

A Brief History of Crowdsourced Digital Publishing at LibriVox.org

Amelia Chesley, Purdue University

This paper traces the digital publishing history of the audiobook archive LibriVox. Founded in 2005, LibriVox.org is a global community of volunteers engaged in the mission of recording all public domain texts as free audiobooks. Importantly, LibriVox relies on the legal existence of unrestricted material that anyone can wonder at and create with: the shared wealth of the public domain. In a time when corporations are investing relentlessly in centralizing content and constraining individuals' freedoms to engage with and share that content, the decentralized and distributed model of LibriVox consistently works to preserve crucial modes of openness and access not only in their finished product, but also in their workflow and production. LibriVox serves as a potential guide for other collaborative publishing projects, especially those built by multicultural and multilingual groups of volunteers, and can also teach us, as scholars and keepers of culture, how to digitize and share in ethical and more sustainable ways.

In August of 2005, Canadian writer and web developer Hugh McGuire sent out a few emails asking friends and acquaintances if they might be willing to collaborate on a podcast recording of Joseph Conrad's 1907 novel *The Secret Agent*. Twelve people joined the project. McGuire posted their recordings of each chapter to a new blog, and that was the beginning of LibriVox—a volunteer-led effort to produce audio versions of public domain texts. In the twelve years since its founding, LibriVox has facilitated and fostered the production of almost eleven thousand free, public domain audio editions, read and recorded by more than eight thousand volunteer readers in 94 different languages.

Inspired by the open-source software movement and emerging forms of crowdsourced content-creation, the globally-distributed community of LibriVox participates in and sustains an open and highly inclusive workflow, welcoming all potential readers and as many versions of any public domain text as volunteers want to create. There is no strict organizational hierarchy at LibriVox—any volunteers willing to propose, manage, and complete audiobook projects or other related public domain projects are encouraged to do so, and others help lead or collaborate on these projects as they are willing and able. In a time when corporations are investing relentlessly in centralizing content and constraining individuals' freedoms to engage with and share that content, the decentralized and distributed model of LibriVox works to preserve crucial modes of openness and access not only in its finished product, but also in its workflow and production processes.

LibriVox volunteers have developed a flexible-yet-resilient system of open, collaborative publishing. In this piece, I trace and examine how the project's volunteers have drawn on and adapted pre-existing technologies and infrastructures such as podcasting, wikis, and open-source software in order to fulfill the LibriVox mission of transforming all public domain texts into free, accessible audiobooks. Building on this history, I also investigate how volunteers manage and negotiate their ongoing collaborative work in the face of persistent questions and controversies stemming primarily from the legal realities of donating one's voice into the public domain and from the social realities of relying on volunteer labor.

The workflow of LibriVox's ambitious project faces continual moments of recalibration and re-justification to account for the challenges of working within a global community of volunteers while also navigating the expectations of millions of listeners. In documenting and discussing the ways LibriVox's initially *ad hoc* techne has settled into protocol, I highlight the valuable rhetorical work done by these volunteers, not only in terms of how it supports the LibriVox project itself, but also for what that work teaches us about online commons-based collaboration.

Because the LibriVox project is so open, I have been able to engage with its community and artifacts as both researcher and as participant. Since January 2016, I have been actively volunteering in the forums as reader, proof listener, and project coordinator. I am also a member of the LibriVox Readers & Listeners Facebook Group and occasional contributor to the LibriVox Community Podcast. As participant and researcher in these dynamic spaces, I draw on a combination of ethnography and

autoethnography (Boellstorff, Nardi, Pearce, & Taylor, 2012), taking care to share my status as a researcher when relevant and to practice reciprocity. Through direct experience with the online community of LibriVox and through analysis of its public production processes, I have begun to illuminate some of the many-layered technosociocultural foundations upon which crowdsourced digital publishing endeavors like LibriVox stand.

Crowds and Networks of Social Peer-Production

Digital, networked technologies have made possible more inclusive and more accelerated modes of collaborating, creating, sharing, and remixing. Many scholars have recognized that digital technologies and their contexts afford a priceless increase in openness and public access (Benkler, 2006; Lanham, 2007; Bollier, 2008; Shirky, 2010; Hayles, 2012; Potts, 2015). Networked technologies combined with open philosophies allowing free (or freer) circulation of information almost inevitably lead to greater transparency, efficiency, and democracy in terms of cultural production and access (Benkler, 2006; Shirky, 2010). Demand for digital curation skills is growing rapidly, along with recognition of the economic value and societal benefits such skills can provide (National Research Council, 2015). Crowdsourcing projects and commons-based, peer-production models allow anyone—from novice to expert—to join and sustain grand social efforts to curate, digitize, publish, and share content across many disciplinary contexts. An example of such inclusive public action can be seen in how “information resources such as repositories, databases, and archives are increasingly being crowdsourced to professional and nonprofessional volunteers,” (Rotman, Procita, Hansen, Parr, & Preece, 2012, p. 1092).

Popular and valued crowdsourced initiatives like Project Gutenberg and Wikipedia (both precursors to LibriVox) have facilitated and encouraged consistent public participation in knowledge production (Benkler, 2006; Jemielniak, 2014). Indeed, the participatory culture of LibriVox is similar in some ways to that of Wikipedia—all listeners, users, and bystanders are invited to contribute in small increments, and their efforts are included in the ever-growing collection of audiobooks. However, the nature of LibriVox’s mission means that plurality and multiple voices are privileged and showcased in ways that Wikipedia, with its pursuit of ever-increasingly-refined consensus, does not tolerate. All three projects have made use of crowdsourced commons-based production models to engage in the ongoing digitization, preservation, and circulation of human knowledge and culture.

The influence of crowdsourced digitization and public knowledge-making efforts have also formed the basis of much research and critique in writing studies and elsewhere (Rosenzweig, 2006; Purdy, 2009; Kill, 2012; Graban, Ramsey-Tobienne, & Myers, 2015; Yancey, 2016). These scholars raise questions about the effect digital knowledge-making and curation should have on the ways we learn and teach effective communication. Much of the value of social production and digitization stems from the collaborative learning opportunities these practices allow and the complex digital communities that emerge around the activities of sharing knowledge (Wenger, White, & Smith, 2009). Similarly, Miles Kimball (2016) recognizes the human drive to share experience using whatever means possible: “At no time in human history have more people [...] been involved in helping to accommodate each other to technology,” (p. 12). Precisely this kind of technologically-enabled accommodation and proliferation of shared resources happens within the LibriVox community as members learn, invent, and practice how to navigate the challenges of global online collaboration and accomplish the work of audiobook production.

LibriVox as Meshwork

The history and activity of LibriVox is distributed across several online spaces and across a vast global network of individuals. Many of the procedures and policies LibriVox has evolved are scattered among various digital records, metadata, and audio files, forming a living meshwork of archived hypertext and human voices. More than a network of joined nodes, a meshwork is constituted by inhabited entanglement where acting, doing, and being take place (Ingold, 2007, p. 80; 2011, p. 63). As part of LibriVox, volunteers inhabit various roles; they work as curators, voice artists, project managers,

audio producers, copyright sleuths, digital content managers, mentors and instructors, researchers, translators, dialect coaches, and/or technical writers. As they fill these roles, volunteers leave traces of their work across the ever-changing digital meshwork.

LibriVox hosts and cultivates the use of a variety of connected tools and spaces as part of its mission to record as many public domain audiobooks as possible. Artifacts encountered thus far in my preliminary map of the technosociocultural landscapes of LibriVox include:

- the LibriVox forums and message boards
- the LibriVox Management Dashboard (for back-end project management; also commonly called the section compiler)
- audio recording and editing software (Audacity, GarageBand, Camtasia, and other programs)
- a wide variety of computers, microphones, headsets, adaptors, and other hardware for recording
- the LibriVox website (including the blog)
- the LibriVox wiki (also called Guides for Listeners & Volunteers)
- links to instructional resources hosted elsewhere (YouTube, other forums, etc.)
- the catalog spaces at LibriVox.org and Archive.org
- the LibriVox Community Podcast archives
- social media presences on Twitter, Facebook, Reddit
- posts from LibriVox volunteers concerning the participatory roles they've defined for themselves

Together these documents, technologies, and spaces begin to define a volunteer-driven audio digitization or publishing network. Harnessing these connections, thousands of volunteers read, produce, and disseminate free audiobooks for millions of listeners, who in turn have the opportunity to become volunteers themselves. Across the LibriVox sites, catalog spaces, social media accounts, and elsewhere, listeners access and engage with finished LibriVox recordings in a variety of ways. The primary LibriVox files hosted at Archive.org have collectively received more than 600 million views (LibriVox Free Audiobook Collection, 2006). Additionally, there are many mirrors and copies created for use via third-party websites and apps. LibriVox recordings have also been re-distributed via radio, television, CDs, podcast segments, and YouTube.

LibriVox's Technosociocultural Histories

A constellation of blog posts, discussion forum threads, and podcast episodes provides detail about the beginnings of the LibriVox project. I first began listening to the LibriVox Community Podcast out of general personal interest, but soon discovered how rich and detailed this content would be for the purposes of my research. Archived podcast episodes provide unique and intimate audio-snapshots of LibriVox over time. In this once weekly and now sporadically produced podcast, volunteers take turns sharing news from the forums, celebrating their work and the work of other volunteers, and reflecting on their participation in the project. Upon realizing the insight these records offered, I downloaded the full archive, spanning ten years of LibriVox's existence (September 2006–December 2016).

For the purposes of my present exploration, I've drawn evidence for my discussion from a large sample of these LibriVox Community Podcast episodes and from specific forum discussions referenced in the podcast. I listened to and created annotations for 110 of 144 episodes, comprising about five years of community history (from September 2006–August 2010). To supplement my annotations, I used the search function on the LibriVox forums to locate the show notes for each episode and any other specific conversations and announcements discussed or referenced in the podcast.

Using what I gleaned from these episodes and forum threads, I began constructing a timeline and overview of LibriVox's history. The following sections briefly describe salient developments from the first ten years of LibriVox and build toward a brief discussion of two specific controversies that regularly percolate through the community.

Any Recording is Better than None

Many core LibriVox spaces emerged, at least in prototype, relatively quickly. Founder Hugh McGuire himself spearheaded the first audiobook projects at LibriVox, but soon realized he couldn't run everything. Eager volunteers with the requisite experience gradually donated server space, coding talents, and time to the project, all drawing on expertise from other arenas such as computer programming or library and information science. Founding members of the project reminisce about the earliest wild west days of collaborating via email and one simple blog, librivox.blogspot.com, for sharing and publishing their audio files (Gonzalez, 2012). A stand-alone domain name, librivox.org, and official LibriVox forums were established, and volunteers soon populated these new forums with orderly sections and helpful structure. A LibriVox wiki was created for consolidating and sharing instructional and policy information with a growing contingent of multicultural volunteers. Once the population of volunteers grew beyond the first handful, email was no longer a feasible way of distributing files. Temporary file-sharing websites like yousendit.com became a standard method of getting files from reader to coordinator and eventually to cataloger. By October 2005, librivox.org sported a new website design, and a catalog database system was under construction.

At first, the quality of audio files was mixed, and there was no vetting process to assure a consistent, pleasant experience for listeners. Referring to his very first LibriVox contribution, Hugh McGuire reframes its poor quality into a symbol of how even a novice can make something useful and share it with the world (Samuels, 2007; Gonzalez, 2012). Eventually, technical specifications were established and readers were encouraged to ensure their recordings met those specifications. The practice of proof listening audio files before cataloging was introduced in January 2006, and gradually became a requirement for all projects. However, proof listeners were instructed to never critique reading style, pacing, pronunciation, or any other subjective quality.

Five core LibriVox values emerged, and remain central to the project:

- LibriVox is a noncommercial, nonprofit and ad-free project.
- LibriVox donates its recordings to the public domain.
- LibriVox is powered by volunteers.
- LibriVox maintains a loose and open structure.
- LibriVox welcomes all volunteers from across the globe, in all languages.

These LibriVox values have guided and shaped the project since its early days. While these principles clearly and succinctly delineate and support the LibriVox mission, they also constrain the project in important yet sometimes controversial ways.

Perennial Controversies

From time to time, the core values and principles of LibriVox conflict with reader and listener expectations, spurring questions, concerns, and debate about the value and/or costs of established policy. Two significant issues and their associated controversies have been especially common throughout the history of LibriVox:

- The issue of whether to attach licenses to LibriVox output, instead of allowing unrestricted use of all published recordings, including commercial re-uses; i.e., the “How dare someone try to sell my volunteer work?” controversy.
- The issue of whether to invite listener ratings or critical feedback for volunteer readers, rather than guarding volunteers against such criticism and accepting all understandable recordings into the catalog regardless of reading style or language ability; i.e., the “Why can't everyone pronounce things the way I prefer?” controversy.

Both of these issues and the ways the LibriVox community approaches the debates surrounding them are discussed briefly below.

Commercial Re-use

LibriVox’s existence relies on willing participants, and also on many pre-existing systems, technologies, and partner organizations. Foremost among these scaffolding systems is the legal existence and precedent of the public domain; without this trove of unrestricted, publicly available, free-for-reuse material, the LibriVox mission would not have the ambitious scope it does. The principles of LibriVox mean that all LibriVox contributions remain in the public domain in the US, where anyone is legally free to do whatever they like with them. No monetary compensation or added copyright attends the time and labor LibriVox volunteers put into their projects. As a result of the public domain principle, anyone is legally free to remix, repackage, or even sell readers’ recordings if they wish.

When volunteers discover repackaged versions of their LibriVox work in other contexts—burned onto CDs for sale on eBay, matched with art or video on monetized YouTube channels, or downloadable for a subscription fee on another website—they often bring their concerns about this seemingly unethical re-use back to LibriVox. Some ask what can be done about the issue, wondering why the community doesn’t license their work using Creative Commons licenses. Others suggest various steps LibriVox could take to protect their catalog. In response, those more familiar with LibriVox policy will remind the upset volunteers that public domain means just that: anyone can repurpose this work for anything, and LibriVox is financially unprepared to challenge such re-use even if it wanted to control the destinations of its products. Accepting this fact is a firm condition of participating with LibriVox, and new volunteers must understand this condition and be willing to donate their work before participating.

Unsolicited Criticism

LibriVox relies solely on volunteer labor and welcomes readers from any and all languages, with any understandable language ability, accent, or style of reading; this means listeners must accept a wide variety of recordings. Of course, not all reading styles are enjoyable to all ears, and listeners regularly approach the community with suggestions for greater quality control. LibriVox, however, recognizing the potential chilling effects of negative feedback, has long enforced a strict policy of “no unasked-for feedback or criticism.” (Mowatt, 2006; Hughes, 2007; Samuels, 2008; Gesine, 2010). Since judgments about readers’ pacing, cadence, tone, pronunciations, accent, pitch, and other stylistic elements are inherently complex and subjective, the LibriVox community has determined that all such efforts at quality control are not warranted in a volunteer-run project.

A corollary here is that no single reader, accent, or style will ever constitute the definitive performance of any particular text. Multiple readings are welcome and encouraged; if one listener disagrees with one reader’s interpretation, that listener is welcome to find another, or even to record their own.

Non-negotiable Navigations

LibriVox handles these and other controversial issues by appealing to their “prime directive” (as several volunteers half-jokingly, half-reverently term their central mission), which is simply to create free public domain audiobooks for the world. If a suggested modification to LibriVox’s policy or processes would clearly ease the process of freely recording and distributing audiobooks, the community may consider it. If the suggestion can’t be shown to help directly with that central mission, it is simply (and often quite firmly) dismissed.

Despite this clear and consistent evaluation process, the concerns about unethical commercial re-use and unsuitable reading styles, pronunciations, accents, etc., will likely never go away. Such controversies, and their regular reappearance at LibriVox, are a consequence of the core strength of the project: its openness. Volunteers new and old will continue to confront these and other common issues as they are rearticulated. A constant influx of new volunteers since 2005 has meant that the inexperienced and unassimilated unearth these controversies again and again. Each new volunteer, in joining and adjusting to the LibriVox community, must confront and negotiate her own principles within those of the project as

a whole. It takes patience and firmness from experienced volunteers to continually re-establish, re-explain, and re-argue the reasons behind LibriVox's policies every time they are challenged.

Future Investigation

Volunteers have recorded and preserved the history of LibriVox in ways that seem haphazard and disconnected, but also reflect the open structure and priorities of the project. The difficulty of accessing digital history via often transient web artifacts scattered across platforms means there may be persistent gaps in what I am able to discover. Interviews with volunteers will allow me to complete a fuller ethnography of these spaces, filling in gaps where the public records of LibriVox lack transparency. I also plan to explore a selection of completed LibriVox projects in greater depth, looking for more detailed evidence of how the community and its technosociocultural contexts have shaped it and its work over time.

Conclusions

With this review of what has made LibriVox into the popular and productive site it has become, I want to emphasize the importance and value of decentralized and distributed models like LibriVox for promoting and safeguarding crucial modes of ethical, resilient openness. The LibriVox community and publishing project has grown and evolved in surprising and wonderful ways that deserve further study. LibriVox's clarity of purpose and open, welcoming processes become potentially useful models for future collaborative, online media projects, and the implications of this successful, sustainable, commons-based, digital publishing model may help prompt important, democratizing shifts in the future of open scholarly publishing.

By actively engaging with a fertile intellectual, cultural, and technological commons, LibriVox volunteers have freely adapted the tools available to them within the constraints of existing legal, cultural, social, and technological systems. Built on a recognition of what crowdsourcing makes possible and on an appreciation for the cultural commons of the public domain, LibriVox provides opportunity for anyone to select and transform beloved old texts into sound, and to share the results across the web, adding their voice to a living, ever-expanding archive.

References

- Benkler, Yochai. (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven, CT: Yale University Press.
- Boellstorff, Tom, Nardi, Bonnie, Pearce, Celia, & Taylor, T.L. (2012). *Ethnography in virtual worlds: A handbook of method*. Princeton, NJ: Princeton University Press
- Bollier, David. (2008). *Viral spiral: How the commoners built a digital republic of their own*. New York, NY: New Press.
- Gesine. (2010, August 4). LibriVox community podcast episode #113: Thank you messages from listeners. LibriVox Community Podcast. Retrieved July 23, 2017, from http://ia600209.us.archive.org/17/items/librivox_community_2010/librivox_community_podcast_113.mp3
- Gonzalez, Robert. (2012). LibriVox community podcast episodes #126–130, LibriVox: The pioneer days, parts 1–5. LibriVox Community Podcast. Retrieved July 23, 2017, from http://wiki.librivox.org/index.php/Podcast_Links_and_Signup
- Graban, Tarez Samra, Ramsey-Tobienne, Alexis, & Myers, Whitney. (2015). In, through, and about the archive: What digitization (dis)allows. In Jim Ridolfo & William Hart-Davidson (Eds.), *Rhetoric and the digital humanities* (pp. 233–244). Chicago, IL: University of Chicago Press.

- Hayles, Katherine. (2012). *How we think: Digital media and contemporary technogenesis*. Chicago, IL: University of Chicago Press.
- Hughes, Kristin. (2007, November 1). LibriVox community podcast #60. LibriVox Community Podcast. Retrieved July 23, 2017, from http://ia902604.us.archive.org/21/items/librivox_community_2007/librivox_community_podcast_60.mp3
- Ingold, Tim. (2007). *Lines: A brief history*. New York, NY: Routledge.
- Ingold, Tim. (2011). *Being alive: Essays on movement, knowledge and description*. New York, NY: Routledge.
- Jemielniak, Dariusz. (2014). *Common knowledge? An ethnography of Wikipedia*. Stanford, CA: Stanford University Press.
- Kill, Melanie. (2012). Teaching digital rhetoric: Wikipedia, collaboration, and the politics of free knowledge. In Brett D. Hirsch (Ed.), *Digital humanities pedagogy: Practices, principles and politics* (pp. 389–405). Cambridge, UK: OpenBook Publishers. Retrieved July 23, 2017, from <http://www.openbookpublishers.com/product/161>
- Kimball, Miles A. (2016). The golden age of technical communication. *Journal of Technical Writing and Communication*, 0(0), 1–29.
- Lanham, Richard. (2007). *The economics of attention: Style and substance in the age of information*. Chicago, IL: University of Chicago Press.
- LibriVox. (2005–). Retrieved July 20, 2017, from <https://librivox.org>
- LibriVox. (2006–2016). LibriVox community podcast. Retrieved July 23, 2017, from http://wiki.librivox.org/index.php/Podcast_Links_and_Signup
- LibriVox Forums. (2005–). Retrieved July 20, 2017, from <https://forum.librivox.org>
- LibriVox Free Audiobook Collection. (2006–). About. Internet Archive. Retrieved July 20, 2017, from <https://archive.org/details/librivoxaudio&tab=about>
- Mowatt, Jim. (2006, December 14). LibriVox community podcast episode #14. LibriVox Community Podcast. Retrieved July 23, 2017 from http://ia902609.us.archive.org/6/items/librivox_community_2006/librivox_community_podcast_14.mp3
- National Research Council. (2015). *Preparing the workforce for digital curation*. Washington, DC: National Academies Press.
- Potts, Liza. (2015). Archive experiences: A vision for user-centered design in the digital humanities. In Jim Ridolfo & William Hart-Davidson (Eds.), *Rhetoric and the digital humanities* (pp. 255–264). Chicago, IL: University of Chicago Press.
- Purdy, James. (2009). When the tenets of composition go public: A study of writing in Wikipedia. *College Composition and Communication*, 61(2), 351–373.
- Rosenzweig, Roy. (2006). Can History be open source? Wikipedia and the future of the past. *The Journal of American History*, 93(1), 117–146.
- Rotman, Dana, Procita, Kezia, Hansen, Derek, Sims Parr, Cynthia, & Preece, Jennifer. (2012). Supporting content curation communities: The case of the encyclopedia of life. *Journal of the American Society for Information Science and Technology*, 63(6), 1092–1107
- Samuels, Cori. (2007, October 25). LibriVox community podcast #59: Approaching 1000 completed works. LibriVox Community Podcast. Retrieved July 23, 2017, from http://ia902604.us.archive.org/21/items/librivox_community_2007/librivox_community_podcast_59.mp3

- Samuels, Cori. (2008, April 24). LibriVox community podcast #79. LibriVox Community Podcast. Retrieved July 23, 2017, from http://ia902301.us.archive.org/28/items/librivox_community/librivox_community_podcast_79.mp3
- Shirky, Clay. (2010). *Cognitive surplus: Creativity and generosity in a connected age*. New York, NY: Penguin Group.
- Wenger, Etienne, White, Nancy, & Smith, John D. (2009). *Digital habitats: Stewarding technology for communities*. Portland, OR: CPsquare.
- Yancey, Kathleen Blake. (2016). Print, digital, and the liminal counterpart (in-between): The lessons of Hill's manual of social and business forms for rhetorical delivery. *Enculturation*. Retrieved July 20, 2017, from <http://enculturation.net/print-digital-and-the-liminal-counterpart>