

Towards Universal Access to Knowledge

The Invisible Labor of Digitizing

By Yoonhee Lee

With digitization technologies, libraries and archives are able to share reproductions of books, photographs, archival documents and other cultural material to a wider audience beyond their physical walls. The excitement around digitization has sparked mass digitization projects and numerous national and international initiatives, which celebrate the idea of universal access to knowledge and information. As initial excitement waned, however, criticisms of these projects began to emerge. In particular, library scholars raised critical questions about copyright, poor metadata and quality, the lack of transparency, and ideologies of technological liberation with the Google Books Search project (Conway; Hoffmann and Bloom; Leetaru; Thylstrup). In addition, archival scholars pointed to privacy concerns and the ethics of circulating archival material online (Kandiuk; Robertson). What is lacking in these discussions, however, is an examination of the labor of digitization. The book scanners whose traces can be seen in the fingers and hands captured in error have been addressed in Andrew Norman Wilson's *Workers Leaving the Googleplex* and *ScanOps* and Aliza Elkin's *Hand Job* zine; however, further discussion and research is needed.

Looking at who does digitization labor and the context in which this labor takes place, I hope to situate the labor of digitizing books and archival material within larger discussions of the growth of invisible work and knowledge work and its exploitation under neoliberal capitalism. I will argue that within an economy of precarity, libraries and archives, as neoliberal institutions, are increasingly devaluing the

Yoonhee Lee is a recent graduate of the Faculty of Information, University of Toronto where she studied archives and records management. She is currently a learning and curriculum support librarian at the University of Guelph.

Keywords: mass digitization, labor, knowledge work, neoliberalism

labor of digitization through the use of the unpaid and underpaid labor and outsourced labor. With a closer examination of outsourced labor, I highlight how mass digitization involves factory-like industrial labor to produce knowledge commodities. In doing so, I demonstrate how the forces of neoliberalism, with its ideology of productivity and efficiency, have permeated into the realm of public service, knowledge, and cultural heritage.

Digitization as knowledge work

Digitization is a costly and labor-intensive endeavor for libraries and archives, but it (re)produces digital surrogates with not only immense cultural and research value, but market value as well. Digitization creates new knowledge products, which generate large investments (Conway 55). While some digitized products are available on the open web, others are only accessible through paid-subscriptions. Digitized collections increase the visibility and viability of knowledge organizations, helping to secure further funding in a climate of austerity (Moravec 187). In this market, users of digitized collections can be understood as consumers who have increasing demands and expectations for digitized material (Cifor and Lee 13). These users extract data from these digitized collections to produce new knowledge for the neoliberal academy.

Under the purview of knowledge, culture, and heritage industries, digitization is often associated with a public service or good. Instead of viewing digitization work through a lens that valorizes this labor, I position it as a form of exploited knowledge work under neoliberal capitalism. As Mosco and McKercher argue, the scope of knowledge work varies, from narrow interpretations that focus on the mobile creative class to expansive interpretations that encompass all workers involved in the creation and circulation of knowledge products. In defining knowledge work, Mosco and McKercher draw on Fritz Malchup's typology of knowledge producers, which consists of creators, analyzers, interpreters, processors, transformers, and transporters (30-31). Within this framework, digitization labor includes all strata of knowledge work from software engineers who design digital collection platforms and systems, librarians and archivists who coordinate and interpret projects, to the scanners and technicians who process and transform digital surrogates.

Often, however, this labor is invisible to the users of digital collections. While researchers interact with archival staff when accessing physical collections, digital archival environments render these workers and their work invisible (Moravec 191). In particular, those who process and create the digital reproductions through scanning and metadata creation are hidden. Poster, Cherry, and Crain

conceptualize invisible labor as work that is “overlooked, ignored, and/or devalued by employers, consumers, workers, and ultimately the legal system itself.” (6). This framework views multiple forms of labor across class and social hierarchies that are invisible within capitalism, including call center workers, crowdsourced engineers, clerical workers, gig workers, and migrant farm workers. The labor of digitization, I argue is a similar invisible form of labor, which is devalued by employers (libraries and archives) and consumers (users) through the use of contingent underpaid and unpaid labor and increasingly outsourced labor. In viewing digitization under this umbrella, the ways in which capitalism has infiltrated the public sector can be examined. Although much of digitization labor is fueled by dedication to access and preservation of knowledge, as Ursula Huws argues, through standardization and commodification, capital transforms this labor of love and public service to exploited labor. Taking inspiration from scholarship that makes visible the labor of knowledge workers, such as Sarah Roberts’s research on commercial content moderators, I foreground and highlight the devalued and hidden labor of digitization.

Hidden underpaid and unpaid labor

Libraries and archives have not been immune to the demands of neoliberal capitalism, which has created an increase in precarious work, characterized by job insecurity, temporary and nonstandard work (Poster, Cherry, and Crain 11). Like other institutions, libraries and archives are turning to temporary contractual labor and unpaid internships. This practice is especially acute in digitization projects, which are seen as an added responsibility to overwhelming workloads. In response to declining government funding and the rhetoric of More Product Less Process (Greene and Meissner), libraries and archives rely on project-based grants to create term-limited contractual positions. In “Implications of archival labor”, Stacey Williams points to how these grant funded workers are devalued through the lack of security and benefits, and excluded from the work of the permanent staff. Often due to conditions of the grant, these contract archivists are unable to participate in professional endeavors or participate in the decision-making processes of the larger department or organization (Williams). While the knowledge products are expected to be preserved and accessible for a long time, digitization laborers can only expect a few months or years of contractual work. Due to this compression of time, project archival work often results in the intensification of work and feelings of isolation and insecurity (Davis, Mattson, McNally and Reynolds). Moreover, similar to the academic labor market (Huws 77-78), precarity has resulted in increased competition, where one’s labor

is easily replaced. This transitory nature of digitization work adds to the devaluation of the labor performed—only the knowledge product, the digital surrogate, remains visible.

In an effort to reduce further costs and maximize efficiency, libraries and archives are also utilizing underpaid and unpaid student labor for digitization projects. Many grants available to institutions are reserved for student employees, framed in the rhetoric of providing valuable training and experience. In Canada, many heritage organizations rely on the Young Canada Works grants which provides wage subsidies for student summer employment. The program is designed to increase students' employability and help host organizations "maintain their operations in key functions with a skilled workforce" (Government of Canada). The ubiquity of these grants reveal how student labor is built into the infrastructure of heritage organizations, which rely on the knowledge, technical skills, and cheap labor of students for digitization work. Student practicums and internships, embedded into library and information school's curricula, also utilize unpaid labor. This free labor is framed as an experiential learning opportunity and a contribution to the goal of more equitable, universal access to knowledge and cultural heritage. While individuals benefit from these opportunities, the reliance on unpaid or underpaid labor reinforces a system, where digitization labor is increasingly deskilled and devalued, despite generating value, prestige, and investment for these organizations (Wildenhaus).

Hidden outsourced labor and partnerships

Another consequence of neoliberalism's infiltration into libraries and archives is the outsourcing of digitization labor under the guise of partnerships. Libraries and archives are encouraged to find cost-effective solutions through public and private partnerships. Public-private partnerships exemplify the ways in which market logic permeates the public sector (Hall 715). This neoliberal strategy can be seen in Canada's Library and Archives (LAC) approach towards digitization, where partnerships to increase LAC's digitization capacity are stressed (Library and Archives Canada). One major public-private partnership was the agreement with Ancestry.ca, one of the largest commercial vendors for family genealogy. In this partnership, Ancestry.ca digitizes and indexes genealogical records held by LAC, and in return Ancestry.ca is able to make these records available to its member subscription base (Library and Archives Canada). Similarly, the recent National Heritage Digitization Strategy recognized the commercial popularity of genealogical records, noting that it would most likely attract private partners. In these strategies, the language of outsourcing is avoided, but rather hidden as collaborative partnerships.

In this way, digitization labor is hidden under the cost-effective service provided by an external partner organization.

To reduce costs some outsourcing partnerships reinforce systems of inequalities. Like other forms of knowledge work that are being deskilled, the routinization of digitization labor make it vulnerable to global outsourcing (Mosco and McKercher 36). Not surprisingly, libraries and archives have outsourced digitization labor abroad. For example, Carnegie Mellon's Million Books Project consisted of a global partnership with India, China, and Egypt, with India and China providing the "manpower" for scanning and indexing. While this global partnership addressed the need for digitized non-English language books, it also took advantage of cheap global labor. In addition to global inequalities, some organizations have found cheap domestic labor through the prison-industrial-complex (Cifor and Lee 14). According to a *Mother Jones* article, the genealogical records of the Mormon church have been digitized and indexed by volunteer prison inmates in Utah (Bauer). The article also pointed to how the Utah government has utilized prison labor to scan the archival documents of the Department of Facilities and Construction Management (Utah Correctional Industries). Prison labor is normalized to the extent that an American Library Association publication presented multiple examples of public libraries utilizing the digitization labor of Oklahoma Correctional Industries as a successful story of overcoming limited resources (Caro 9-10). This example highlights how neoliberal forces have devalued the labor of digitization and the people who perform this labor in the name of efficiently and rapidly expanding access to knowledge.

Mass digitization factories

Although the conditions in which these outsourced digitization laborers work in are not visible, two prominent mass digitization partnerships, Google Books Search and the Internet Archive may provide some insight. As market demands for digitized material increased, the need to digitize at scale emerged. Google Books Search offered libraries and archives a technological solution, but their machines still required human operators to turn pages. Reportedly, the Google book scanner machine could digitize at the rate of 1,000 pages per hour, but this output depends on the machine-like work of the human book scanner (Somers). According to Wilson, these yellow-badged workers, predominantly people of color, were excluded from the privileges afforded to standard Google workers, such as the use of cafes, bikes, and shuttles. Moreover, Wilson describes the poor working conditions of low wages, high turnover rate, and a behavioral point system—an image that contradicts the Google Mountain View

campus as a place of creativity, flexibility, and innovation.

It is not surprising that Google as a for-profit company exploits its workers, demanding industrial outputs of digitized knowledge products. However, this form of labor is replicated in the not-for-profit digitization initiatives, highlighting the extent to which neoliberal forces have structured nonprofits as well. The Open Content Alliance's (OCA) digitization project positioned itself as an alternative to the commercial Google Books project with more transparency and openness (Leetaru). However, the human and technical infrastructure of scanning is similar (Thylstrup 12). The Internet Archive, which administered the project, digitizes its clients' books at its scanning centers across North America.

Similar to Google's operation, the Internet Archive relies on the labor of relatively low-waged human scanners. According to an Internet Archive job posting from 2006, at the San Francisco scanning center, part-time book scanners made \$11/hour; another source reports an average hourly wage of \$17/hour (Murrell 73). The job posting stressed the need for tolerance for repetitive tasks, attention to detail, and a basic knowledge of computers and email. In their dissertation research, Mary Murrell contends that digitization requires the technical skills and judgement of humans to produce quality images, as books "often violate the standards assumed in industrialized workflows" (72). These book scanners are tasked with not only performing rote, mechanical labor, but skilled labor as well; however, they are not acknowledged or valued for their contributions. While the Internet Archive book scanners likely work in better conditions than Google scanners, the work still involves low-paid, repetitive mechanical labor performed at a rapid pace. At one point, the Toronto scanning center consisted of human scanners turning pages in two shifts from 8:30 AM to 11:00 PM (Hutchinson). An article in *The Toronto Star*, described the project as an "industrial assembly line," questioning the company's portrait of the operation as medieval monks and scribes (Calami). Arguably, the Internet Archive's digitization operation resembles the factory, rather than the medieval scriptorium. This image of industrial labor reveals how nonprofit organizations under neoliberalism replicate the exploitative labor of businesses to remain competitive for their client partners.

The Internet Archive's book scanners, however, value their work and take pride in their contribution to open knowledge. According to Brewster Kahle, founder of the Internet Archive, most scanners are college graduates who "just love books and want to see them live on." (Kahle and Vadillo 2). The employees themselves have expressed their enjoyment of working with books in various newspaper interviews. One employee who was laid off after three years, claimed "It was the first time I felt I contributed to society" (Casey). Another employee

with an English literature degree described the book scanning job as one of the best they've had (Kesmodel and Vara). Despite the repetitive nature of the work, many Internet Archive scanners are passionate about their role in bringing out the organization's mission of universal access to all knowledge. This passion even surprised Kahle and the scanning centre supervisors; they assumed that no one would be able to tolerate repetitive work for long periods of time (Murrell 73). As this passion makes scanners more productive and efficient, it is codified in the job description: "love of books a plus" (Internet Archive). In positioning digitization work as a labor of love, the Internet Archive is able to mask the factory-like conditions of labor, where employees are expected to perform at a machine-like pace with minimal job security. In fact, the book scanning positions have turned into volunteer or intern positions, promising volunteers a "chance to help bring digital knowledge to others both near and far!" (Internet Archive). The mission to bringing about universal access to knowledge conceals the infrastructure of factory-like conditions of labor.

Ethical access to knowledge

The exploitative working conditions surrounding digitization stand in stark contrast with the ideologies of universal access and social justice touted by libraries and archives. As Cifor and Lee argue, the archival field has yet to grapple with how neoliberal structures are reproducing inequalities and devaluing archival labor. However, there are growing critical discussions around digitization labor and initiatives to build ethical labor practices while advancing access to knowledge. One example is the Latin American Digital Initiative (LADI), which advocates for community-oriented partnerships. Alpert-Abrams, Bliss, and Carbajal outline the consciously anti-colonial and anti-neoliberal approach they took in carrying out a digital initiative of the LLILAS Benson Latin American Collection library. LADI utilizes the post-custodial model of archives, in which LLIAS Benson provides digital stewardship of archival material held by partner organizations and archives in Latin America. In this framework, the outsourcing of digitization labor is recognized as a way to direct funds to community partners in Latin America, providing fair compensation, open communication, and training and support. Alpert-Abrams, Bliss, and Carbajal see their efforts as a way of resisting and critically reflecting on neoliberal archival practice. In this way, these partnerships are not centred on cost-reduction, but true collaboration and equal partnership in stewarding access to knowledge.

Another way forward is to draw on the work of digital humanities scholars, practitioners, and students who are discussing labor issues within this emerging field. Digital humanities scholars are recognizing

how students are often tasked with the processing of humanities data, which includes scanning and proofing digitized texts, but rarely are given opportunities to participate in scholarly analysis or project design (Anderson et al.). Although digital humanities labor, and similarly digitization labor, involves design, coordination, policy development, and other skilled labor, students are often relegated to routine tasks, such as scanning and metadata entry. The UCLA Digital Humanities program created a Student Collaborators' Bill of Rights which defines this type of digitization work as "mechanical labor," which students should be compensated for (Di Pressi et al.). The Bill of Rights states principles and values involving student labor and collaboration that should guide digital humanities projects. The library and archives field too can explicitly address unpaid student labor through the development of a bill of rights and advocate for paid internships (Wildenhaus).

Conversations around archival labor is growing. For example, recently, the Issues and Advocacy section of the Society of American Archivists launched a survey to gather data about temporary labor (@courecore). The Digital Library Forum (DLF) has also established a Working Group on Labor in Digital Libraries and has published a research agenda centred on gathering more information about organized labor, different categories of labor, and experiences of laborers within digital libraries. Collectively, these professional associations, working groups, and organizations can advocate for changes in library and archive labor practices. Through critical discussion of how forces of neoliberalism and capital has created a system of precarious underpaid, unpaid, and outsourced labor, libraries and archives can imagine ways of resisting these forces and advocate for the value of the labor of digitization. In this way, libraries and archives can build greater access to knowledge and community heritage, while also building equitable and ethical infrastructure for knowledge work.

Works Cited

- Alpert-Abrams, Hannah, et al. "Post-Custodial Archiving for the Collective Good: Examining Neoliberalism in US-Latin American Archival Partnerships." *Journal of Critical Library and Information Studies*, vol. 2, no. 1, 2019.
- Anderson, Katrina, et al. "Student Labor and Training in Digital Humanities." *Digital Humanities Quarterly*, vol. 010, no. 1, Feb. 2016.
- Bauer, Shane. "Did an Inmate Digitize Your Genealogical Records?" *Mother Jones*, Aug. 2015, www.motherjones.com/politics/2015/08/mormon-church-prison-geneology-family-search/.
- Calami, Peter. "Archivists Embrace Digital Page." *The Toronto Star*, 16 Apr. 2007, https://www.thestar.com/news/2007/04/16/archivists_embrace_digital_page.html.
- Caro, Susanne. *Digitizing your collection: public library success stories*. American

Library Association, 2016.

- Casey, Liam. "Toronto Online Book Archive Forced to Fire 75% of Staff." *The Toronto Star*, 8 July 2011, https://www.thestar.com/news/gta/2011/07/08/toronto_online_book_archive_forced_to_fire_75_of_staff.html.
- Cifor, Marika, and Jamie A. Lee. "Towards an Archival Critique: Opening Possibilities for Addressing Neoliberalism in the Archival Field." *Journal of Critical Library and Information Studies*, no. 1, Jan. 2017, <https://doi.org/10.24242/jclis.v1i1.10>.
- Conway, Paul. "Digital Transformations and the Archival Nature of Surrogates." *Archival Science*, vol. 15, no. 1, 2015, pp. 51–69.
- Davis, Rosemary, Mattson, Rachel, McNally, Elliot, and Alison Reynolds. "Grant Cycles, Deadlines, and Labor Advocacy: The Changing Work of Project Archivists." Mid-Atlantic Regional Archives Conference, April 2016, Pittsburg, PA. Conference Presentation.
- @courcore (Courtney Dean). "Happy to announce that the @ArchivesIssues survey on temp labor in archives is finally live. We hope it will serve as a good starting place and provide some data to help ground recent conversations. Please share widely! <http://tinyurl.com/templaborsurvey>." Twitter, 19 April 2019, twitter.com/courcore/status/1102665652412661760.
- Di Pressi, Haley et al. "A Student Collaborators' Bill of Rights." *HumTech*, 2019. humtech.ucla.edu/news/a-student-collaborators-bill-of-rights/.
- DLF Working Group on Labor in Digital Libraries. "Research agenda: Valuing labor in digital libraries." wiki.diglib.org/Labor.
- Elkin, A. *Hand Job* zine. alizaelk.in/digitize/.
- Government of Canada. "Employer guide 2019-2020 - Young Canada Works." *Young Canada Works*, www.canada.ca/en/canadian-heritage/services/funding/young-canada-works/employers/employer-guide.html.
- Greene, Mark, and Dennis Meissner. "More Product, Less Process: Revamping Traditional Archival Processing." *The American Archivist*, vol. 68, no. 2, 2005, pp. 208–263.
- Hall, Stuart. "The Neo-Liberal Revolution." *Cultural Studies*, vol. 25, no. 6, Nov. 2011, pp. 705–28. *Crossref*, doi:10.1080/09502386.2011.619886.
- Hoffmann, Anna Lauren, and Raina Bloom. "Digitizing Books, Obscuring Women's Work: Google Books, Librarians, and Ideologies of Access." *ADA: A Journal of Gender, New Media, and Technology*, no. 9, May 2016, doi:10.7264/N3BC3WTH.
- Hutchinson, Alex. "Human Resources: The Job You Didn't Even Know You Had." *The Walrus*, 12 Mar. 2009, thewalrus.ca/human-resources/.
- Internet Archive. "Volunteer positions." 9 February 2014, web.archive.org/web/20140209074311/https://archive.org/about/volunteerpositions.php.
- . "Job opportunities in the Open Access Text Archive." 21 July 2006, web.archive.org/web/20060721224623/http://www.archive.org/about/textsjobs.php#bookScanner.
- Kahle, Brewster, and Ana Parejo Vardillo. "The Internet Archive: An Interview with Brewster Kahle." *19: Interdisciplinary Studies in the Long Nineteenth Century*, vol. 2015, no. 21, Dec. 2015. doi:10.16995/ntn.760.
- Kandiuk, Mary. "The Rhetoric of Digitization and the Politicization of Canadian Heritage." *Library Trends*, vol. 65, no. 2, 2016, pp. 165–79. doi:10.1353/lib.2016.0029.

- Kesmodel, David, and Vauhini Vara. "Building an Online Library One Volume at a Time." *Wall Street Journal*, 10 Nov. 2005.
- Leetaru, Kalev. "Mass Book Digitization: The Deeper Story of Google Books and the Open Content Alliance." *First Monday*, vol. 13, no. 10, Oct. 2008. doi:10.5210/fm.v13i10.2101.
- Library and Archives Canada. "Digitization at Library and Archives Canada." 8 March 2018, www.bac-lac.gc.ca/eng/about-us/about-collection/Pages/digitization-lac.
- . "Library and Archives Canada partners with Ancestry.ca." 10 November 2008, www.collectionscanada.gc.ca/whats-new/013-362-e.html.
- Moravec, Michelle. "Feminist Research Practices and Digital Archives." *Australian Feminist Studies*, vol. 32, no. 91–92, 2017, pp. 186–201.
- Mosco, Vincent, and Catherine McKercher. *The Laboring of Communication: Will Knowledge Workers of the World Unite?* Lexington Books, 2008.
- Murrell, Mary E. *The Open Book: Digital Form in the Making*. 2012 University of California, Berkeley, PhD Dissertation.
- National Heritage Digitization Strategy. "The Strategy." nhds.ca/about/the-strategy/.
- Poster, Winifred, et al., editors. *Invisible Labor: Hidden Work in the Contemporary World*. University of California Press, 2016.
- Roberts, Sarah. T. *Behind the screen: the hidden digital labor of commercial content moderation*. 2014. University of Illinois at Urbana-Champaign, PhD Dissertation.
- Robertson, Tara. "Not All Information Wants to Be Free: The Case Study of On Our Backs." *Applying Library Values to Emerging Technology: Decision-Making in the Age of Open Access, Maker Spaces, and the Ever-Changing Library (Publications in Librarianship #72)*, American Library Association, 2018, pp. 225–239.
- Somers, James. "Torching the Modern-Day Library of Alexandria." *The Atlantic*, Apr. 2017. *The Atlantic*, www.theatlantic.com/technology/archive/2017/04/the-tragedy-of-google-books/523320/.
- Thylstrup, Nanna Bonde. *The Politics of Mass Digitization*. The MIT Press, 2018. *Utah Correctional Industries Website*. <https://uci.utah.gov/index.php/business-operations/data-scanning-services>. Accessed 5 Apr. 2019.
- Wildenhaus, Karly. "Wages for Intern Work: Denormalizing Unpaid Positions in Archives and Libraries." *Journal of Critical Library and Information Studies*, vol. 2, no. 2, 2018, <http://libraryjuicepress.com/journals/index.php/jclis/article/view/88>.
- Williams, S. "Implications of Archival Labor." *On Archivy*, 11 April 2016, medium.com/on-archivy/implications-of-archival-labor-b606d8d02014.
- Wilson, Andrew Norman. *ScanOps*. www.andrewnormanwilson.com/ScanOps.html.
- . *Workers Leaving the Googleplex*. www.andrewnormanwilson.com/WorkersGoogleplex.html.