new spatio-temporal infrastructures of cultural memory institutions, where notions such as serendipity and the infrapolitics of platforms have taken precedence over accuracy and sovereign institutional politics. The third part thus calls into question arguments that imagine mass digitization as instrumentalized projects that either undermine or produce values of serendipity, as well as overarching narratives of how mass digitization produces uncomplicated forms of individualized empowerment and freedom. Instead, the chapter draws attention to the new cultural logics of platforms that affect the cultural politics of mass digitization projects.

Crucially, then, this book seeks neither to condemn nor celebrate mass digitization, but rather to unpack the phenomenon and anchor it in its contemporary political reality. It offers a story of the ways in which mass digitization produces new cultural memory institutions online that may be entwined in the cultural politics of their analog origins, but also raises new political questions to the collections.

Setting the Stage: Assembling the Motley Crew of Mass Digitization

The dream and practice of mass digitizing cultural works has been around for decades and, as this section attests, the projects vary significantly in shape, size, and form. While rudimentary and nonexhaustive, this section gathers a motley collection of mass digitization initiatives, from some of the earliest digitization programs to later initiatives. The goal of this section is thus not so much to meticulously map mass digitization programs, but rather to provide examples of projects that might illuminate the purpose of this book and its efforts to highlight the infrastructural politics of mass digitization. As the section attests, mass digitization is anything but a streamlined process. Rather, it is a painstakingly complex process mired in legal, technical, personal, and political challenges and problems, and it is a vision whose grand rhetoric often works to conceal its messy reality.

It is pertinent to note that mass digitization suffers from the combined gendered and racialized reality of cultural institutions, tech corporations, and infrastructural projects: save a few exceptions, there is precious little

diversity in the official map of mass digitization, even in those projects that emerge bottom-up. This does not mean that women and minorities have not formed a crucial part of mass digitization, selecting cultural objects, prepping them (for instance ironing newspapers to ensure that they are flat), scanning them, and constructing their digital infrastructures. However, more often than not, their contributions fade into the background as tenders of the infrastructures of mass digitization rather than as the (predominantly white, male) “face” of mass digitization. As such, an important dimension of the politics of these infrastructural projects is their reproduction of established gendered and racialized infrastructures already present in both cultural institutions and the tech industry.³ This book hints at these crucial dimensions of mass digitization, but much more work is needed to change the familiar cast of cultural memory institutions, both in the analog and digital realms.

With these introductory remarks in place, let us now turn to the long and winding road to mass digitization as we know it today. Locating the exact origins of this road is a subjective task that often ends up trapping the explorer in the mirror halls of technology. But it is worth noting that of course there existed, before the Internet, numerous attempts at capturing and remediating books in scalable forms, for the purposes both of preservation and of extending the reach of library collections. One of the most revolutionary of such technologies before the digital computer or the Internet was microfilm, which was first held forth as a promising technology of preservation and remediation in the middle of the 1800s.⁴ At the beginning of the twentieth century, the Belgian author, entrepreneur, visionary, lawyer, peace activist, and one of the founders of information science, Paul Otlet, brought the possibilities of microfilm to bear directly on the world of libraries. Otlet authored two influential think pieces that outlined the benefits of microfilm as a stable and long-term remediation format that could, ultimately, also be used to extend the reach of literature, just as he and his collaborator, inventor and engineer Robert Goldschmidt, co-authored a work on the new form of the book through microphotography, *Sur une forme nouvelle du livre: le livre microphotographique*.⁵ In his analyses, Otlet suggested that the most important transformations would not take place in the book itself, but in substitutes for it. Some years later, beginning in 1927 with the Library of Congress microfilming more than three million pages of books and manuscripts in the British Library, the remediation

of cultural works in microformat became a widespread practice across the world, and microfilm is still in use to this day.⁶ Otlet did not confine himself to thinking only about microphotography, however, but also pursued a more speculative vein, inspired by contemporary experiments with electromagnetic waves, arguing that the most radical change of the book would be wireless technology. Moreover, he also envisioned and partly realized a physical space, *Mundaneum*, for his dreams of a universal archive. Paul Otlet and Nobel Peace Prize Winner Henri La Fontaine conceived of *Mundaneum* in 1895 as part of their work on documentation science. Otlet called the *Mundaneum* "... an Idea, an Institution, a Method, a Body of work materials and collections, a Building, a Network." In more concrete, but no less ambitious terms, the *Mundaneum* was to gather together all the world's knowledge and classify it according to a universal system they developed called the "Universal Decimal Classification." In 1910, Otlet and Fontaine found a place for their work in the Palais du Cinquanteaire, a government building in Brussels. Later, Otlet commissioned Le Corbusier to design a building for the *Mundaneum* in Geneva. The cooperation ended unsuccessfully, however, and it later led a nomadic life, moving from The Hague to Brussels and then in 1993 to the city of Mons in Belgium, where it now exists as a museum called the *Mundaneum Archive Center*. Fatefully, Mons, a former mining district, also houses Google's largest data center in Europe and it did not take Google long to recognize the cultural value in entering a partnership with the *Mundaneum*, the two parties signing a contract in 2013. The contract entailed among other things that Google would sponsor a traveling exhibit on the *Mundaneum*, as well as a series of talks on Internet issues at the museum and the university, and that the *Mundaneum* would use Google's social networking service, Google Plus, as a promotional tool. An article in the *New York Times* described the partnership as "part of a broader campaign by Google to demonstrate that it is a friend of European culture, at a time when its services are being investigated by regulators on a variety of fronts."⁷ The collaboration not only spurred international interest, but also inspired a group of influential tech activists and artists closely associated with the creative work of shadow libraries to create the critical archival project *Mondotheque.be*, a platform for "discussing and exploring the way knowledge is managed and distributed today in a way that allows us to invent other futures and different narrations of the past,"⁸ and a resulting digital publication project, *The*

Radiated Book, authored by an assembly of activists, artists, and scholars such as Femke Snelting, Tomislav Medak, Dušan Barok, Geraldine Juárez, Shin Joung Yeo, and Matthew Fuller.⁹

Another early precursor of mass digitization emerged with Project Gutenberg, often referred to as the world's oldest digital library. Project Gutenberg was the brainchild of author Michael S. Hart, who in 1971, using technologies such as ARPANET, Bulletin Board Systems (BBS), and Gopher protocols, experimented with publishing and distributing books in digital form. As Hart reminisced in his later text, "The History and Philosophy of Project Gutenberg,"¹⁰ Project Gutenberg emerged out of a donation he received as an undergraduate in 1971, which consisted of \$100 million worth of computing time on the Xerox Sigma V mainframe at the University of Illinois at Urbana-Champaign. Wanting to make good use of the donation, Hart, in his own words, "announced that the greatest value created by computers would not be computing, but would be the storage, retrieval, and searching of what was stored in our libraries."¹¹ He therefore committed himself to converting analog cultural works into digital text in a format not only available to, but also accessible/readable to, almost all computer systems: "Plain Vanilla ASCII" (ASCII for "American Standard Code for Information Interchange"). While Project Gutenberg only converted about 50 works into digital text in the 1970s and the 1980s (the first was the Declaration of Independence), it today hosts up to 56,000 texts in its distinctly lo-fi manner.¹² Interestingly, Michael S. Hart noted very early on that the intention of the project was never to reproduce authoritative editions of works for readers—"who cares whether a certain phrase in Shakespeare has a ':' or a ';' between its clauses"—but rather to "release etexts that are 99.9% accurate in the eyes of the general reader."¹³ As the present book attests, this early statement captures one of the central points of contestation in mass digitization: the trade-off between accuracy and accessibility, raising questions both of the limits of commercialized accelerated digitization processes (see chapter 2 on Google Books) and of class-based and postcolonial implications (see chapter 4 on shadow libraries).

If Project Gutenberg spearheaded the efforts of bringing cultural works into the digital sphere through manual conversion of analog text into lo-fi digital text, a French mass digitization project affiliated with the construction of the Bibliothèque nationale de France (BnF) initiated in 1989 could

be considered one of the earliest examples of actually digitizing cultural works on an industrial scale.¹⁴ The French were thus working on blueprints of mass digitization programs before mass digitization became a widespread practice as part of the construction of a new national library, under the guidance of Alain Giffard and initiated by François Mitterand. In a letter sent in 1990 to Prime Minister Michel Rocard, President Mitterand outlined his vision of a digital library, noting that “the novelty will be in the possibility of using the most modern computer techniques for access to catalogs and documents of the Bibliothèque nationale de France.”¹⁵ The project managed to digitize a body of 70,000–80,000 titles, a sizeable amount of works for its time. As Alain Giffard noted in hindsight, “the main difficulty for a digitization program is to choose the books, and to choose the people to choose the books.”¹⁶ Explaining in a conversation with me how he went about this task, Giffard emphasized that he chose “not librarians but critics, researchers, etc.” This choice, he underlined, could be made only because the digitization program was “the last project of the president and a special mission” and thus not formally a civil service program.¹⁷ The work process was thus as follows:

I asked them to prepare a list. I told them, “Don’t think about what exists. I ask of you a list of books that would be logical in this concept of a library of France.” I had the first list and we showed it to the national library, which was always fighting internally. So I told them, “I want this book to be digitized.” But they would never give it to us because of territory. Their ship was not my ship. So I said to them, “If you don’t give me the books I shall buy the books.” They said I could never buy them, but then I started buying the books from antiques suppliers because I earned a lot of money at that time. So in the end I had a lot of books. And I said to them, “If you want the books digitized you must give me the books.” But of the 80,000 books that were digitized, half were not in the collection. I used the staff’s garages for the books, 80,000 books. It is an incredible story.¹⁸

Incredible indeed. And a wonderful anecdote that makes clear that mass digitization, rather than being just a technical challenge, is also a politically contingent process that raises fundamental questions of territory (institutional as well as national), materiality, and culture. The integration of the digital *très grande bibliothèque* into the French national mass digitization project Gallica, later in 1997, also foregrounds the infrastructural trajectory of early national digitization programs into later glocal initiatives.¹⁹

The question of pan-national digitization programs was precisely at the forefront of another early prominent mass digitization project, namely the

Universal Digital Library (UDL), which was launched in 1995 by Carnegie Mellon computer scientist Raj Reddy and developed by linguist Jaime Carbonell, physicist Michael Shamos, and Carnegie Mellon Foundation dean of libraries Gloriana St. Clair. In 1998, the project launched the Thousand Book Project. Later, the UDL scaled its initial efforts up to the Million Book Project, which they successfully completed in 2007.²⁰ Organizationally, the UDL stood out from many of the other digitization projects by including initial participation from three non-Western entities in addition to the Carnegie Mellon Foundation—the governments of India, China, and Egypt.²¹ Indeed, India and China invested about \$10 million in the initial phase, employing several hundred people to find books, bring them in, and take them back. While the project ambitiously aimed to provide access “to all human knowledge, anytime, anywhere,” it ended its scanning activities 2008. As such, the Universal Digital Library points to another central infrastructural dimension of mass digitization: its highly contingent spatio-temporal configurations that are often posed in direct contradistinction to the universalizing discourse of mass digitization. Across the board, mass digitization projects, while confining themselves in practice to a limited target of how many books they will digitize, employ a discourse of universality, perhaps alluding vaguely to how long such an endeavor will take but in highly uncertain terms (see chapters 3 and 5 in particular).

No exception from the universalizing discourse, another highly significant mass digitization project, the Internet Archive, emerged around the same time as the Universal Digital Library. The Internet Archive was founded by open access activist and computer engineer Brewster Kahle in 1996, and although it was primarily oriented toward preserving born-digital material, in particular the Internet (*Wired* calls Brewster Kahle “the Internet’s de facto librarian”²²), the Archive also began digitizing books in 2005, supported by a grant from the Alfred Sloan Foundation. Later that year, the Internet Archive created the infrastructural initiative, Open Content Alliance (OCA), and was now embedded in an infrastructure that included over 30 major US libraries, as well as major search engines (by Yahoo! and Microsoft), technology companies (Adobe and Xerox), a commercial publisher (O’Reilly Media, Inc.), and a not-for-profit membership organization of more than 150 institutions, including universities, research libraries, archives, museums, and historical societies.²³ The Internet Archive’s mass

digitization infrastructure was thus from the beginning a mesh of public and private cooperation, where libraries made their collections available to the Alliance for scanning, and corporate sponsors or the Internet Archive conversely funded the digitization processes. As such, the infrastructures of the Internet Archive and Google Books were rather similar in their setups.²⁴ Nevertheless, the initiative of the Internet Archive's mass digitization project and its attendant infrastructural alliance, OCA, should be read as both a technical infrastructure responding to the question of *how* to mass digitize in technical terms, and as an infrapolitical reaction in response to the forces of the commercial world that were beginning to gather around mass digitization, such as Amazon²⁵ and Google. The Internet Archive thus positioned itself as a transparent open source alternative to the closed doors of corporate and commercial initiatives. Yet, as Kalev Leetaru notes, the case was more complex than that. Indeed, while the OCA was often foregrounded as more transparent than Google, their technical infrastructural components and practices were in fact often just as shrouded in secrecy.²⁶ As such, the Internet Archive and the OCA draw attention to the important infrapolitical question in mass digitization, namely *how*, *why*, and *when* to manage visibilities in mass digitization projects.

Although the media sometimes picked up stories on mass digitization projects already outlined, it wasn't until Google entered the scene that mass digitization became a headline-grabbing enterprise. In 2004, Google founders Larry Page and Sergey Brin traveled to Frankfurt to make a rare appearance at the Frankfurt Book Fair. Google was at that time still considered a "scrappy" Internet company in some quarters, as compared with tech giants such as Microsoft.²⁷ Yet Page and Brin went to Frankfurt to deliver a monumental announcement: Google would launch a ten-year plan to make available approximately 15 million digitized books, both in- and out-of-copyright works.²⁸ They baptized the program "Google Print," a project that consisted of a series of partnerships between Google and five English-language libraries: the University of Michigan at Ann Arbor, Stanford, Harvard, Oxford (Bodleian Library), and the New York City Public Library. While Page's and Brin's announcement was surprising to some, many had anticipated it; as already noted, advances toward mass digitization proper had already been made, and some of the partnership institutions had been negotiating with Google since 2002.²⁹ As with many of the previous mass digitization projects, Google found inspiration for their digitization project

in the long-lived utopian ideal of the universal library, and in particular the mythic library of Alexandria.³⁰ As with other Google endeavors, it seemed that Page was intent on realizing a utopian ideal that scholars (and others) had long dreamed of: a library containing everything ever written. It would be realized, however, not with traditional human-centered means drawn from the world of libraries, but rather with an AI approach. Google Books would exceed human constraints, taking the seemingly impossible vision of digitizing all the books in the world as a starting point for constructing an omniscient Artificial Intelligence that would know the entire human symbol system and allow flexible and intuitive recollection. These constraints were physical (how to digitize and organize all this knowledge in physical form); legal (how to do it in a way that suspends existing regulation); and political (how to transgress territorial systems). The invocation of the notion of the universal library was not a neutral action. Rather, the image of Google Books as a library worked as a symbolic form in a cultural scheme that situated Google as a utopian, and even ethical, idealist project. Google Books seemingly existed by virtue of Goethe's famous maxim that "To live in the ideal world is to treat the impossible as if it were possible."³¹ At the time, the industry magazine *Bookseller* wrote in response to Google's digitization plans: "The prospect is both thrilling and frightening for the book industry, raising a host of technical and theoretical issues."³² And indeed, while some reacted with enthusiasm and relief to the prospect of an organization being willing to suffer the cost of mass digitization, others expressed economic and ethical concerns. The Authors Guild, a New York-based association, promptly filed a copyright infringement suit against Google. And librarians were forced to revisit core ethical principles such as privacy and public access.

The controversies of Google Books initially played out only in US territory. However, another set of concerns of a more territorial and political nature soon came to light. The French President at the time, Jacques Chirac, called France to cultural-political arms, urging his culture minister, Renaud Donnedieu de Vabres, and Jean-Noël Jeanneney, then-head of France's Bibliothèque nationale, to do the same with French texts as Google planned to do with their partner libraries, but by means of a French search engine.³³ Jeanneney initially framed this French cultural-political endeavor as a European "contre-attaque" against Google Books, which, according to Jeanneney, could pose "une domination écrasante de l'Amérique dans la

définition de l'idée que les prochaines générations se feront du monde." ("a crushing American domination of the formation of future generations' ideas about the world")³⁴ Other French officials insisted that the French digitization project should be seen not primarily as a cultural-political reaction *against* Google, but rather as a cultural-political incentive within France and Europe to make European information available online. "I really stress that it's not anti-American," an official at France's Ministry of Culture and Communication, speaking on the condition of anonymity, noted in an interview. "It is not a reaction. The objective is to make more material relevant to European heritage available. ... Everybody is working on digitization projects." Furthermore, the official did not rule out potential cooperation between Google and the European project.³⁵ There was no doubt, however, that the move to mass digitization "was a political drive by the French," as Stephen Bury, head of European and American collections at the British Library, emphasized.³⁶

Despite its mixed messages, the French reaction nevertheless underscored the controversial nature of mass digitization as a symbolic, as well as technical, aspiration: mass digitization was a process that not only neutrally scanned and represented books but could also produce a new mode of world-making, actively structuring archives as well as their users.³⁷ Now questions began to surface about where, or with whom, to place governance over this new archive: who would be the custodian of the keys to this new library? And who would be the librarians? A series of related questions could also be asked: who would determine the archival limits, the relations between the secret and the non-secret or the private and the public, and whether these might involve property or access rights, publication or reproduction rights, classification, and putting into order? France soon managed to rally other EU countries (Spain, Poland, Hungary, Italy, and Germany) to back its recommendation to the European Commission (EC) to construct a European alternative to Google's search engine and archive and to set this out in writing. Occasioned by the French recommendation, the EC promptly adopted the idea of Europeana—the name of the proposed alternative—as a "flagship project" for the budding EU cultural policy.³⁸ Soon after, in 2008, the EC launched Europeana, giving access to some 4.5 million digital objects from more than 1,000 institutions.

Europeana's Europeanizing discourse presents a territorializing approach to mass digitization that stands in contrast to the more universalizing

tone of Mundaneum, Gutenberg, Google Books, and the Universal Digital Library. As such, it ties in with our final examples, namely the sovereign mass digitization projects that have in fact always been one of the primary drivers in mass digitization efforts. To this day, the map of mass digitization is populated with sovereign mass digitization efforts from Holland and Norway to France and the United States. One of the most impressive projects is the Norwegian mass digitization project at the National Library of Norway, which since 2004 has worked systematically to develop a digital National Library that encompasses text, audio, video, image, and websites. Impressively, the National Library of Norway offers digital library services that provide online access (to all with a Norwegian IP address) to full-text versions of all books published in Norway up until the year 2001, access to digital newspaper collections from the major national and regional newspapers in all libraries in the country, and opportunities for everyone with Internet access to search and listen to more than 40,000 radio programs recorded between 1933 and the present day.³⁹ Another ambitious national mass digitization project is the Dutch National Library's effort to digitize all printed publications since 1470 and to create a National Platform for Digital Publications, which is to act both as a content delivery platform for its mass digitization output and as a national aggregator for publications. To this end, the Dutch National Library made deals with Google Books and Proquest to digitize 42 million pages just as it entered into partnerships with cross-domain aggregators such as Europeana.⁴⁰ Finally, it is imperative to mention the Digital Public Library of America (DPLA), a national digital library conceived of in 2010 and launched in 2013, which aggregates digital collections of metadata from around the United States, pulling in content from large institutions like the National Archives and Records Administration and HathiTrust, as well as from smaller archives. The DPLA is in great part the fruit of the intellectual work of Harvard University's Berkman Center for Internet and Society and the work of its Steering Committee, which consisted of influential names from the digital, legal, and library worlds, such as Robert Darnton, Maura Marx, and John Palfrey from Harvard University; Paul Courant of the University of Michigan; Carla Hayden, then of Baltimore's Enoch Pratt Free Library and subsequently the Librarian of Congress; Brewster Kahle; Jerome McGann; Amy Ryan of the Boston Public Library; and Doron Weber of the Sloan Foundation. Key figures in the

DPLA have often to great rhetorical effect positioned DPLA vis-à-vis Google Books, partly as a question of public versus private infrastructures.⁴¹ Yet, as the then-Chairman of DPLA John Palfrey conceded, the question of what constitutes “public” in a mass digitization context remains a critical issue: “The Digital Public Library of America has its critics. One counterargument is that investments in digital infrastructures at scale will undermine support for the traditional and the local. As the chairman of the DPLA, I hear this critique in the question-and-answer period of nearly every presentation I give. ... The concern is that support for the DPLA will undercut already eroding support for small, local public libraries.”⁴² While Palfrey offers good arguments for why the DPLA could easily work in unison with, rather than jeopardize, smaller public libraries, and while the DPLA is building infrastructures to support this claim,⁴³ the discussion nevertheless highlights the difficulties with determining when something is “public,” and even national.

While the highly publicized and institutionalized projects I have just recounted have taken center stage in the early and later years of mass digitization, they neither constitute the full cast, nor the whole machinery, of mass digitization assemblages. Indeed, as chapter 4 in this book charts, at the margins of mass digitization another set of actors have been at work building new digital cultural memory assemblages, including projects such as Monoskop and Lib.ru. These actors, referred to in this book as shadow library projects (see chapter 4), at once both challenge and confirm the broader infrapolitical dimensions of mass digitization, including its logics of digital capitalism, network power, and territorial reconfigurations of cultural memory between universalizing and glocalizing discourses. Within this new “ecosystem of access,” unauthorized archives as Libgen, Giga-pedia, and Sci-Hub have successfully built “shadow libraries” with global reach, containing massive aggregations of downloadable text material of both scholarly and fictional character.⁴⁴ As chapter 4 shows, these initiatives further challenge our notions of public good, licit and illicit mass digitization, and the territorial borders of mass digitization, just as they add another layer of complexity to the question of the politics of mass digitization.

Today, then, the landscape of mass digitization has evolved considerably, and we can now begin to make out the political contours that have

shaped, and continue to shape, the emergent contemporary knowledge infrastructures of mass digitization, ripe as they are with contestation, cooperation, and competition. From this perspective, mass digitization appears as a preeminent example of how knowledge politics are configured in today's world of "assemblages" as "multisited, transboundary networks" that connect subnational, national, supranational, and global infrastructures and actors, without, however, necessarily doing so through formal interstate systems.⁴⁵ We can also see that mass digitization projects did not arise as a result of a sovereign decision, but rather emerged through a series of contingencies shaped by late-capitalist and late-sovereign forces. Furthermore, mass digitization presents us with an entirely new cultural memory paradigm—a paradigm that requires a shift in thinking about cultural works, collections, and contexts, from cultural records to be preserved and read by humans, to ephemeral machine-readable entities. This change requires a shift in thinking about the economy of cultural works, collections, and contexts, from scarce institutional objects to ubiquitous flexible information. Finally, it requires a shift in thinking about these same issues as belonging to national-global domains to conceiving them in terms of a set of political processes that may well be placed in national settings, but are oriented toward global agendas and systems.

Interrogating Mass Digitization

Mass digitization is often elastic in definition and elusive in practice. Concrete attempts have been made to delimit what mass digitization is, but these rarely go into specifics. The two characteristics most commonly associated with mass digitization are the relative lack of selectivity of materials, as compared to smaller-scale digitization projects, and the high speed and high volume of the process in terms of both digital conversion and metadata creation, which are made possible through a high level of automation.⁴⁶ Mass digitization is thus concerned not only with preservation, but also with what kind of knowledge practices and values technology allows for and encourages, for example, in relation to de- and recontextualization, automation, and scale.⁴⁷

Studies of mass digitization are commonly oriented toward technology or information policy issues close to libraries, such as copyright, the quality of digital imagery, long-term preservation responsibility, standards

and interoperability, and economic models for libraries, publishers, and booksellers, rather than, as here, the exploration of theory.⁴⁸ This is not to say that existing work on mass digitization is not informed by theoretical considerations, but rather that the majority of research emphasizes policy and technical implementation at the expense of a more fundamental understanding of the cultural implications of mass digitization. In part, the reason for this is the relative novelty of mass digitization as an identifiable field of practice and policy, and its significant ramifications in the fields of law and information science.⁴⁹ In addition to scholarly elucidations, mass digitization has also given rise to more ideologically fuelled critical books and articles on the topic.⁵⁰

Despite its disciplinary branching, work on mass digitization has mainly taken place in the fields of information science, law, and computer science, and has primarily problematized the “hows” of mass digitization and not the “whys.”⁵¹ As with technical work on mass digitization, most nontechnical studies of mass digitization are “problem-solving” rather than “critical,” and this applies in particular to work originating from within the policy analysis community. This body seeks to solve problems within the existing social order—for example, copyright or metadata—rather than to interrogate the assumptions that underlie mass digitization programs, which would include asking what kinds of knowledge production mass digitization gives rise to. How does mass digitization change the ideological infrastructures of cultural heritage institutions? And from what political context does the urge to digitize on an industrial scale emerge? While the technical and problem-solving corpus on mass digitization is highly valuable in terms of outlining the most important stakeholders and technical issues of the field, it does not provide insight into the deeper structures, social mechanisms, and political implications of mass digitization. Moreover, it often fails to account for digitization as a force that is deeply entwined with other dynamics that shape its development and uses. It is this lack that the present volume seeks to mitigate.

Assembling Mass Digitization

Mass digitization is a composite and fluctuating infrastructure of disciplines, interests, and forces rooted in public-private assemblages, driven by ideas of value extraction and distribution, and supported by new forms of

social organization. Google Books, for instance, is both a commercial project covered by nondisclosure agreements *and* an academic scholarly project open for all to see. Similarly, Europeana is both a public digitization project directed at “citizens” *and* a public-private partnership enterprise ripe with profit motives. Nevertheless, while it is tempting to speak about specific mass digitization projects such as Google Books and Europeana in monolithic and contrastive terms, mass digitization projects are anything but tightly organized, institutionally delineated, coherent wholes that produce one dominant reading. We do not find one “essence” in mass digitized archives. They are not “enlightenment projects,” “library services,” “software applications,” “interfaces,” or “corporations.” Nor are they rooted in one central location or single ideology. Rather, mass digitization is a complex material and social infrastructure performed by a diverse constellation of cultural memory professionals, computer scientists, information specialists, policy personnel, politicians, scanners, and scholars. Hence, this volume approaches mass digitization projects as “assemblages,” that is, as contingent arrangements consisting of humans, machines, objects, subjects, spaces and places, habits, norms, laws, politics, and so on. These arrangements cross national-global and public-private lines, producing what this volume calls “late-sovereign,” “posthuman,” and “late-capitalist” assemblages.

To give an example, we can look at how the national and global aspects of cultural memory institutions change with mass digitization. The national museums and libraries we frequent today were largely erected during eras of high nationalism, as supreme acts of cultural and national territoriality. “The early establishment of a national collection,” as Belinda Tiffen notes, “was an important step in the birth of the new nation,” since it signified “the legitimacy of the nation as a political and cultural entity with its own heritage and culture worthy of being recorded and preserved.”⁵² Today, as the initial French incentive to build Europeana shows, we find similar nationalization processes in mass digitization projects. However, nationalizing a digital collection often remains a performative gesture than a practical feat, partly because the information environment in the digital sphere differs significantly from that of the analog world in terms of territory and materiality, and partly because the dichotomy between national and global, an agreed-upon construction for centuries, is becoming more and more difficult to uphold in theory and practice.⁵³ Thus, both Google

Books and Europeana link to sovereign frameworks such as citizens and national representation, while also undermining them with late-capitalist transnational economic agreements.

A related example is the posthuman aspect of cultural memory politics. Cultural memory artifacts have always been thought of as profoundly human collections, in the sense that they were created by and for human minds and human meaning-making. Previously, humans also organized collections. But with the invention of computers, most cultural memory institutions also introduced a machine element to the management of accelerating amounts of information, such as computerized catalog systems and recollection systems. With the advent of mass digitization, machines have gained a whole new role in the cultural memory ecosystem, not only as managers, but also as interpreters. Thus, collections are increasingly digitized to be read by machines instead of humans, just as metadata is now becoming a question of machine analysis rather than of human contextualization. Machines are taking on more and more tasks in the realm of cultural memory that require a substantial amount of cognitive insight (just as mass digitization has created the need for new robot-like, and often poorly paid, human tasks, such as the monotonous work of book scanning). Mass digitization has thereby given rise to an entirely new cultural-legal category titled “non-consumptive research,” a term used to describe the large-scale analysis of texts, and which has been formalized by the Google Books Settlement, for instance, in the following way: “research in which computational analysis is performed on one or more books, but not research in which a researcher reads or displays.”⁵⁴

Lastly, mass digitization connects the politics of cultural memory to transnational late capitalism, and to one of its expressions in particular: digital capitalism.⁵⁵ Of course, cultural memory collections have a long history with capitalism. The nineteenth century held very fuzzy boundaries between the cultural functions of libraries and the commercial interests that surrounded them, and, as historian of libraries Francis Miksa notes, Melvin Dewey, inventor of the Dewey Decimal System, was a great admirer of the corporate ideal, and was eager to apply it to the library system.⁵⁶ Indeed, library development in the United States was greatly advanced by the philanthropy of capitalism, most notably by Andrew Carnegie.⁵⁷ The question, then, is not so much whether mass digitization has brought cultural memory institutions, and their collections and users, into a capitalist

system, but *what kind* of capitalist system mass digitization has introduced cultural memory to: digital capitalism.

Today, elements of the politics of cultural memory are being reassembled into novel knowledge configurations. As a consequence, their connections and conjugations are being transformed, as are their institutional embeddings. Indeed, mass digitization assemblages are a product of our time. They are new forms of knowledge institutions arising from a sociopolitical environment where vertical territorial hierarchies and horizontal networks entwine in a new political mesh: where solid things melt into air, and clouds materialize as material infrastructures, where boundaries between experts and laypeople disintegrate, and where machine cognition operates on a par with human cognition on an increasingly large scale. These assemblages enable new types of political actors—networked assemblages—which hold particular forms of power despite their informality vis-à-vis the formal political system; and in turn, through their practices, these actors partly build and shape those assemblages.

Since concepts always respond to “a specific social and historical situation of which an intellectual occasion is part,”⁵⁸ it is instructive to revisit the 1980s, when the theoretical notion of assemblage emerged and slowly gained cross-disciplinary purchase.⁵⁹ Around this time, the stable structures of modernist institutions began to give ground to postmodern forces: sovereign systems entered into supra-, trans-, and international structures, “globalization” became a buzzword, and privatizing initiatives drove wedges into the foundations of state structures. The centralized power exercised by disciplinary institutions was increasingly distributed along more and more lines, weakening the walls of circumscribed centralized authority.⁶⁰ This disciplinary decomposition took place on all levels and across all fields of society, including institutional cultural memory containers such as libraries and museums. The forces of privatization, globalization, and digitization put pressures not only on the authority of these institutions but also on a host of related authoritative cultural memory elements, such as “librarians,” “cultural works,” and “taxonomies,” and cultural memory practices such as “curating,” “reading,” and “ownership.” Librarians were “disintermediated” by technology, cultural works fragmented into flexible data, and curatorial principles were revised and restructured just as reading was now beginning to take place in front of screens, meaning-making to

be performed by machines, and ownership of works to be substituted by contractual renewals.

Thinking about mass digitization as an “assemblage” allows us to abandon the image of a circumscribed entity in favor of approaching it as an aggregate of many highly varied components and their contingent connections: scanners, servers, reading devices, cables, algorithms; national, EU, and US policymakers; corporate CEOs and employees; cultural heritage professionals and laypeople; software developers, engineers, lobby organizations, and unsalaried labor; legal settlements, academic conferences, position papers, and so on. It gives us pause—every time we say “Google” or “Europeana,” we might reflect on what we actually mean. Does the researcher employed by a university library and working with Google Books also belong to Google Books? Do the underpaid scanners? Do the users of Google? Or, when we refer to Google Books, do we rather only mean to include the founders and CEOs of Google? Or has Google in fact become a metaphor that expresses certain characteristics of our time? The present volume suggests that all these components enter into the new phenomenon of mass digitization and produce a new field of potentiality, while at the same time they retain their original qualities and value systems, at least to some extent. No assemblage is whole and imperturbable, nor entirely reducible to its parts, but is simultaneously an accumulation of smaller assemblages and a member of larger ones.⁶¹ Thus Google Books, for example, is both an aggregation of smaller assemblages such as university libraries, scanners (both humans and machines), and books, *and* a member of larger assemblages such as Google, Silicon Valley, neoliberal lobbies, and the Internet, to name but a few.

While representations of assemblages such as the analyses performed in this volume are always doomed to misrepresent empirical reality on some level, this approach nevertheless provides a tool for grasping at least some of mass digitization’s internal heterogeneity, and the mechanisms and processes that enable each project’s continued assembled existence. The concept of the assemblage allows us to grasp mass digitization as comprised of ephemeral projects that are uncertain by nature, and sometimes even made up of contradictory components.⁶² It also allows us to recognize that they are more than mere networks: while ephemeral and networked, something enables them to cohere. Bruno Latour writes, “Groups are not

silent things, but rather the provisional product of a constant uproar made by the millions of contradictory voices about what is a group and who pertains to what.”⁶³ It is the “taming and constraining of this multivocality,” in particular by communities of knowledge and everyday practices, that enables something like mass digitization to cohere as an assemblage.⁶⁴ This book is, among other things, about those communities and practices, and the politics they produce and are produced by. In particular, it addresses the politics of mass digitization as an infrapolitical activity that retreats into, and emanates from, digital infrastructures and the network effects they produce.

Politics in Mass Digitization: Infrastructure and Infrapolitics

If the concept of “assemblage” allows us to see the relational set-up of mass digitization, it also allows us to inquire into its political infrastructures. In political terms, assemblage thinking is partly driven by dissatisfaction with state-centric dominant ontologies, including reified units such as state, society, or capitalism, and the unilinear focus on state-centric politics over other forms of politics.⁶⁵ The assemblage perspective is therefore especially useful for understanding the politics of late-sovereign and late-capitalist data projects such as mass digitization. As we will see in part 2, the epistemic frame of sovereignty continues to offer an organizing frame for the constitution and regulation of mass digitization and the virtues associated with it (such as national representation and citizen engagement). However, at the same time, mass digitization projects are in direct correspondence with neoliberal values such as privatization, consumerism, globalization, and acceleration, and its technological features allow for a complete restructuring of the disciplinary spaces of libraries to form vaster and even global scales of integration and economic organization on a multinational stage.

Mass digitization is a concrete example of what cultural memory projects look like in a “late-sovereign” age, where globalization tests the political and symbolic authority of sovereign cultural memory politics to its limits, while sovereignty as an epistemic organizing principle for the politics of cultural memory nonetheless persists.⁶⁶ The politics of cultural memory, in particular those practiced by cultural heritage institutions, often still cling to fixed sovereign taxonomies and epistemic frameworks. This focus

is partly determined by their institutional anchoring in the framework of national cultural policies. In mass digitization, however, the formal political apparatus of cultural heritage institutions is adjoined by a politics that plays out in the margins: in lobbies, software industries, universities, social media, etc. Those evaluating mass digitization assemblages in macropolitical terms, that is, those who are concerned with political categories, will glean little of the real politics of mass digitization, since such politics at the margins would escape this analytic matrix.⁶⁷ Assemblage thinking, by contrast, allows us to acknowledge the political mechanisms of mass digitization beyond disciplinary regulatory models, in societies where “where forces ... not categories, clash.”⁶⁸

As Ian Hacking and many others have noted, the capacious usage of the notion of “politics” threatens to strip the word of meaning.⁶⁹ But talk of a politics of mass digitization is no conceptual gimmick, since what is taking place in the construction and practice of mass digitization assemblages plainly is political. The question, then, is how best to describe the politics at work in mass digitization assemblages. The answer advanced by the present volume is to think of the politics of mass digitization as “infrapolitics.”

The notion of infrapolitics has until now primarily and profoundly been advanced as a concept of hidden dissent or contestation (Scott, 1990).⁷⁰ This volume suggests shifting the lens to focus on a different kind of infrapolitics, however, one that not only takes the shape of resistance but also of maintenance and conformity, since the story of mass digitization is both the story of contestation *and* the politics of mundane and standard-seeking practices.⁷¹ The infrapolitics of mass digitization is, then, a kind of politics “premised not on a subject, but on the *infra*,” that is, the “underlying rules of the world,” organized around glocal infrastructures.⁷² The infrapolitics of mass digitization is the building and living of infrastructures, both as spaces of contestation and processes of naturalization.

Geoffrey Bowker and Susan Leigh Star have argued that the establishment of standards, categories, and infrastructures “should be recognized as the significant site of political and ethical work that they are.”⁷³ This applies not least in the construction and development of knowledge infrastructures such as mass digitization assemblages, structures that are upheld by increasingly complex sets of protocols and standards. Attaching “politics” to “infrastructure” endows the term—and hence mass digitization under this rubric—with a distinct organizational form that connects various stages

and levels of politics, as well as a distinct temporality that relates mass digitization to the forces and ideas of industrialization and globalization.

The notion of infrastructure has a surprisingly brief etymology. It first entered the French language in 1875 in relation to the excavation of railways.⁷⁴ Over the following decades, it primarily designated fixed installations designed to facilitate and foster mobility. It did not enter English vocabulary until 1927, and as late as 1951, the word was still described by English sources as “new” (OED).⁷⁵ When NATO adopted the term in the 1950s, it gained a military tinge. Since then, “infrastructure” has proliferated into ever more contexts and disciplines, becoming a “plastic word”⁷⁶ often used to signify any vital and widely shared human-constructed resource.⁷⁷

What makes infrastructures central for understanding the politics of mass digitization? Primarily, they are crucial to understanding how industrialism has affected the ways in which we organize and engage with knowledge, but the politics of infrastructures are also becoming increasingly significant in the late-sovereign, late-capitalist landscape.

The infrastructures of mass digitization mediate, combine, connect, and converge upon different institutions, social networks, and devices, augmenting the actors that take part in them with new agential possibilities by expanding the radius of their action, strengthening and prolonging the reach of their performance, and setting them free for other activities through their accelerating effects, time often reinvested in other infrastructures, such as, for instance, social media activities. The infrastructures of mass digitization also increase the demand for globalization and mobility, since they expand the radius of using/reading/working.

The infrastructures of mass digitization are thus media of politics and politics, at times visible and at others barely legible or felt, and home both to dissent as well as to standardizing measures. These include legal infrastructures such as copyright, privacy, and trade law; material infrastructures such as books, wires, scanners, screens, server parks, and shelving systems; disciplinary infrastructures such as metadata, knowledge organization, and standards; cultural infrastructures such as algorithms, searching, reading, and downloading; societal infrastructures such as the realms of the public and private, national and global. These infrastructures are, depending, both the prerequisites for and the results of interactions between the spatial, temporal, and social classes that take part in the construction of mass

digitization. The infrapolitics of mass digitization is thus geared toward both interoperability and standardization, as well as toward variation.⁷⁸

Often when thinking of infrastructures, we conceive of them in terms of durability and stability. Yet, while some infrastructures, such as railways and Internet cables, are fairly solid and rigid constructions, others—such as semantic links, time-limited contracts, and research projects—are more contingent entities which operate not as “fully coherent, deliberately engineered, end-to-end processes,” but rather as morphous contingent assemblages, as “ecologies or complex adaptive systems” consisting of “numerous systems, each with unique origins and goals, which are made to interoperate by means of standards, socket layers, social practices, norms, and individual behaviors that smooth out the connections among them.”⁷⁹ This contingency has direct implications for infrapolitics, which become equally flexible and adaptive. These characteristics endow mass digitization infrastructures with vulnerabilities but also with tremendous cultural power, allowing them to distribute agency, and to create and facilitate new forms of sociality and culture.

Building mass digitization infrastructures is a costly endeavor, and hence mass digitization infrastructures are often backed by public-private partnerships. Indeed infrastructures—and mass digitization infrastructures are no exceptions—are often so costly that a certain mixture of political or individual megalomania, state reach, and private capital is present in their construction.⁸⁰ This mixed foundation means that a lot of the political decisions regarding mass digitization literally take place *beneath* the radar of “the representative institutions of the political system of nation-states,” while also more or less aggressively filling out “gaps” in nation-state systems, and even creating transnational zones with their own policies.⁸¹ Hence the notion of “infra”: the infrapolitics of mass digitization hover at a frequency that lies *below* and beyond formal sovereign state apparatus, organized, as they are, around glocal—and often private or privatized—material and social infrastructures.

While distinct from the formalized sovereign political system, infrapolitical assemblages nevertheless often perform as late-sovereign actors by engaging in various forms of “sovereignty games.”⁸² Take Google, for instance, a private corporation that often defines itself as at odds with state practice, yet also often more or less informally meets with state leaders, engages in diplomatic discussions, and enters into agreements with state

agencies and local political councils. The infrapolitical forces of Google in these sovereignty games can on the one hand exert political pressure on states—for instance in the name of civic freedom—but in Google’s embrace of politics, its infrapolitical forces can on the other hand also squeeze the life out of existing parliamentary ways, promoting instead various forms of apolitical or libertarian modes of life. The infrapolitical apparatus thus stands apart from more formalized politics, not only in terms of political arena, but also the constraints that are placed upon them in the form, for instance, of public accountability.⁸³ What is described here can in general terms be called the infrapolitics of neoliberalism, whose scenery consists of lobby rooms, policy-making headquarters, financial zones, public-private spheres, and is populated by lobbyists, bureaucrats, lawyers, and CEOs.

But the infrapolitical dynamics of mass digitization also operate in more mundane and less obvious settings, such as software design offices and standardization agencies, and are enacted by engineers, statisticians, designers, and even users. Infrastructures are—increasingly—essential parts of our everyday lives, not only in mass digitization contexts, but in all walks of life, from file formats and software programs to converging transportation systems, payment systems, and knowledge infrastructures. Yet, what is most significant about the majority of infrapolitical institutions is that they are so mundane; if we notice them at all, they appear to us as boring “lists of numbers and technical specifications.”⁸⁴ And their maintenance and construction often occurs “behind the scenes.”⁸⁵ There is a politics to these naturalizing processes, since they influence and frame our moral, scientific, and aesthetic choices. This is to say that these kinds of infrapolitical activities often retire or withdraw into a kind of self-evidence in which the values, choices, and influences of infrastructures are taken for granted and accorded a kind of obviousness, which is universally accepted. It is therefore all the more “politically and ethically crucial”⁸⁶ to recognize the infrapolitics of mass digitization, not only as contestation and privatized power games, but also as a mode of existence that values professionalized standardization measures and mundane routines, not least because these infrapolitical modes of existence often outlast their material circumstances (“software outlasts hardware” as John Durham Peters notes).⁸⁷ In sum, infrastructures and the infrapolitics they produce yield subtle but significant world-making powers.

Power in Mass Digitization

If mass digitization is a product of a particular social configuration and political infrastructure, it is also, ultimately, a site and an instrument of power. In a sense, mass digitization is an event that stages a fundamental confrontation between state and corporate power, while pointing to the reconfigurations of both as they become increasingly embedded in digital infrastructures. For instance, such confrontation takes place at the negotiating table, where cultural heritage directors face the seductive and awe-inspiring riches of Silicon Valley, as well as its overwhelmingly intricate contractual layouts and its intimidating entourage of lawyers. Confrontation also takes place at the level of infrastructural ideology, in the meeting between twentieth-century standardization ideals and the playful and flexible network dynamics of the twenty-first century, as seen for instance in the conjunction of institutionally fixed taxonomies and algorithmic retrieval systems that include feedback mechanisms. And it takes place at the level of users, as they experience a gain in some powers and the loss of others in their identity transition from national patrons of cultural memory institutions to globalized users of mass digitization assemblages.

These transformations are partly the results of society's increasing reliance on network power and its effects. Political theorists Michael Hardt and Antonio Negri suggested almost two decades ago that among other things, global digital systems enabled a shift in power infrastructures from robust national economies and core industrial sectors to interactive networks and flexible accumulation, creating a "form of network power, which requires the wide collaboration of dominant nation-states, major corporations, supra-national economic and political institutions, various NGOs, media conglomerates and a series of other powers."⁸⁸ From this landscape, according to their argument, emerged a new system of power in which morphing networks took precedence over reliable blocs. Hardt and Negri's diagnosis was one of several similar arguments across the political spectrum that were formed within such a short interval that "the network" arguably became the "defining concept of our epoch."⁸⁹ Within this new epoch, the old centralized blocs of power crumbled to make room for new forms of decentralized "bastard" power phenomena, such as the extensive corporate/state mass surveillance systems revealed by Edward

Snowden and others, and new forms of human rights such as “the right to be forgotten,” a right for which a more appropriate name would be “the right to not be found by Google.”⁹⁰ Network power and network effects are therefore central to understanding how mass digitization assemblages operate, and why some mass digitization assemblages are more powerful than others.

The power dynamics we find in Google Books, for instance, are directly related to the ways in which digital technologies harness network effects: the power of Google Books grows exponentially as its network expands.⁹¹ Indeed, as Siva Vaidhyanathan noted in his critical work on Google’s role in society, what he referred to as the “Googlization of books” was ultimately deeply intertwined with the “Googlization of everything.”⁹² The networks of Google thus weren’t external to both the success and the challenges of Google, but deeply endemic to it, from portals and ranking systems to anchoring (elite) institutions, and so on. The better Google Books becomes at harnessing network effects, the more fundamental its influence is in the digital sphere. And Google Books is very good at harnessing digital network power. Indeed, Google Books reached its “tipping point” almost before it launched: it had by then already attracted so many stakeholders that its mere existence decreased the power of any competing entities—and the fact that its heavy user traffic is embedded in Google only strengthened its network effects. Google Books’s tipping point tells us little about its quality in an abstract sense: “tipping points” are more often attained by proprietary measures, lobbying, expansion, and most typically by a mixture of all of the above, than by sheer quality.⁹³ This explains not only the success of Google Books, but also its traction with even its critics: although Google Books was initially criticized heavily for its poor imagery and faulty metadata,⁹⁴ its possible harmful impact on the public sphere,⁹⁵ and later, over privacy concerns,⁹⁶ it had already created a power hub to which, although they could have navigated around it, masses of people were nevertheless increasingly drawn.

Network power is endemic not only to concrete digital networks, but also to globalization at large as a process that simultaneously gives rise to feelings of freedom of choice and loss of choice.⁹⁷ Mass digitization assemblages, and their globalization of knowledge infrastructures, thus crystalize the more general tendencies of globalization as a process in which people participate by choice, but not necessarily voluntarily; one in which we are

increasingly pushed into a game of social coordination, where common standards allow more effective coordination yet also entrap us in their pull for convergence. Standardization is therefore a key technique of network power: on the one hand, standardization is linked with globalization (and various neoliberal regimes) and the attendant widespread contraction of the state, while on the other hand, standardization implies a reconfiguration of everyday life.⁹⁸ Standards allow for both minute data analytics and overarching political systems that “govern at a distance.”⁹⁹ Standardization understood in this way is thus a mode of capturing, conceptualizing, and configuring reality, rather than simply an economic instrument or lubricant. In a sense, standardization could even be said to be habit forming: through standardization, “inventions become commonplace, novelties become mundane, and the local becomes universal.”¹⁰⁰

To be sure, standardization has long been a crucial tool of world-making power, spanning both the early and late-capitalist eras.¹⁰¹ “Standard time,” as John Durham Peters notes, “is a sine qua non for international capitalism.”¹⁰² Without the standardized infrastructure of time there would be no global transportation networks, no global trade channels, and no global communication networks. Indeed, globalization is premised on standardization processes.

What kind of standardization processes do we find, then, in mass digitization assemblages? Internet use alone involves direct engagement with hundreds of global standards, from Bluetooth to Wi-Fi standards, from protocol standards to file standards such as Word and MP4 and HTTP.¹⁰³ Moreover, mass digitization assemblages confront users with a series of additional standards, from cultural standards of tagging to technical standards of interoperability, such as the European Data Model (EDM) and Google’s schema.org, or legal standards such as copyright and privacy regulations. Yet, while these standards share affinities with the standardization processes of industrialization, in many respects they also deviate from them. Instead, we experience in mass digitization “a new form of standardization,”¹⁰⁴ in which differentiation and flexibility gain increasing influence without, however, dispensing with standardization processes.

Today’s standardization is increasingly coupled with demands for flexibility and interoperability. Flexibility, as Joyce Kolko has shown, is a term that gained traction in the 1970s, when it was employed to describe putative solutions to the problems of Fordism.¹⁰⁵ It was seen as an antidote to

Fordist “rigidity”—a serious offense in the neoliberal regime. Thus, while the digital networks underlying mass digitization are geared toward standardization and expansion, since “information technology rewards scale, but only to the extent that practices are standardized,”¹⁰⁶ they are also becoming increasingly flexible, since too-rigid standards hinder network effects, that is, the growth of additional networks. This is one reason why mass digitization assemblages increasingly and intentionally break down the so-called “silo” thinking of cultural memory institutions, and implement standard flexibility and interoperability to increase their range.¹⁰⁷ One area of such reconfiguration in mass digitization is the taxonomic field, where stable institutional taxonomic structures are converted to new flexible modes of knowledge organization like linked data.¹⁰⁸ Linked data can connect cultural memory artifacts as well as metadata in new ways, and the move from a cultural memory web of interlinked documents to a cultural memory web of interlinked data can potentially “amplify the impact of the work of libraries and archives.”¹⁰⁹ However, in order to work effectively, linked data demands standards and shared protocols.

Flexibility allows the user a freer range of actions, and thus potentially also the possibility of innovation. These affordances often translate into user freedom or empowerment. Yet flexibility does not necessarily equal fundamental user autonomy or control. On the contrary, flexibility is often achieved through decomposition, modularization, and black-boxing, allowing some components to remain stable while others are changed without implications for the rest of the system.¹¹⁰ These components are made “fluid” in the sense that they are dispersed of clear boundaries and allowed multiple identities, and in that they enable continuity and dissolution.

While these new flexible standard-setting mechanisms are often localized in national and subnational settings, they are also globalized systems “oriented towards global agendas and systems.”¹¹¹ Indeed, they are “glocal” configurations with digital networks at their cores. The increasing significance of these glocal configurations has not only cultural but also democratic consequences, since they often leave users powerless when it comes to influencing their cores.¹¹² This more fundamental problematic also pertains to mass digitization, a phenomenon that operates in an environment that constructs and encourages less Habermasian public spheres than “relations of sociability,” from which “aggregate outcomes emerge not from an act of collective decision-making, but through the accumulation

of decentralized, individual decisions that, taken together, nonetheless conduce to a circumstance that affects the entire group."¹¹³ For example, despite the flexibility Google Books allows us in terms of search and correlation, we have very little sway over its construction, even though we arguably influence its dynamics. The limitations of our influence on the cores of mass digitization assemblages have implications not only for how we conceive of institutional power, but also for our own power within these matrixes.

II Mapping Mass Digitization

2 The Trials, Tribulations, and Transformations of Google Books

Introduction

In a 2004 article in the cultural theory journal *Critical Inquiry*, book historian Roger Chartier argued that the electronic world had created a triple rupture in the world of text: by providing new techniques for inscribing and disseminating the written word, by inspiring new relationships with texts, and by imposing new forms of organization onto them. Indeed, Chartier foresaw that “the originality and the importance of the digital revolution must therefore not be underestimated insofar as it forces the contemporary reader to abandon—consciously or not—the various legacies that formed it.”¹ Chartier’s premonition was inspired by the ripples that digitization was already spreading across the sea of texts. People were increasingly writing and distributing electronically, interacting with texts in new ways, and operating and implementing new textual economies.² These textual transformations gave rise to a range of emotional reactions in readers and publishers, from catastrophizing attitudes and pessimism about “the end of the book” to the triumphalist mythologizing of liquid virtual books that were shedding their analog ties like butterflies shedding their cocoons.

The most widely publicized mass digitization project to date, Google Books, precipitated the entire emotional spectrum that could arise from these textual transversals: from fears that control over culture was slipping from authors and publishers into the hands of large tech companies, to hopeful ideas about the democratizing potential of bringing knowledge that was once locked up in dusty tomes at places like Harvard and Stanford, and to a utopian mythologizing of the transcendent potential of mass digitization. Moreover, Google Books also affected legal and professional

transformations of the infrastructural set-up of the book, creating new precedents and a new professional ethos. The cultural, legal, and political significance of Google Books, whether positive or negative, not only emphasizes its fundamental role in shaping current knowledge landscapes, it also allows us to see Google Books as a prism that reflects more general political tendencies toward globalization, privatization, and digitization, such as modulations in institutional infrastructures, legal landscapes, and aesthetic and political conventions. But how did the unlikely marriage between a tech company and cultural memory institutions even come about? Who drove it forward, and around and within which infrastructures? And what kind of cultural memory politics did it produce? The following sections of this chapter will address some of these problematics.

The New Librarians

It was in the midst of a turbulent restructuring of the world of text, in October 2004 at the Frankfurt International Book Fair, that Larry Page and Sergey Brin of Google announced the launch of Google Print, a cooperation between Google and leading Anglophone publishers. Google Print, which later became Google Partner Program, would significantly alter the landscape and experience of cultural memory, as well as its regulatory infrastructures. A decade later, the traditional practices of reading, and the guardianship of text and cultural works, had acquired entirely new meanings. In October 2004, however, the publishing world was still unaware of Google's pending influence on the institutional world of cultural memory. Indeed, at that time, Amazon's mounting dominance in the field of books, which began a decade earlier in 1995, appeared to pose much more significant implications. The majority of publishers therefore greeted Google's plans in Frankfurt as a welcome alternative to Jeff Bezos's growing online behemoth.

Larry Page and Sergey Brin withheld a few details from their announcement at Frankfurt, however; Google's digitization plans would involve not only cooperation with publishers, but also with libraries. As such, what would later become Google Books would in fact consist of two separate, yet interrelated, programs: Google Print (which would later become Google Partner Program) and Google Library Project. In all secrecy, Google had for many months prior to the Frankfurt Book Fair worked with select libraries

in the US and the UK to digitize their holdings. And in December 2004 the true scope of Google's mass digitization plans were revealed: what Page and Brin were building was the foundation of a groundbreaking cultural memory archive, inspired by the myth of Alexandria.³ The invocation of Alexandria situated the nascent Google Books project in a cultural schema that historicized the project as a utopian, even moral and idealist, project that could finally, thanks to technology, exceed existing human constraints—legal, political, and physical.⁴

Google's utopian discourse was not foreign to mass digitization enthusiasts. Indeed, it was the *langue du jour* underpinning most large-scale digitization projects, a discourse nurtured and influenced by the seemingly borderless infrastructure of the web itself (which was often referred to in universalizing terms).⁵ Yet, while the universalizing discourse of mass digitization was familiar, it had until then seemed like aspirational talk at best, and strategic policy talk in the face of limited public funding, complex copyright landscapes, and lumbering infrastructures, at worst. Google, however, faced the task with a fresh attitude of determination and a will to disrupt, as well as a very different form of leverage in terms of infrastructural set-up. Google was already the world's preferred search engine, having mastered the tactical skill of navigating its users through increasingly complex information landscapes on the web, and harvesting their metadata in the process to continuously improve Google's feedback systems. Essentially ever-larger amounts of information (understood here as "users") were passing through Google's crawling engines, and as the masses of information in Google's server parks grew, so did their computational power. Google Books, then, as opposed to most existing digitization projects, which were conceived mainly in terms of "access," was embedded in the larger system of Google that understood the power and value of "feedback," collecting information and entering it into feedback loops between users, machines, and engineers. Google also understood that information power didn't necessarily lie in owning all the information they gave access to, but rather in controlling the informational processes themselves.

Yet, despite Google's advances in information seeking behaviors, the idea of Google Books appeared as an odd marriage. Why was a private company in Silicon Valley, working in the futuristic and accelerating world of software and fluid information streams, intent on partnering up with the slow-paced world of cultural memory institutions, traditionally more concerned with

the past? Despite the apparent clash of temporal and cultural regimes, however, Google was in fact returning home to its point of inception. Google was born of a research project titled the Stanford Integrated Digital Library Project, which was part of the NSF's Digital Libraries Initiative (1994–1999). Larry Page and Sergey Brin were students then, working on the Stanford component of this project, intending to develop the base technologies required to overcome the most critical barriers to effective digital libraries, of which there were many.⁶ Page's and Brin's specific project, titled Google, was presented as a technical solution to the increasing amount of information on the World Wide Web.⁷ At Stanford, Larry Page also tried to facilitate a serious discussion of mass digitization at Stanford, and of whether or not it was feasible. But his ideas received little support, and he was forced to leave the idea on the drawing board in favor of developing search technologies.⁸

In September 1998, Sergey Brin and Larry Page left the library project to found Google as a company and became immersed in search engine technologies. However, a few years later, Page resuscitated the idea of mass digitization as a part of their larger self-professed goal to change the world of information by increasing access, scaling the amount of information available, and improving computational power. They convinced Eric Schmidt, the new CEO of Google, that the mass digitization of cultural works made sense not only from an information perspective, but also from a business perspective, since the vast amounts of information Google could extract from books would improve Google's ability to deliver information that was hitherto lacking, and this new content would eventually also result in an increase in traffic and clicks on ads.⁹

The Scaling Techniques of Mass Digitization

A series of experiments followed on how to best approach the daunting task. The emergence and decay of these experiments highlight the ways in which mass digitization assemblages consist not only of thoughts, ideals, and materials, but also a series of cultural techniques that entwine temporality, materiality, and even corporeality. This perspective on mass digitization emphasizes the mixed nature of mass digitization assemblages: what at first glance appears as a relatively straightforward story about new technical inventions, at a closer look emerges as complex entanglements of human

and nonhuman actors, with implications not only for how we approach it as a legal-technical entity but also an infrapolitical phenomenon. As the following section shows, attending to the complex cultural techniques of mass digitization (its “how”) enables us to see that its “minor” techniques are not excluded from or irrelevant to, but rather are endemic to, larger questions of the infrapolitics of digital capitalism. Thus, Google’s simple technique of scaling scanning to make the digitization processes go faster becomes entangled in the creation of new habits and techniques of acceleration and rationalization that tie in with the politics of digital culture and digital devices. The industrial scaling of mass digitization becomes a crucial part of the industrial apparatus of big data, which provide new modes of inscription for both individuals and digital industries that in turn can be capitalized on via data-mining, just as it raises questions of digital labor and copyright.

Yet, what kinds of scaling techniques—and what kinds of investments—Google would have to leverage to achieve its initial goals were still unclear to Google in those early years. Larry Page and co-worker Marissa Mayer therefore began to experiment with the best ways to proceed. First, they created a makeshift scanning device, whereby Marissa Mayer would turn the page and Larry Page would click the shutter of the camera, guided by the pace of a metronome.¹⁰ These initial mass digitization experiments signaled the industrial nature of the mass digitization process, providing a metronomic rhythm governed by the implacable regularity of the machine, in addition to the temporal horizon of eternity in cultural memory institutions (or at least of material decay).¹¹ After some experimentation with scale and time, Google bought a consignment of books from a second-hand book store in Arizona. They scanned them and subsequently experimented with how to best index these works not only by using information from the book, but also by pulling data about the books from various other sources on the web. These extractions allowed them to calculate a work’s relevance and importance, for instance by looking at the number of times it had been referred to.¹²

In 2004 Google was also granted patent rights to a scanner that would be able to scan the pages of works without destroying them, and which would make them searchable thanks to sophisticated 3D scanning and complex algorithms.¹³ Google’s new scanner used infrared camera technology that detected the three-dimensional shape and angle of book pages when the

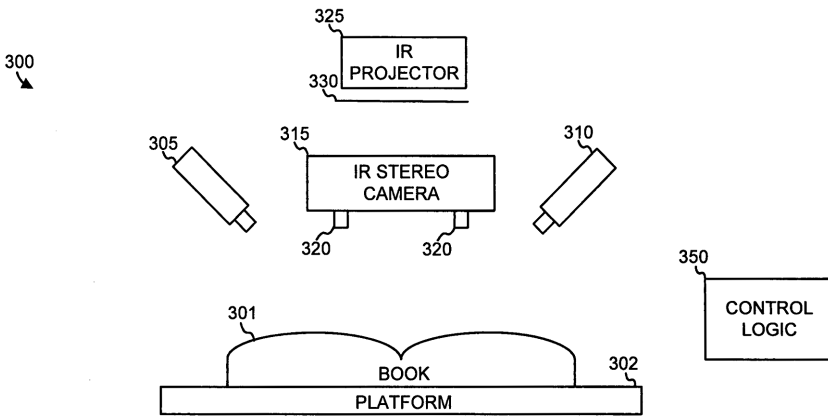


Figure 2.1

François-Marie Lefevre and Marin Saric. "Detection of grooves in scanned images." U.S. Patent 7508978B1. Assigned to Google LLC.

book was placed in the scanner. The information from the book was then transmitted to Optical Character Recognition (OCR), which adjusted image focus and allowed the OCR software to read images of curved surfaces more accurately.

These new scanning technologies allowed Google to unsettle the fixed content of cultural works on an industrial scale and enter them into new distribution systems. The untethering and circulation of text already existed, of course, but now text would mutate on an industrial scale, bringing into coexistence a multiplicity of archiving modes and textual accumulation. Indeed, Google's systematic scaling-up of already existing technologies on an industrial and accelerated scale posed a new paradigm in mass digitization, to a much larger extent than, for instance, inventions of new technologies.¹⁴ Thus, while Google's new book scanners did expand the possibilities of capturing information, Google couldn't solve the problem of automating the process of turning the pages of the books. For that they had to hire human scanners who were asked to manually turn pages. The work of these human scanners was largely invisible to the public, who could only see the books magically appearing online as the digital archive accumulated. The scanners nevertheless left ghostly traces, in the form of scanning errors such as pink fingers and missing and crumbled pages—visual traces that underlined the historically crucial role of human labor in

industrializing and automating processes.¹⁵ Indeed, the question of how to solve human errors in the book scanning process led to a series of inventive systems, such as the patent granted to Google in 2009 (filed in 2003), which describes a system that would minimize scanning errors with the help of music.¹⁶ Later, Google open sourced plans for a book scanner named “Linear Book Scanner” that would turn the pages automatically with the help of a vacuum cleaner and a cleverly designed sheet metal structure, after passing them over two image sensors taken from a desktop scanner.¹⁷

Eventually, after much experimentation, Google consolidated its mass digitization efforts in collaboration with select libraries.¹⁸ While some institutions immediately and enthusiastically welcomed Google’s aspirations as aligning with their own mission to improve access to information, others were more hesitant, an institutional vacillation that hinted ominously at controversy to come. Some libraries, such as the University of Michigan, greeted the initiative with enthusiasm, whereas others, such as the Library of Congress, saw a red flag pop up: copyright, one of the most fundamental elements in the rights of texts and authors.¹⁹ The Library of Congress questioned whether it was legal to scan and index books without a rights holder’s permission. Google, in response, argued that it was within the fair use provisions of the law, but the argument was speculative in so far as there was no precedent for what Google was going to do. While some universities agreed with Google’s views on copyright and shared its desire to disrupt existing copyright practices, others allowed Google to make digital copies of their holdings (a precondition for creating an index of it). Hence, some libraries gave full access, others allowed only the scanning of books in the public domain (published before 1923), and still others denied access altogether. While the reticence of libraries was scattered, it was also a precursor of a much more zealous resistance to Google Books, an opposition that was mounted by powerful voices in the cultural world, namely publishers and authors, and other commercial infrastructures of cultural memory.

While Google’s announcement of its cooperation with publishers at the Frankfurt Book Fair was received without drama—even welcomed by many—the announcement of its cooperation with libraries a few months later caused a commercial uproar. The most publicized point of contestation was the fact that Google was now not only displaying books in cooperation with publishers, but also building a library of its own, without

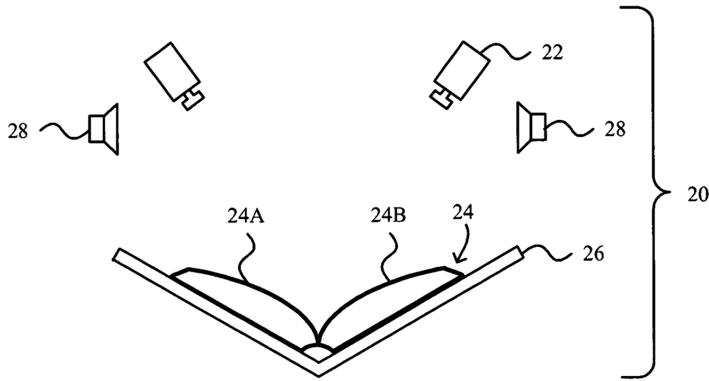


FIG. 1

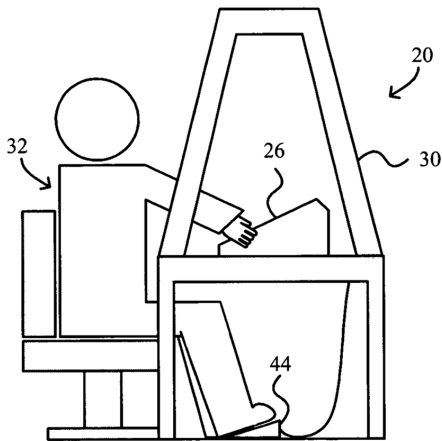


FIG. 2

Figure 2.2

Joseph K. O'Sullivan, Alexander Proudfoot, and Christopher R. Uhlik. "Pacing and error monitoring of manual page turning operator." U.S. Patent 7619784B1. Assigned to Google LLC, Google Technology Holdings LLC.

remunerating publishers and authors. Why would readers buy books if they could read them free online? Moreover, the Authors Guild worried that Google's digital library would increase the risk of piracy. At a deeper level, the case also emphasized authors' and publishers' desire to retain control over their copyrighted works in the face of the threat that the Library Project (unlike the Partner Program) was posing: Google was digitizing without the copyright holder's permission. Thus, to them, the Library Project fundamentally threatened their copyrights and, on a more fundamental level, existing copyright systems. Both factors, they argued, would make book buying a superfluous activity.²⁰ The harsher criticisms framed Google Books as a book thief rather than as a global philanthropist.²¹ Google, on its behalf, launched a defense of their actions based on the notion of "fair use," which as the following section shows, eventually became the fundamental legal question.

Infrastructural Transformations

Google Books became the symbol of the painful confusion and territorial battles that marred the publishing world as it underwent a transformation from analog to digital. The mounting and diverse opposition to Google Books was thus not an isolated affair, but rather a persistent symptom—increasingly loud stress signals emitting from the infrastructural joints of the analog realm of books as it buckled under the strain of digital logic. As media theorist John Durham Peters (drawing on media theorist Harold Innis) notes, the history of media is also an "occupational history" that tells the tales of craftspeople mastering medium-specific skills tactically battling for monopolies of knowledge and guarding their access.²² And in the occupational history of Google Books, the craftspeople of the printed book were being challenged by a new breed of artificers who were excelling not so much in how to print, which book sellers to negotiate with, or how to sell books to people, but rather in the medium-specific tactical skills of the digital, such as building software and devising search technologies, skills they were leveraging to their own gain to create new "monopolies of knowledge" in the process.

As previously mentioned, the concerns expressed by publishers and authors in regards to remuneration was accompanied by a more abstract sense of a loss of control over their works and how this loss of control

would affect the copyrights. These concerns did not arise out of thin air, but were part of a more general discourse on digital information as something that *cannot* be secured and controlled in the same way as analog commodities can. Indeed, it seemed that authors and publishers were part of a world entirely different from Google Books: while publishers and authors were still living in and defending a “regime of scarcity,”²³ Google Books, by contrast, was busy building a “realm of plenitude and infinite replenishment.” As such, the clash between the traditional infrastructures of the analog book and the new infrastructures of Google Books was symptomatic of the underlying radical reorganization of information from a state of trade and exchange to a state of constant transmission and contagion.²⁴

Foregrounding the fair use defense²⁵, Google argued that the public benefits of scanning outweighed the negative consequences for authors.²⁶ Influential legal scholars such as Lawrence Lessig, among others, supported this argument, suggesting that inclusion in a search engine in a way that does not erode the value of the book was of such societal importance that it should be deemed legal.²⁷ The copyright owners, however, insisted that the burden should be on Google to request permission to scan each work.²⁸

Google and copyright owners reached a proposed settlement on October 28, 2008. The proposal would allow Google not only to continue its scanning activities and to show free snippets online, but would also give Google exclusive rights to sell digital copies of out-of-print books. In return, Google would provide all libraries in the United States with one free subscription to the digital database, but Google could also sell additional subscriptions. Moreover, Google was to pay \$125 million, part of which would go to the construction of a Book Rights Registry that identified rights holders and handled payments to lawyers.²⁹ Yet before the settlement was even formally treated, a mounting opposition to it was launched in public.

The proposed settlement was received with harsh words, for instance by Internet archivist Brewster Kahle and legal scholar Lawrence Lessig, who opposed the settlement with words ranging from “insanity” to “cultural asphyxiation” and “information monopoly.”³⁰ Privacy proponents also spoke out against Google Books, bringing attention to the implications of Google being able to follow and track reading habits, among other things.³¹ The organization Privacy Authors, including writers such as Jonathan Lethem, Bruce Schneier, and Michael Chabon, and publishers, argued that

although Google Books was an “extremely exciting” project, it failed in its current form to protect the privacy of readers, thus creating a “real risk of disclosure” of sensitive information to “prying governmental entities and private litigants,” potentially giving rise to a “chilling effect,” hurting not only readers but also authors and publishers, not least those writing about sensitive or controversial topics.³² The Association of Libraries also raised a set of concerns, such as the cost of library subscriptions and privacy.³³ And most predictably, companies such as Amazon and Microsoft, who also had a stake in mass digitization, opposed the settlement; Microsoft even funded some nuanced research efforts into its implications.³⁴ Finally, and most damningly, the Department of Justice decided to get involved with an antitrust argument.

By this point, opposition to the Google Books project, as it was outlined in the proposed settlement, wasn't only motivated by commercial concerns; it was now also motivated by a public that framed Google's mass digitization project as a parasitical threat to the public sphere itself. The framing of Google as a potential menace was a jarring image that stood in stark contrast to Larry Page's and Sergey Brin's philanthropic attitudes and to Google's famous “Don't be evil” slogan. The public reaction thus signaled a change in Google's reputation as the company metamorphosed in the public eye from a small underdog company to a multinational corporation with a near-monopoly in the search industry. Google's initially inspiring approach to information as a realm of plenitude now appeared in the public view more similar to the actions of megalomaniac land-grabbers.

Google, however, while maintaining its universalizing mission regarding information, also countered the accusations of monopoly building, arguing that potential competitors could just step up, since nothing in the agreements entered into by the libraries and Google “precludes any other company or organization from pursuing their own similar effort.”³⁵ Nevertheless Judge Denny Chin denied the settlement in March 2011 with the following statement: “The question presented is whether the ASA is fair, adequate, and reasonable. I conclude that it is not.”³⁶ Google left the proposed settlement behind, and appealed the decision of their initial case with new amicus briefs focusing on their argument that book scanning was fair use. They argued that they were not demanding exclusivity on the information they scanned, that they didn't prohibit other actors from

digitizing the works they were digitizing, and that their main goal was to enrich the public sphere with more information, not to build an information monopoly. In July 2013 Judge Denny Chin issued a new opinion confirming that Google Books was indeed fair use.³⁷ Chin's opinion was later consolidated in a major victory for Google in 2015 when Judge Pierre Leval in the Second Circuit Court legalized Google Books with the words "Google's unauthorized digitizing of copyright-protected works, creation of a search functionality, and display of snippets from those works are non-infringing fair uses."³⁸ Leval's decision marked a new direction, not only for Google Books, but also for mass digitization in general, as it signaled a shift in cultural expectations about what it means to experience and disseminate cultural artifacts.

Once again, the story of Google Books took a new turn. What was first presented as a gift to cultural memory institutions and the public, and later as theft from and threat to these same entities, on closer inspection revealed itself as a much more complex circulatory system of expectations, promises, risks, and blame. Google Books thus instigated a dynamic and forceful connection between Google and cultural memory institutions, where the roles of giver and receiver, and the first giver and second giver/returner, were difficult to decode. Indeed, the binding nature of the relationship between Google Books and cultural memory institutions proved to be much more complex than the simple physical exchange of books and digital files. As the next section outlines, this complex system of cultural production was held together by contractual arrangement—central joints, as it were, connecting data and works, public and private, local and global, in increasingly complex ways. For Google Books, these contractual relations appear as the connective tissues that make these assemblages possible, and which are therefore fundamental to their affective dimensions.

The Infrapolitics of Contract

In common parlance a contract is a legal tool that formalizes a "mutual agreement between two or more parties that something shall be done or forborne by one or both," often enforceable by law.³⁹ Contractual systems emerged with the medieval merchant regime, and later evolved with classical liberalism into an ideological revolt against paternalist systems as nothing less than freedom, a legal construct that could destroy the sentimental

bonds of personal dependence.⁴⁰ As the classic liberal social scientist William Graham Sumner argued, “[c]ontract ... is rational ... realistic, cold, and matter-of-fact.” The rational nature of contracts also affected their temporality, since a contract endures only “so long as the reason for it endures,” and their spatiality, relegating any form of sentiment from the public sphere to “the sphere of private and personal relations.”⁴¹

Sentiments prevailed, however, as the contracts tying together Google and cultural memory institutions emerged. Indeed, public and professional evaluations of the agreements often took an affective, even sexualized, form. The economist Paul Courant situated libraries “in bed with Google”⁴²; library consultant and media experts Jeff Ubois and Peter B. Kaufman recounted *how* they got in bed with Google—“[w]e were approached singly, charmed in confidence, the stranger was beguiling, and we embraced”⁴³; communication scholar Evelyn Bottando announced that “libraries not only got in bed with Google. They got married”⁴⁴; and librarian Jessamyn West finally pondered on the relationship ruins, “[s]till not sure, after all that, how we got this all so wrong. Didn’t we both want the same thing? Maybe it really wasn’t us, it was them. Most days it’s hard to remember what we saw in Google. Why did we think we’d make good partners?”⁴⁵

The evaluative discourse around Google Books dispels the idea of contracts as dispassionate transactions for services and labor, showing rather that contracts are infrapolitical apparatuses that give rise to emotions and affect; and that, moreover, they are systems of doctrines, relations, and social artifacts that organize around specific ideologies, temporalities, materialities, and techniques.⁴⁶ First and foremost, contracts give rise to new kinds of infrastructures in the field of cultural memory: they mediate, connect, and converge cultural memory institutions globally, giving rise to new institutional networks, in some cases increasing globalization and mobility for both users and objects, and in other cases restricting the same. The Google Books contracts display both technical and symbolic aspects: as technical artifacts they establish intricate frameworks of procedures, commitments, rights, and incentives for governing the transactions of cultural memory artifacts and their digitized copies. As symbolic artifacts they evoke normative principles, expressing different measures of good will toward libraries, but also—as all contracts do—introduce the possibility of distrust, conflict and betrayal.⁴⁷

Despite their centrality to mass digitization assemblages, and although some of them have been made available to the public,⁴⁸ the content of these particular contracts still suffer from the epistemic gap incurred in practical and symbolic form by Google's Agreements and Non-Disclosure Agreements (NDA), a kind of agreement most libraries are required to sign when entering the agreement. Like all contracts, the individual contracts signed by the partnership libraries vary in nature and have different implications. While many of Google's agreements may be publically available, they have often only been made public through requests and transparency mechanisms such as the Freedom of Information Act. As the Open Rights Alliance notes in their publication of the agreement entered between the British Library and Google, "We asked the British Library for a copy of the agreement with Google, which was not uploaded to their transparency website with other similar contracts, as it didn't involve monetary exchange. This may be a loophole transparency activists want to look at. After some toing and froing with the Freedom of Information Act we got a copy."⁴⁹

While the culture of contractual secrecy is native to the business world, with its safeguarding of business processes, and is easily navigated by business partners, it is often opposed to the ethos of state-subsidized cultural institutions who "draw their financial and moral support from a public that expects transparency in their activities, ranging from their materials acquisitions to their business deals."⁵⁰ For these reasons, library organizations have recommended that nondisclosure agreements should be avoided if possible, and minimized if they are necessary.⁵¹ Google, in response, noted on its website that: "[t]hough not all of the library contracts have been made public, we can say that all of them are non-exclusive, meaning that all of our library partners are free to continue their own scanning projects or work with others while they work with Google to digitize their books."⁵²

Regardless of their contractual content and later publication, the contracts are a vital instrument in Google's broader management of visibility. As Mikkel Flyverbom, Clare Birchall, and others have argued, this practice of visibility management—which they define as "the many ways in which organizations seek to curate and control their presence, relations, and comprehension vis-à-vis their surroundings" through practices of transparency, secrecy, opacity, surveillance, and disclosure—is in the digital age a complex issue closely tied to the question of governance and power. While each

publication act may serve to create an uncomplicated picture of transparency, it nevertheless happens in a paradoxical global regulatory environment that on the one hand encourages “sunshine” laws that demand that governments, corporations, and civil-sector organizations provide access to information, yet on the other hand also harbors regulatory agencies that seek mechanisms and rules by which to keep information hidden. Thus, as Flyverbom et al. conclude, the “everyday practices of organizing invariably implicate visibility management,” whose valences are “attached to transparency and opacity” that are not simple and straightforward, but rather remain “dependent upon the actor, the context, and the purpose of organizations and individuals.”⁵³

Steven Levy recounts how Google began its scanning operations in “near-total stealth,” a “cloak-and-dagger” approach that stood in contrast to Google’s public promotion of transparency as a new mode of existence. As Levy argues, “[t]he secrecy was yet another expression of the paradox of a company that sometimes embraced transparency and other times seemed to model itself on the NSA.”⁵⁴ Yet, while secrecy practices may have suited some of Google’s operations, they sit much more uneasily with their book scanning programs: “If Google had a more efficient way to scan books, sharing the improved techniques could benefit the company in the long run—inevitably, much of the output would find its way onto the web, bolstering Google’s indexes. But in this case, paranoia and a focus on short-term gain kept the machines under wraps.”⁵⁵ The nondisclosure agreements show that while boundaries may be blurred between Google Books and libraries, we may still identify different regulatory models and modes of existence within their networks, including the explicit *library ethos* (in the Weberian sense of the term) of public access, not only to the front end but also to some areas of the back end, and the business world’s secrecy practices.⁵⁶

Entering into a mass digitization public-private partnership (PPP) with a corporation such as Google is thus not only a logical and pragmatic next step for cultural memory institutions, it is also a political step. As already noted, Google Books, through its embedding in Google, injects cultural memory objects into new economic and cultural infrastructures. These infrastructures are governed less by the hierarchical world of curators, historians, and politicians, and more by feedback networks of tech companies, users, and algorithms. Moreover, they forge ever closer connections

to data-driven market logics, where computational rather than representational power counts. Mass digitization PPPs such as Google Books are thus also symptoms of a much more pervasive infrapolitical situation, in which cultural memory institutions are increasingly forced to alter their identities from public caretakers of cultural heritage to economic actors in the EU internal market, controlled by the framework of competition law, time-limited contracts, and rules on state aid.⁵⁷ Moreover, mastering the rules of these new infrastructures is not necessarily an easy feat for public institutions.⁵⁸ Thus, while Google claims to hold a core commitment regarding free digital access to information, and while its financial apparatus could be construed as making Google an eligible partner in accordance with the EU's policy objectives toward furthering public-private partnerships in Europe,⁵⁹ it is nevertheless, as legal scholar Maurizio Borghi notes, relevant to take into account Google's previous monopoly-building history.⁶⁰

The Politics of Google Books

A final aspect of Google Books relates to the universal aspiration of Google Books's collection, its infrapolitics, and what it empirically produces in territorial terms. As this chapter's previous sections have outlined, it was an aspiration of Google Books to transcend the cultural and political limitations of physical cultural memory collections by gathering the written material of cultural memory institutions into one massive digitized collection. Yet, while the collection spans millions of works in hundreds of languages from hundreds of countries,⁶¹ it is also clear that even large-scale mass digitization processes still entail procedures of selection on multiple levels from libraries to works. These decisions produce a political reality that in some respects reproduces and accentuates the existing politics of cultural memory institutions in terms of territorial and class-based representations, and in other respects give rise to new forms of cultural memory politics that part ways with the political regimes of traditional curatorial apparatuses.

One obvious area in which to examine the politics produced by the Google Books assemblage is in the selection of libraries that Google chooses to partner with.⁶² While the full list of Google Books partners is not disclosed on Google's own webpage, it is clear from the available list that, up to now, Google Books has mainly partnered with "great libraries," such

as elite university libraries and national libraries. The rationale for choosing these libraries has no doubt been to partner up with cultural memory institutions that preside over as much material as possible, and which are therefore able to provide more pieces of the puzzle than, say, a small-town public library that only presides over a fraction of their collections. Yet, while these libraries provide Google Books with an impressive and extensive collection of rare and valuable artifacts that give the impression of a near-universal collection, they nevertheless also contain epistemological and historical gaps. Historian and digital humanist Andrew Prescott notes, for example, the limited collections of literature written by workers and other lower-class people in the early eighteenth century in elite libraries. This institutional lack creates a pre-filtered collection in Google Books, favoring “[t]hose writers of working class origins who had a success story to report, who had become distinguished statesmen, successful businessmen, religious leaders and so on,” that is, the people who were “able to find commercial publishers who were interested in their story.”⁶³ Google’s decision to partner with elite libraries thus inadvertently reproduces the class-based biases of analog cultural memory institutions.

In addition to the reproduction of analog class-based bias in its digital collection, the Google Books corpus also displays a genre bias, veering heavily toward scientific publications. As mathematicians Eitan Pechenik et al. show, the contents of the Google Books corpus in the period of the 1900s is “increasingly dominated by scientific publications rather than popular works,” and “even the first data set specifically labeled as fiction appears to be saturated with medical literature.”⁶⁴ The fact that Google Books is constellated in such a manner thus challenges a “vast majority of existing claims drawn from the Google Books corpus,” just as it points to the need “to fully characterize the dynamics of the corpus before using these data sets to draw broad conclusions about cultural and linguistic evolution.”⁶⁵

Last but not least, Google Books’s collection still bespeaks its beginnings: it still primarily covers Anglophone ground. There is hardly any literature that reviews the geographic scope in Google Books, but existing work does suggest that Google is still heavily oriented toward US-based libraries.⁶⁶ This orientation does not necessarily give rise to an Anglophone linguistic hegemony, as some have feared, since many of the Anglophone libraries hold considerable collections of foreign language books. But it does invariably

limit its collections to the works in foreign languages that the elite libraries deemed worthy of preserving. The gaps and biases of Google Books reveal it to be less of a universal and monolithic collection, and more of an impressive, but also specific and contingent, assemblage of works, texts, and relations that is determined by the relations Google Books has entered into in terms of class, discipline, and geographical scope.

Google Books is not only the result of selection processes on the level of partnering institutions, but also on the level of organizational infrastructure. While the infrastructures of Google Books in fact depart from those of its parent company in many regards to avoid copyright infringement charges, there is little doubt, however, that people working actively on Google's digitization activities (included here are both users and Google employees) are also globally distributed in networked constellations. The central organization for cultural digitization, the Google Cultural Institute, is located in Paris, France. Yet the people affiliated with this hub are working across several countries. Moreover, people working on various aspects of Google Books, from marketing to language technology, to software developments and manual scanning processes, are dispersed across the globe. And it is perhaps in this way that we tend to think of Google in general—as a networked global company—and for good reasons. Google has been operating internationally almost for as long as it has been around. It has offices in countries all over the globe, and works in numerous languages. Today it is one of the most important global information institutions, and as more and more people turn to Google for its services, Google also increasingly reflects them—indeed they enter into a complex cognitive feedback mechanism system. Google depends on the growing diversity of its “inhabitants” and on its financial and cultural leverage on a global scale, and to this effect it is continuously fine-tuning its glocalization strategies, blending the universal and the particular. This glocal strategy does not necessarily create a universal company, however; it would be more correct to say that Google's glocality brings the globe to Google, redefining it as an “American” company.⁶⁷ Hence, while there is little doubt that Google, and in effect Google Books, increasingly tailors to specific consumers,⁶⁸ and that this tailoring allows for a more complex global representation generated by feedback systems, Google's core nevertheless remains lodged on American soil. This is underlined by the fact that Google Books still effectively belongs to US jurisdiction.⁶⁹ Google Books is thus on the one hand a globalized company

in terms of both content and institutional framework; yet it also remains an *American* multinational corporation, constrained by US regulation and social standards, and ultimately reinforcing the capacities of the American state. While Google Books operates as a networked glocal project with universal aspirations, then, it also remains fenced in by its legal and cultural apparatuses.

In sum, just as a country's regulatory and political apparatus affects the politics of its cultural memory institutions in the analog world, so is the politics of Google Books co-determined by the operations of Google. Thus, curatorial choices are made not only on the basis of content, but also of the location of server parks, existing company units, lobbying efforts, public policy concerns, and so on. And the institutional identity of Google Books is profoundly late-sovereign in this regard: on one hand it thrives on and operates with horizontal network formations; on the other, it still takes into account and has to operate with, and around, sovereign epistemologies and political apparatuses. These vertical and horizontal lines ultimately rewire the politics of cultural memory, shifting the stakes from sovereign territorial possessions to more functional, complex, and effective means of control.