POP
PUBLIC.
OPEN.
PARTICIPATORY.

Pop!　　Issue Nº1　　Issue Nº2　　About

Public-Private Partnerships and the Digitization of the Textual and Cultural Record

TULLY BARNETT

Flinders University
tully.barnett@flinders.edu.au

october 31, 2020

Introduction

On 19 March 2020, just over a week after the World Health Organization categorized COVID-19 a pandemic and as the first set of nations beyond China were instituting restrictive lockdowns on public and commercial activities, the Internet Archive launched its National Emergency Library. Describing the project as a "temporary collection of books that supports emergency remote teaching, research activities, independent scholarship, and intellectual stimulation while universities, schools, training centers, and libraries are closed," the National Emergency Library sought to make a positive, material, and public intervention in the lives of people experiencing the difficulties of lockdown or "shelter-in-place" orders and cut off from many of their usual sources of books (Internet Archive 2020). Unlike other cultural organizations such as art galleries, museums, and theatre companies, many of which put new collections online for the public to access during the lockdown (Wilson-Barnao 2020), the Internet Archive did not add any new material to its publicly available holdings, nor make any new works available for the first time. Rather, it made its extant collections available in a different way. What was controversial here is that many of the works in the National Emergency Library were still under copyright. The Internet Archive did not hold the copyright for the material it made available, nor did it pay a royalty fee or contribute to a public lending rights scheme of the sort in action in 28 countries around the world

Abstract:

A range of private-public partnerships, volunteer and crowdsourced labour, local, national, international and commercial agendas, and competing copyright systems intersect with the important but complicated project of digitization of the world's cultural material. These agendas are sometimes open, social, and collaborative (Edmond 2016; Liu 2016) but are also often closed,

(but not in the United States). Rather, what the National Emergency Library project did was change the conditions of access to the 1.4 million works in the Internet Archive's Open Library by revising the terms of the controlled digital lending, or "one-in-one-out," approach to ebook lending that limits the number of people who can access a copy at any one time and uses a waiting list system to manage the process, operating much as a physical library would with its limited collection of physical copies of books. Libraries internationally have been using variations of controlled digital lending to manage the loaning of ebook collections as a way of handling publishers' concerns about the risks of ebooks in library collections to publishers' revenues. The Internet Archive also makes the case that the unprecedented conditions of a pandemic require new approaches to information access. This is evident in its name—National *Emergency* Library—and from the messaging around the project in press releases and social media that indicated a limited timeframe for these changes. Indeed, the project closed on 16 June 2020, two weeks earlier than initially advised, and returned to the one borrower at a time controlled digital lending model that the Internet Archive's Open Library had been using for some years. The reason for the early closure was that between March and June, the National Emergency Library received significant backlash from copyright and authors' advocacy organizations. The Copyright Alliance, for example, describes the National Emergency Library as "a wolf in sheep's clothing" and an example of "trying to manipulate others to achieve their own gain" (Bramlet 2020). The Authors Guild published an open letter, signed by thousands, not only asking the Internet Archive to "remove the hundreds of thousands of in-copyright books" in the National Emergency Library but also arguing that even the controlled digital lending approach used in the Open Library was illegal. The letter questioned the Internet Archive's status as a library and the legal benefits that flow from using that term (Authors Guild 2020). On 1 June, four publishers filed a lawsuit accusing the Internet Archive of "wilful mass copyright infringement" (Dwyer 2020). This dispute is a key example of the continuing tensions at play in the very idea of books and libraries are in the digital age, in notions of access, preservation, commercialization of content, recompense for creative labour and of public rights to information, and of competing impulses towards openness and access on the one hand and towards the protection of the rights of content creators in the age of FAANG (Facebook, Amazon, Apple, Netflix, and Google) on the other. The Internet Archive and its Open Library and National Emergency Library projects are case studies of public-private partnerships in digital culture in that while the Internet Archive has been granted library status it is clearly many other kinds of organization, or institution, at the practical and conceptual levels as well. And there are tensions between these identities.

profit-driven, or sensitive agendas (Thylstrup 2019). This paper investigates a series of case studies of electronic access to books and cultural heritage, each incorporating some notion of a public-private partnership and some notion of the importance of open access agendas. It explores the metaphors that we use to conceptualize a form of open social scholarship based on an open social digitization agenda, including the HathiTrust's Digital Library, Google Books, and Microsoft's partnership with the British Library. The paper asks how the principles of open social scholarship contribute to a better and more nuanced understanding of digitization as a cultural practice and asks how a better understanding of the networks, partnerships, and paperwork (agreements, policies etc) of digitization could inform developments in open social scholarship. The public-private partnerships that surround the digitization of the cultural record assist underfunded cultural organizations to do important digitization work but also introduce complexities and contradictions in open social access to information. In addition, the mix of volunteer and crowdsourced labour operating in collaboration with cultural sector professional staff create intricate chains of value over time, space, and sector. The presence of large-scale corporations inside these arrangements has added another layer of complexity over time, and one that many public stakeholders are still struggling to navigate. The case studies examined here illuminate new facets of the large-scale international agenda of digitization at the social and cultural

This paper follows these threads to investigate a series of case studies of electronic access to books and cultural heritage, each incorporating some notion of a public-private partnership and some notion of the importance of open access or public good agendas, using as case studies projects like the HathiTrust's Digital Library, Google Books, and Microsoft's partnership with the British Library in the ill-fate Live Search Books project. The paper asks how the principles of open social scholarship contribute to a better and more nuanced understanding of digitization as a cultural practice and asks how a better understanding of the networks, partnerships, and paperwork (agreements, policies etc) of digitization could inform developments in open social scholarship. The public-private partnerships that surround the digitization of significant components of the cultural record, both textual and object-based, are a great opportunity to preserve and to bring new audiences to cultural works that are often difficult to access. Indeed, much discussion of the original partnership between Google and the university and public libraries it entered into arrangements with in 2004 highlights the significance of the opportunities the libraries saw in solving a very real conundrum: how to resource large scale digitization efficiently and effectively. The mix of volunteer, crowdsourced, and professional cultural sector labour creates intricate chains of value over time, space, and sector, and create nuanced overlappings between stakeholder groups that enrich organizations and their collections. The presence of large-scale corporations inside these arrangements has added another layer of complexity over time, and one that many public stakeholders are still struggling to navigate. The case studies illuminate new facets of the large-scale international agenda of digitization at the social and cultural level and provide insight to help establish a framework for digitization that is open, social, and collaborative. The principles of open scholarship provide a useful first step in unpacking this.

For open social scholarship (Arbuckle et al 2019), a key focus is the relationship between specialist and non-specialist knowledge production and dissemination where "knowledge persists beyond the borders of the university and has the ability to negotiate diverse spaces, institutions and communities" (El Khatib et al 2019). While the focus of research on open social scholarship is on the social lives of university-generated research and its capacity to reach beyond university contexts, the public-private partnerships upon which foundational components, methodological tools, or objects of study are based need attention. The ground upon which the specialist and non-specialist work is a determining factor in the openness and the socialness of the scholarship that can emerge from it.

Similarly, digitization of the cultural record (rather than digit*aliz*ation) has progressed in a particularly piecemeal way with local, national, and international perspectives leading different approaches and with different

expressions across public/private, specialist/non-specialist, and preserved/accessible lines. While the nomenclature and conceptual framework for these practices is yet to solidify, in general *digitization* can be used to refer to the translation of physical objects—such as books but also including other material cultural objects—into digital form, by scanning or by photography, with the addition of metadata and other descriptive work so the digital objects can be accessed, read, and shared electronically. *Digitalization*, one the other hand, might more usefully refer to the condition in which we find ourselves in as we shift to online modes and spaces for the production, circulation, and reception of cultural matter. Distinguishing between these modes is a way of acknowledging the combination of the digitized and the born digital and the spaces in between, and the increasingly digital nature of all work, especially in arts and culture.

There are as many models for digitization as there are digitization projects in the world (Holley 2010; Dahlstrom et al 2012; Tanner 2016). Large scale projects offer a mass approach to digitization. Over the last few decades, digitization work has proceeded by whatever means available in terms of funding, labour, hardware, and software solutions, in the light of austerity and "efficiency dividends" that have befallen cultural organizations in recent times, but also disinterest and confusion from the government sector with rare exceptions usually driven by political expediency (White 2017). While the production, circulation and reception of born-digital material has much to tell us about cultural consumption, here I am focussed on the particularities of cultural objects that existed as tangible material works before becoming digitized through a process that involved humans in professional or semi-professional settings of some kind. Examples of knowledge infrastructure occurring across the boundaries of public-private partnerships reveal some of the stress fractures in the sector.

The Google Books project has received a great deal of attention in the media and in scholarship, less on its own terms but more because of its disruption to the complicated literary marketplace and publishing industry that solidified in a particular way in the 1980s and found it difficult to adapt to the promises and challenges of a constantly evolving digital era (Ray Murray and Squires 2013; Barnett 2016). While publishing was never a fair industry equitably distributing rewards for hard work, the polarising of success and the rise of the blockbuster in the 1980s in particular created a reward system that not only was largely unviable to the majority of published authors but left them particularly vulnerable to the changes the digital era would bring. The early days of the Google Books project began in conversations between the tech giant—in 2004 not quite in the dominant position it is now—and libraries who had struggled for some time to figure out how to resource digitization projects to keep current in the so-called "late age of print" (Striphas 2009). Resourcing was not the only impetus for

the relationship, as I explain below. Google's relationship(s) with libraries is not the only public-private partnership that reveals details about the depth and breadth of the massive international digitization project, though it is certainly the most visible thanks to a long running set of lawsuits. In December 2004, the Internet Archive announced its own collaborations with significant international libraries, only a few weeks after Google had announced its Google Print Project at the Frankfurt Book Fair in October 2004. Headlines from October 2005 announce that "Yahoo Works with Academic Libraries on a New Project to Digitize Books" (Carlson and Young 2005), and closer examination shows the Internet Archive's fingerprints there too, as Yahoo joins with the Internet Archive, the University of California, the University of Toronto, Adobe, the European Archive, the National Archives of England, O'Reilly Media, and Hewlett-Packard Labs to form the Open Content Alliance. The *New York Times* had called them "an unusual alliance of corporations, non-profit groups and universities" (Hafner 2005). A month later, Microsoft announced it had entered into "a strategic partnership" with the British Library to develop Live Search Books (November 2005), a project that would come under examination when abandoned prematurely just three years later. Public-private partnerships such as these need scrutiny to determine what it is they are contributing to the imaginary of cultural stewardship.

Unpacking the project of global mass digitization requires an understanding of the infrastructures upon which the collections that support that digitization were built. This means not only the formats, institutions, and even buildings in which they are housed, but also collections policies and partnership agreements that have governed extant collections and their digitization over time and also the national cultural, educational, and infrastructural policies, explicit or implicit, at work in the framing and resourcing of the work of those cultural institutions. For example, the post-war nation-building agenda of the resourcing of cultural institutions in many of the nations contributing to UNESCO, at least, emerges as one factor shaping the rhetoric of access to local collections. How institutions such as these have circulated notions of preservation and access that are evident not only in the language of cultural organizations engaging in digitization, but also the private organizations and corporations partnering with them, tells us much about the underlying elements of digitization's cultural work and objectives.

Digitization is controversial in part because of issues of copyright, funding, technology, and access (Thylstrup 2019). These issues gain headlines and airwaves, but the implications of the decisions we are making about what, when, and how we digitize or about what we do with the end product remain largely unexamined. Different industry sectors and academic disciplines have responded to this in different ways. For example,

the field of library and information management has much to say about the components of digitization and the practices involved in the creation of digital cultural infrastructure projects (Darnton 2013), while Digital Humanities has tended to focus on the end use of digitized corpuses rather than the effects of the material conditions of digitized objects, though this is changing (Liu; Smithies). By using the term public-private partnerships, a term with a particular prevalence in current neoliberal corporate speak, I want to draw attention to the interconnections involved in prioritising, resourcing, and communicating of digitization projects in organizations large and small and to the place of corporate management and new public management approaches governing the existence of these components. These public-private partnerships operate not only between the big actors (Google, large universities, and large public libraries) but also in the small-to-medium cultural organization category. My argument here is predicated on the notion that the materiality of the cultural artefact, and how that materiality is represented in the digital copies generated by digitization work, is part of the meaning-making, reception, and interpretive framework. To have a real understanding of what digitization is and how it is working in our cultural environments, we must consider these entanglements and the public-private arrangements that have often accidentally set the parameters for the practice.

Google

In 2004, before the Internet Archive, Google began its own controversial book digitization project in collaboration with a range of major libraries around the United States and then the world. For the libraries it was an opportunity to push forward the utopian agenda to make human knowledge accessible to all, utilizing the affordances of digital technology to overcome distance, ability and health factors, opening hours, and austerity agendas. Little archival or historical documentation is currently available in the public domain to shed light on the nature of those early motivations or the characterization of the project by either side, due to confidentiality agreements and persisting sensitivities around the early contracts, though some are accessible due to the Wayback Machine. Much more distance in time and interest is needed before the early details and the later consequences can be fully explored. However, some evidence is available that tells us something about this initial framing, or its public expression. The early participants in the partnerships had a range of stated and unstated objectives in developing the collaborations. The press releases and interviews from the time reveal much of the enthusiasm the libraries possessed, in coming to partner with Google on the digitization work, and some of the more public-facing and publicly expressed motivations for engaging in the partnerships, though other components in the decision

making, framing, and executing of these partnerships are not evident in those early declarations. The publicly available communications about the partnerships typically foreshadow many concerns that have only grown in the last fifteen years around the role that publicly funded knowledge institutions such as public and university libraries play in knowledge production and communication, especially given the growing power and revenue of journal publishing giants like Elsevier. The early public statements foreground on the part of libraries the importance of providing platforms for access, opening often private or otherwise inaccessible collections to broader users, overcoming geographical location or enrolment status or ability profile, or for preservation purposes. For Google, the early statements foreground the importance of access to knowledge for students, researchers, and people far from libraries. Unsurprisingly, there is no record of any reference to data-mining or training AI on the corpuses to be digitized from libraries, uses that would later emerge as important components of the rationale for digitization (Gray 2020; Thylstrup 2019).

Considering larger partnerships, it is also important not just to look at the rhetoric and contractual details at play in the early phase of the project, but also where those arrangements went in practice and how the rhetoric may have changed over time, as well as what other factors intervened to influence the entanglements. For example, lawsuits between the Authors Guild and Google Inc unfolding between 2005 (when both the Authors Guild and the Association of American Publishers launched separate legal action against Google, through several ill-fated attempts at settlement agreements) and 2016 (when the US Supreme Court rejected a request to review the case, putting an end to a decade of legal action that had subsumed the broader discussion around digitization) had a number of effects. It created or at least exacerbated tension between the libraries and Google. Again, until the paperwork reflecting the signing and acquitting of original agreements between Google and the libraries becomes publicly accessible, and early participants are interviewed frankly without the threat of a non-disclosure agreement, the exact trajectory of these relationships and the various sticking points within them is unlikely to be uncovered. And it is impossible to know how this relationship would have progressed without the stressor of the Authors Guild lawsuits. The lawsuits played a role in determining the relationship between the various actors and the publicly accessible outcomes. Looking at public statements of some of those involved in the early days of the Google Books agreements shows a journey over time in emotional response to the project.

Robert Darnton, book historian, and from 2007-2016 Director of the Harvard University Library, was an early supporter of the Google Books partnerships with university libraries, seeing in the project a means to reach

the utopian goals of widening access to the universities' knowledge infrastructure. According to Darnton:

we want to open up our collections and make them available to readers everywhere. How to get there? The only workable tactic may be vigilance: see as far ahead as you can; and while you keep your eye on the road, remember to look in the rearview mirror (Darnton 12 Feb 2009).

This enthusiasm faded somewhat over time (Darnton 2011), worn down perhaps by the lawsuits between Google and the Authors Guild, and in 2011 the Authors Guild sued the HathiTrust, the institutional entity built out of Google's former library partners, which describes itself as "a digital preservation repository and highly functional access platform" (HathiTrust). But it wasn't just legal headaches tarnishing the original collaborative vision but also a fundamental shift in the way the we understand the cultural work of digitization. In 2013 Darnton's enthusiasm resurfaced, angled in a different direction, in support of the newly launched Digital Public Library of America (Darnton 2013), a non-profit organization that says it "empowers people to learn, grow, and contribute to a diverse and better-functioning society by maximising access to our shared history, culture and knowledge" (DPLA website).

What's interesting to note here, though, is the way in the intervening years between the first collaboration between university libraries and Google in 2004, through the creation of the HathiTrust as a way of combatting the Google tide, to Darnton's role in setting up the Digital Public Library of America, the metaphors for understanding the work of digitization, the objects and collections generated by digitization, and the institutions that house digitization work, changed significantly. There has been a return to the concept of the 'library' to frame the work of digitized and digitalized textual objects. The notion of a digital library became the safest way to communicate the work of collaborations between university collections and digital access providers. Much of this played out in the pages of *The New York Review of Books:* "The Library in the New Age" (12 June 2008), "Google and the Future of Books" (12 February 2009), "Google and the New Digital Future" by Robert Darnton 17 December 2009, "Can We Create a National Digital Library?" (28 October 2010), "The Library: Three Jeremiads" (23 December 2010), "Google's Loss: The Public's Gain" (28 April 2011), "A World Digital Library Is Coming True!" (22 May 2014), "Great New Possibilities for the Library of Congress!" (13 August 2015). And that's just a few. What we see here is a shift from a logic of books as standalone resources and search and access as their main purpose (Hillis et al 2012), to libraries and collective public knowledge for the public good, from an extraction logic to a building logic.

For its part, Google's rhetoric around the motivations of mass digitization also emphasize various notions of access on the surface at least. Google

presents itself as a generous partner in the honourable ambition of providing access to all of the world's information rather than as an international business with corporate interests in digitizing content. The official commentary from Google repeatedly stresses collaboration: "we've worked with publishers and authors around the world" and in a press release, one of Google's founders, Sergey Brin, says, "Google's mission is to organize the world's information, and we're excited to be working with libraries to help make this mission a reality." Here Brin's use of the word "organizing," so apparently neutral, takes on rather sinister implications. One might ask according to what principles or policy this organizing occurs and what agency is involved in it. But it is easy to raise these questions with the benefit of hindsight. Google also emphasizes the role it can play helping struggling authors to publicize and sell their works and gain readerships. They argue that the Google Books project "expands the market for publishers to sell books. It also creates more opportunities for authors to earn money and be read" (Google 2009).

Librarians participating in the project tended to echo Google's utopian rhetoric, at least in the early days of the project. In fact, Google lists quotes from its partner libraries on the Google Books website in support of the work of the project. For example, Hubert A. Villard, Director of the Cantonal and University Library of Lausanne, says: "We are quite literally opening our library to the world. The opportunities for education are phenomenal and we are delighted to be working with Google on this project." John P. Wilkin, Associate University Librarian, University of Michigan, similarly endorses the collaboration, stating "Our mission as a great public university to advance knowledge — on campus and beyond," and citing their work with Google as an agent of that mission. For more information on the political undercurrents of digitization in the US and Europe, see Nanna Bonde Thylstrup's *The politics of mass digitization* (2019).

Microsoft

Not wanting to be left too far behind Google, Microsoft first announced its partnership with the British Library in 2005 to digitize 25 million pages from about 100,000 out-of-copyright print book collection. The agreement was for the scans to be accessible through both Microsoft's Live Search search engine and through the British Library's own catalogue system. And the plan was to add more books each year as their copyright expired. Live Search Books was launched in December 2006. The Internet Archive had been contracted by Microsoft to undertake much of the scanning work. In 2006, the Live Search Academic website FAQ read:

> Windows Live Academic is a new addition to the Windows Live

> Search family of services that allows users to search through academic information. Currently, users can search content in academic journals in the fields of Computer Science, Electrical Engineering, and Physics. We will be adding more subject areas in the near future, based on user feedback and demand (Microsoft 2006a).

A 24 May 2006 webcrawl by the Wayback Machine found this response to a question:

> Why don't you have content from all fields?
>
> Academic search has launched in a beta version so we can receive feedback from our users - ultimately allowing us to introduce a product that will provide the best possible user experience. We understand that for researchers to have a productive search experience, they need to search a comprehensive index in their field of study. Therefore, we decided to launch our beta version with journal content from Computer Science, Engineering (mostly electrical and electronics), and a good selection of Physics journals. We believe that our deep index in these chosen fields will serve the needs of our users well, so they can give us the feedback we need to improve the search experience. After launch, we will add content in phases from more subject areas. Our goal is to have the most comprehensive, largest academic index possible.
>
> (Microsoft 2006b)

The intention was clearly to ramp up digitization, access, and search, with initial announcements giving a target of 100,000 books in the first instance. According a news item in *The Guardian*, the purpose of the "strategic partnership" was "not only about digitization and preservation, but also about delivering a great experience for people accessing this amazing collection through British Library and MSN websites," a quote attributed to Bill Gates (4 November 2005). But in May 2008, Microsoft announced the cessation of the partnership and its intentions to close down its Live Search Books and Live Search Academic projects. Sixty thousand books had been digitized from the British Library collections. The British Library was left with a partially completed dataset of digitized works but has, through its British Library Labs project and Andrew Mellon funding, repurposed the digitized collection for creative and cultural projects with strong outcomes (Ridge 2018) including the development of an international Library Labs and GLAM Labs network (Mahey et al 2019).

Internet Archive's Open Library

The Internet Archive also commenced a book digitization program in the middle of the first decade of the twenty-first century. Prior to launching the National Emergency Library in March 2020, the Internet Archive's Open Library has used a controlled digital lending approach to managing the provision of public access to scanned in-copyright works within what it sees as the agreed legal framework. The books have been scanned over the last fifteen years, since the Internet Archive began partnering with libraries to scan books in their collections in 2005 and it launched the Open Library in 2011, a service that offers a two-week loan of digital copies of books scanned from originals held in partner libraries, with limits to how many people can borrow at any one time, and waiting lists to manage the loan system. That is, the Open Library operated in a way similar to a physical library for which loaning is limited and controlled. This is in contrast to, say, Project Gutenberg which maintains text-only copies of out-of-copyright books (no page scans) or the HathiTrust which maintains digitized page scan copies of out-of-copyright books that can be accessed by anyone who is interested in them at any time. The only limit there is the bandwidth of the servers on which they are housed. In this way the Internet Archive sought to apply the library model to the circulation of copyrighted works to the public, consistent with practice in other library contexts. While it was controversial, the use of a waitlisting system and restricted lending capacities rather than a complete open access free-for-all system kept the controversy in check, though a series of open letters, take down notices and legal action in the UK in particular bubble along (Flood 2019).

State and National Archives

There's one more public-private(ish) partnership I would like to add into this mix to think about its implications for open social digitization. When I visited the State Records of South Australia's Open Day in 2018, travelling out to Adelaide's northern suburbs enticed by a session demonstrating digitization at work in their organization, I discovered two volunteers sitting at the digitization station. When I came upon them, they were in the middle of digitizing coroners' reports and told us some of the interesting items they'd found in between pages of the bound reports, including a hair and bullet. They were older women, retired, and both spoke with American accents. They were representatives of the Church of Jesus Christ of Latter-day Saints spending time in Adelaide as part of their mission work digitizing collections. The church runs FamilySearch, a non-profit organization that provides access to genealogical records through a website and proprietary software. Operating under the name Genealogical Society of Utah since 1894, the organization began digitizing family history records in 1998, after a long history of collecting and preserving analogue records (Kriesberg 2016). This partnership between cultural institutions large and

small and the Mormon Church is happening around the world (ABC Radio Adelaide 2018; Little 2012; Kriesberg 2016; Laite 2020). The FamilySearch company gives multiple reasons for undertaking this work, such as the importance of family and community service to the Church (ABC Radio Adelaide 2018), but also the controversial idea of "baptism of the dead" (Laite 2020; Almeida et al 2011; Bauer 2015):

> The Mormon Church teaches that their members are responsible to be baptized for dead loved ancestors. If a person dies having never been baptized in this life, a Mormon relative can be baptized in his place. Then the dead person may have a chance after death to believe the gospel, repent, and be saved. Joseph Smith, the founder of Mormonism, taught that seeking the dead in this manner is a Mormon's greatest responsibility (Almeida et al 2011 p. 146)

The importance of volunteer labour to the cultural sector, especially in times of increasing austerity, and the complexities of negotiating public-private relationships given the legal complications of the Google Books project and the mercurial nature of corporate interests especially in the tech industry's well-known competitive striving means that relationships with groups like the Church of Jesus Christ of Latter-day Saints are an important component for the growth of access to resources. There needs to be a degree of transparency about the arrangements made for digitization work. This is an international component of the resourcing and infrastructure of digitization and the future of collections, their accessibility, their preservation, and their use. It is a cheap and efficient means of solving the problem of digitization in the context of decreasing funding for cultural collections. But it is also a part of the infrastructural conditions of those textual, cultural, and historical records and objects, and their participation in digital collections is underwritten by a particular religious agenda. My argument is that these conditions have to be taken into account to understand digitization and the digital cultural record. These issues extend to the particular agreements in place between Ancestry.com and public archives as well as other digitization of public material for commercial use as well (Kriesberg 2017; Manžuch 2017).

Volunteer and crowdsourced labour in public-private partnerships

The case of volunteer scanners from the Mormon Church is a particularly interesting example of managing the labour of digitization from a policy perspective, but across all of the case studies of public-private partnerships in digitization in this paper and beyond are different forms of the sharing of labour resourcing of tasks, within and between the public and private

organizations but also entirely beyond them, as in the example of crowdsourcing. How are these labour-intensive projects resourced, especially given the requirements for speed as well as quality, and the sheer enormity of the work? The most obvious case study for the issue of labour in digitization projects is the Google Books project, especially in the parts of the project that occurred on Google's main Mountain View California site. Despite the significant and robust non-disclosure agreements at work in all elements of the Google Books project, some components of the labour conditions of Google Books staff have emerged. Video artist Andrew Norman Wilson brought some of the issues around labour conditions of Google Books employees to light in 2011 after he briefly worked there as a sub-contractor doing video work. In his 11-minute video "Workers Leaving the Googleplex," Wilson narrates his experience being fired for filming some of the staff of the Google Books project leaving their building at the end of their shift (Wilson 2011; Zeffiro 2019). Wilson's experience reveals the extent of the different working conditions the book scanners experience compared to those other workers who have access to the legendary benefits Google employees supposedly receive. The free shuttle services, electric scooters, talks by famous authors, foosball tables, colourful beanbags, and free food are not enjoyed by those employed in the Google Book scanning projects. Google uses low-paid labour and gives poor working conditions to this workforce. These poor working conditions and the very high scan rates required have resulted in the numerous scan errors inside the digitized copies: the fingers of the scanners, pages caught in motion mid-scan, the inclusion of ephemera, multi-leaved pages not folded out for scanning, and so on (Barnett 2016; Zeffiro 2019). Wilson filmed the Google Books scan team outside of the building in which the commercial-in-confidence work is occurring and after they have clocked off from their shifts, and therefore beyond the technologies and workflows that are ostensibly the subject of Google's nondisclosure agreements, and yet Wilson was fired for it. The 'private' component of this partnership is clear—Google's rhetoric of openness and access is always at odds with its secrecy and resource hoarding—but the 'public' component needs some more exploration.

The high-profile case only tells part of the story of Google's digitization efforts and its labour requirements. Only some of the book scanning occurred on Google sites; much of it occurred in the venues of the partner libraries such as the University of Michigan Library, Stanford Library, Harvard Library, New York Public Library, and Oxford's Bodleian Library in the UK. These institutions have remits as caretakers of knowledge for the public good. There is not a great deal on the public record about the labour conditions of the scanning projects that occurred in these sites on the public record, in part because of the nondisclosure agreements and in part because the identity of libraries as champions of public good has been caught up

with series of lawsuits over a decade. As the relationship with Google turned sour, there was a real hit to the identity and purpose of libraries in the digital age. All of the utopian rhetoric spouted by libraries in announcing these partnerships, some years after the lessons of the dot.com bubble burst, was once again consumed by the partnerships.

Another source of labour for digitization projects is crowdsourcing, which grows as a strategy for solving the problem of too much work and too little labour in an era and context of digital sharing technologies. Crowdsourcing has become an area of interest, growth, and opportunity across the humanities and the Galleries, Libraries, Archives, and Museums (GLAM) sector. It is both a strategy for solving labour problems but also for building relationships between cultural collections and the publics they serve (Mahey et al 2019, p.87). Melissa Terras argues that

> We are now at a stage where crowdsourcing has joined the ranks of established digital methods for gathering and classifying data for use in answer the types of questions of interest to humanities scholars, although there is much research that still needs to be done about user response to crowdsourcing requests, and how best to build and deliver projects (2016, p. 432).

Crowdsourcing can refer to both paid and unpaid labour in digital humanities projects. Ellen Cushing's 2013 examination of microtasking economy supported by Amazon's Mechanical Turk platform, one of a number of low-cost outsource labour platforms, highlights some of the questions around using the platform to pay very low rates to people in other geographical domains to undertake tasks like OCR correction, data cleaning, and metadata tagging. But, in addition to these paid arrangements, many crowdsourced relationships to digitized cultural heritage are voluntary. Projects like the National Library of Australia's Trove (Holley 2010; Ayers and Andro 2013) maintain a loyal user-base of volunteers particularly for their digitized newspaper OCR text correction program.

Finally, I want to point to the case of the use of prison labour in digitization work. Bauer reports that inmates are used to digitize a range of government and non-government documents at a rate of between US$0.60 and US$1.75 per hour except for the Church of Jesus Christ of Latter-day Saints, which seeks volunteers amongst the prison population to digitize genealogical records (Bauer 2015).

I want to suggest that the labour of digitization in public-private partnerships, where universities, libraries (public or private), and cultural organizations partner with private companies to digitize (or otherwise transform) works in the public domain for private and commercial purposes complicates the very foundations of the public domain, public

goods (and public good), and the work university libraries are striving to undertake. Here public works are essentially reprivatized.

Conclusion

There are many other agents at work in the public-private partnership space in book, archival, and cultural object digitization beyond the signed agreements between major organizations. Included in this mix are the range of privately developed platforms at play, amongst other things. There are clearly consequences for open social knowledge and open social scholarship of the public-private partnerships at the heart of our major digitization projects. There is an inherent tension in almost all digitization efforts between the private nature of books (both in and out of copyright) and the private company motives of, for example, Google and Microsoft that are enacted through the infrastructure of the public good (public libraries, university libraries, and other cultural organizations). This complicates the work of digitization and the task of the scholar trying to make sense of its practical and conceptual impacts on the day-to-day life of culture and scholarship, of the balance between and unique requirements of access and preservation, and for the future of the historical record. Work that remains to be done here includes looking into the role of private corporations in undertaking digitization in other nations around the world, the unique conditions of the public-private partnerships enacted in seeking cost-effective digitizations, the labour conditions involved in the digitization, and the consequences for the public good. Digitizations need provenance information that include the infrastructure labour conditions that are implicated in their creation.

Google has essentially re-privatized public knowledge and the private knowledge of other individuals while much of the legal wrangling focussed on a very narrow set of legal questions around fair use of in-copyright works. And in this there was a sleight of hand, deflecting from questions over whether Google should *hold* these collections and what *private* purposes they should be permitted to put them to.

The Internet Archive's National Emergency Project was clearly an attempt to push at the boundaries of access, digital libraries, and controlled digital lending as much as it was an attempt to give to a locked down public informational and creative resources in a time of crisis. But it also reveals an ongoing set of tensions around the digital object in settings that may be deemed libraries. The Authors Guild's concerns about controlled digital lending are clearly warranted given the abysmally low pay points of publishing authors, but this fraught line between the Internet Archive and authors is just one relationship in a complex network of cause and effect around the digital object in a digital literary marketplace.

Acknowledgements

Bibliography

ABC Radio Adelaide. 2018. "Why does the Mormon Church want state records? And what do they do with them?" *ABC News* 5 July Retrieved May 13, 2020, from https://www.abc.net.au/news/2018-07-05/why-does-the-mormon-church-want-state-records/9943288

Almeida, Alessandra BS, Rafael Dueire Lins, and Gabriel de F. Pereira e Silva. 2011. "Thanatos: automatically retrieving information from death certificates in Brazil." In *Proceedings of the 2011 Workshop on Historical Document Imaging and Processing*, pp. 146-153. 2011.

Arbuckle, Alyssa, Luis Meneses, and Ray Siemens. 2019. "Introduction, Beyond Open: Implementing Social Scholarship." *KULA: knowledge creation, dissemination, and preservation studies* 3.1, 8. DOI: http://doi.org/10.5334/kula.55

Authors Guild. 2020. "Authors Guild Sends Open Letter to Internet Archive and Brewster Kahle Demanding Open Library's 'National Emergency Library' Shut Down." Retrieved May 13, 2020, from https://www.authorsguild.org/industry-advocacy/ag-sends-open-letter-demanding-national-emergency-library-shutdown/

Ayres, Marie-Louise, and Mathieu Andro. 2013. " 'Singing for their supper': Trove, Australian newspapers, and the crowd." In *IFLA WLIC 2013*. 2013.Retrieved 24 July 2020 from https://hal.inrae.fr/hal-02746003/document

Barnett, Tully. 2016. "The Human Trace in Google Books" *Border Crossings*, Wakefield Press, Kent Town, South Australia, 53-71.

Bauer, Shaun. 2015. "Your Family's Genealogical Records May Have Been Digitized by a Prisoner" 13 August. Retrieved July 23, 2020 from https://www.motherjones.com/politics/2015/08/mormon-church-prison-geneology-family-search/

Bramlet, Eileen. 2020."Internet Archive's 'Emergency Library': A Wolf in Sheep's Clothing" Copyright Alliance. Retrieved May 13, 2020, from

https://copyrightalliance.org/ca_post/internet-archive-emergency-library-for-covid-19/

Carlson, Scott. and Young, Jeffrey, R., (2005). "Yahoo Works with Academic Libraries on a New Project to Digitize Books". *Chronicle of Higher Education*, *52*(8).

Cushing, Ellen. (2013). "Amazon Mechanical Turk: The Digital Sweatshop" *Utne Reader*. January/February 2013. Retrieved 24 July 2020 from https://www.utne.com/science-and-technology/amazon-mechanical-turk-zm0z13jfzlin

Dahlström, Mats., Joacim Hansson, and Ulrika Kjellman, (2012). "'As We May Digitize'—Institutions and Documents Reconfigured". *Liber Quarterly*, 21(3-4) 455–74. DOI: http://doi.org/10.18352/lq.8036

Darnton, Robert. 2008. "The Library in the New Age" *The New York Review of Books* 12 June Retrieved May 23, 2020 from https://www.nybooks.com/articles/2008/06/12/the-library-in-the-new-age/

Darnton, Robert. 2009a. "Google & the Future of Books", *The New York Review of Books* Retrieved May 13, 2020 from https://www.nybooks.com/articles/2009/02/12/google-the-future-of-books/

Darnton, Robert. 2009b. "Google and the New Digital Future" *The New York Review of Books* 17 December Retrieved May 23, 2020 from https://www.nybooks.com/articles/2009/12/17/google-and-the-new-digital-future/.

Darnton, Robert. 2010a. "Can We Create a National Digital Library?" *The New York Review of Books* (28 October) Retrieved May 23, 2020 from https://www.nybooks.com/articles/2010/10/28/can-we-create-national-digital-library/

Darnton, Robert. 2010b. "The Library: Three Jeremiads" *The New York Review of Books* 23 December. Retrieved May 25, 2020 from https://www.nybooks.com/articles/2010/12/23/library-three-jeremiads/

Darnton, Robert. 2011a. "Six Reasons Google Books Failed", *The New York Review of Books: NYR Daily* Retrieved May 25, 2020 from https://www.nybooks.com/daily/2011/03/28/six-reasons-google-books-failed/

Darnton, Robert. 2011b. "Google's Loss: The Public's Gain" *The New York Review of Books* (28 April). Retrieved May 23, 2020 from https://www.nybooks.com/articles/2011/04/28/googles-loss-publics-gain/

Darnton, Robert 2013. "The National Digital Public Library is Launched!" *The New York Review of Books*. 25 April. Retrieved May 13, 2020, from </https:> Darnton, Robert. 2014. "A World Digital Library Is Coming True!" *The New York Review of Books* (22 May), Retrieved May 13, 2020 from https://www.nybooks.com/articles/2014/05/22/world-digital-library-coming-true/

Darnton, Robert. 2015. "Great New Possibilities for the Library of Congress!" *The New York Review of Books* (13 August). Retrieved May 23, 2020 from https://www.nybooks.com/articles/2015/08/13/great-new-possibilities-library-congress/

DPLA (n.d). "DPLA About Us" Retrieved July 19, 2020 https://dp.la/about

Dwyer, Colin. 2020 "Publishers Sue Internet Archive for 'Mass Copyright Infringement'" *NPR*. 3 June. Retrieved July 23, 2020 from https://www.npr.org/2020/06/03/868861704/publishers-sue-internet-archive-for-mass-copyright-infringement

El Khatib, Randa, Lindsey Seatter, Tracey El Hajj, Conrad Leibel, Alyssa Arbuckle, Ray Siemens, and Caroline Winter. 2019. "Open Social Scholarship Annotated Bibliography". *KULA: knowledge creation, dissemination, and preservation studies*, *3*(1).

Google 2009. "Google Books Settlement Agreement with Authors and Publishers," A video uploaded by Google on 23.6.2009 and spoken by "Nathan, an Engineer for Google Books" http://www.youtube.com/watch?feature=player_embedded&v=J16juV1acCI Accessed 14 May 2020. http://www.youtube.com/watch?feature=player_embedded&v=J16juV1acCI

Gray, Joanne. 2020. *Google Rules: The History and Future of Copyright Under the Influence of Google*. Oxford University Press.

Hafner, Katie. 2005. "In Challenge to Google, Yahoo Will Scan Books" *The New York Times* 3 October . Retrieved May 13, from https://www.nytimes.com/2005/10/03/business/in-challenge-to-google-yahoo-will-scan-books.html

HathiTrust (n.d). "Our Digital Library" Retrieved 19 July, 2020. https://www.hathitrust.org/digital_library

Hillis, Ken., Michael Petit, and Kylie Jarrett. 2012. *Google and the Culture of Search*. Routledge.

Holley, Rose. 2010. "Trove: Innovation in access to information in Australia". *Ariadne*, (64).

Internet Archive. 2020. Announcing a National Emergency Library to Provide Digitized Books to Students and the Public. *Internet Archive Blogs*. 24 March. Retrieved May 13, 2020, from http://blog.archive.org/2020/03/24/announcing-a-national-emergency-library-to-provide-digitized-books-to-students-and-the-public/

Kriesberg, Adam. 2017. "The future of access to public records? Public–private partnerships in US state and territorial archives." *Archival Science* 17.1, 5-25.

Laite, Julia. 2020. "The Emmet's Inch: Small History in a Digital Age." *Journal of Social History* Volume 53, Issue 4, Summer 2020, 963–989

Little, Hannah n.d.. "Microfilm, Mormons and the Technology of the Archive." *eSHARP* Issue 12 Retrieved 13 May, 2020, from https://www.gla.ac.uk/media/Media_102866_smxx.pdf

Liu, Alan. 2016. 'Drafts for Against the Cultural Singularity.' HCommons DOI 10.17613/M6SS3B.

Mahey, Mahendra., Al-Abdulla, A., Ames, S., Bray, P., Candela, G., Chambers, S., Derven, C., Dobreva-McPherson, M., Gasser, K., Karner, S., Kokegei, K., Laursen, D., Potter, A., Straube, A., Wagner, S-C. and Wilms, L., with forewords by: Al-Emadi, T. A., Broady-Preston, J., Landry, P. and Papaioannou, G. (2019) *Open a GLAM Lab*. Digital Cultural Heritage Innovation Labs, Book Sprint, Doha, Qatar, 23-27 September, 2019. Retrieved July 19, 2020. https://glamlabs.pubpub.org/

Manoff, Marlene. (2006) 'The Materiality of Digital Collections: Theoretical and Historical Perspectives.' *Portal*: 6.3 2006: 311-325.

Manžuch, Zinaida. (2017) "Ethical Issues In Digitization Of Cultural Heritage,"Journal of Contemporary Archival Studies: Vol. 4 , Article4.Available at: http://elischolar.library.yale.edu/jcas/vol4/iss2/4

Microsoft. 2006a. "Welcome to Windows Live Academic" 12 April. Retrieved May23, 2020 from https://web.archive.org/web/20060412104432/http://academic.live.com/.

Microsoft. 2006b. "Welcome to Academic Search" 24 May. Retrieved May 23, from https://web.archive.org/web/20060524173547/http://academic.live.com/

Ortega, Élika. 2017. "Not a Case of Words: Textual Environments and Multimateriality in Between Page and Screen." *Electronic Book Review* Retrieved May 13, 2020 from https://electronicbookreview.com/essay/not-a-case-of-words-textual-environments-and-multimateriality-in-between-page-and-screen/

Ray Murray, Padmini. and Claire. Squires. 2013. "The digital publishing communications circuit." *Book 2.0* 3.1, 3-23.

Ridge, Mia. 2018. Breathing life into digital collections at the British Library. *Access*, *32*(3), 40.

Rissam, Roopika. 2019. "The stakes of digital labor in the twenty-first-century academy: The revolution will not be Turkified." In *Humans at Work in the Digital Age*. Routledge. 239-249.

Smithies, James. 2017. The Digital Humanities and the Digital Modern. London. Palgrave Macmillan.

Tanner, Simon. 2016. "Using impact as a strategic tool for developing the digital library via the Balanced Value Impact Model." Library Leadership & Management 30.4.

Thylstrup, Nanna. Bonde. 2019. *The Politics of Mass Digitization*. MIT Press.

Terras, Melissa. 2016. Crowdsourcing in the Digital Humanities. In: Schreibman, S and Siemens, R, (eds.) *Companion to Digital Humanities II*. (pp. 420-439). Wiley-Blackwell: Oxford, UK.

White, Judith., 2017. *Culture heist: Art versus money*. Brandl & Schlesinger.

Wilson, Andrew Norman. 2011"Workers Leaving the Googleplex." AndrewNormanWilson.com.

Wilson-Barnao, Caroline. (2020) "Virtual zoos, museums and galleries: 14 sites with great free art and entertainment" The Conversation. 26 March. https://theconversation.com/virtual-zoos-museums-and-galleries-14-sites-with-great-free-art-and-entertainment-134153

Zeffiro, Andrea. 2019. "Digitizing labor in the Google Books Project: Gloved

fingertips and severed hands." *Humans at Work in the Digital Age.*
Routledge. 133-153.