

UC Davis

Library Staff Publications

Title

Project LEND Service Plan

Permalink

<https://escholarship.org/uc/item/9450t5m9>

Authors

Majors, Rice

Waibel, Günter

Mitchell, Erik

Publication Date

2025-04-01

Project LEND Service Plan

Project LEND working group

Rice Majors, UC Davis, principal investigator

Günter Waibel, California Digital Library, co-principal investigator

Erik Mitchell, UC San Diego, co-principal investigator

Peter Brantley, UC Davis

Rebecca Chambers, UC Berkeley

Michael Furlough, HathiTrust

Lin Li, UC Irvine

Micah May, Digital Public Library of America

Anne Marie Piper, UC Irvine

Roger Smith, UC San Diego

Molly Van Houweling, UC Berkeley

Neil Weingarten, UC Davis

Ange Zaytsev, HathiTrust

April 2025

Executive Summary

Project LEND was a two-year research project that looked at user needs in higher education in the United States for digital books, as well as the technology, collections, and legal dimensions of how best to meet those needs. Our intention was to foster a national conversation around these topics and to lay the groundwork for systems to be instantiated to better meet user needs.

Our research found great potentiality for increased coordination and interoperability between stakeholder entities (especially publishers, distributors, and libraries), who could leverage an open architecture to address a wide range of use cases in education and research contexts.

We identify a number of potential actions for publishers, distributors, and libraries, including:

1. Libraries to expand digital content with intentionality
2. Libraries and publishers to enrich metadata for under-described content
3. Publishers and distributors to enable broader usage of their content
4. Libraries to educate users on copyright law and ebook formats
5. Libraries to engage in the testing and development of new platforms

Our research also found that technology, user expectations, copyright and transformative use, and digital book access are all rapidly evolving. We especially note that user perceptions of what is possible are evolving as technological advances (such as the potentiality of artificial intelligence) inspire users and help them think beyond the present capabilities of digital book use.

Finally, we recommend further dialogue with users who engage in research modalities where they use a corpus of books (e.g., data science research, digital humanities) both to understand these users better and to determine the appetite for potential improvements.

“Digital book”
refers to book-like information available online, whether published natively as an ebook or converted into a digital format from a print format. When we mean commercially-available ebooks in particular, we will say **“ebook.”**

Background and History

Project LEND (Library Expansion of Networked Delivery) explores an increasingly obvious reality: that **users in higher education contexts make use of books in many different ways**, and that reading a book cover to cover is now just one of many uses that scholars seek to make of books.

The University of California Libraries have been focused on this issue for well over a decade, especially through participation in the Google Books partnership. The UC Libraries gained additional insight into user needs with regards to digitized books during the COVID-19 pandemic, when UC had access to HathiTrust's emergency temporary access service (ETAS). The pandemic functioned as an unplanned large-scale natural experiment, during which scholars pivoted largely to digital modalities of conducting research and education. Concurrent with the conclusion of ETAS, in 2022, the UC Libraries **completed an environmental scan** of users that surfaced many potential use cases that weren't currently possible or optimized, either because of technology, interpretation of copyright law, or available digital collections.

With generous support from the Mellon Foundation, Project LEND was initially launched to conduct systematic user research on the uses of digitized books and articulate pathways to supporting existing and emerging research and educational needs of users.

A foundational assumption of Project LEND is that a national conversation should be happening regarding the use of digital books to identify potential next steps for implementation. This informed our decisions about project leadership, which included faculty and senior library leaders from across the UC system as well as domain experts from HathiTrust and the Digital Public Library of America. We also set up a Library Community Group of senior leaders at other HathiTrust member libraries to check our assumptions and direction, helping us avoid

inadvertently designing something that worked principally for the University of California.

Our research was divided into four domains: user perspectives, legal context, technology, and collections. Given the range of use cases that had emerged in the UC Libraries' environmental scan, we centered the user throughout our work — not assuming that we already knew what users may want. This approach was validated as artificial intelligence and especially large language models (LLMs) made rapid advances even as we were engaging with users, who quickly surfaced new use cases that could apply to digital content.

Project LEND has culminated in three published work products: a **response** to the U.S. Copyright Office's Notice of Inquiry on Artificial Intelligence and Copyright, a user-research white paper, and this document — a service plan for potential future uses of digitized content.

Our **findings from user research** are far-ranging, as users brought up many topics in focus groups and interviews that were central to our research questions and others that were more orthogonal. This led to a decision to publish all of these findings in their own separate document, so that this rich user data can be digested by and acted upon by others.

This service plan document serves as the summative output of Project LEND, which we hope will set the stage for national services to arise that instantiate the various use cases for lawful use of digital books that the plan describes.

Our Vision for Digital Books

How will students, faculty, and researchers discover and use information in the future?

Will they find everything they need in one comprehensive place, or will they be forced to search a multitude of independent, often proprietary systems, uncertain if they've

uncovered the best resources? Will they have the flexibility to save texts and annotations, and access materials in a seamless, user-friendly environment that works across devices and platforms (including switching between devices)? Will researchers have confidence that their workflows comport with copyright law? Will they be able to engage with large, diverse datasets featuring accurate optical character recognition (OCR) and metadata? And will they have powerful computational tools at their disposal, such as text summarization, custom natural language processing, and data visualization options?

These critical questions are at the heart of Project LEND. **Project LEND's vision is that in the future users will be able to access a vast array of information in a streamlined, robust, and accessible environment.** This environment would allow users to search across all library assets; conduct full-text searches within books; create, save and share annotations; and engage with digital content in meaningful ways — all within a lawful framework. In the future envisioned by Project LEND, publishers and libraries will support access to comprehensive and accurate content from multiple sources and equip researchers with the tools they need to derive insights and generate new knowledge.

This vision emerges out of important lessons learned during the COVID-19 pandemic about both user behavior and the potential to improve research and education outcomes dramatically by leveraging the ubiquity of digital content. We can clearly see the impacts that a lack of coordination (among publishers, libraries, et al) have had on users. Improvements in metadata and available technology, together with Project LEND's robust data on user needs, position us to make effective improvements for the benefit of all.

Achieving the vision of Project LEND will require systemic change and collaborative effort. Publishers, distributors, copyright holders, users, and libraries share many of the same interests, including especially a shared interest in people robustly using digital books. Enabling more ways to use digital books will enhance, not erode, the benefits for each stakeholder group. Fostering an environment where significant collaboration occurs is an integral part of our shared success.

To realize this vision, libraries must act more strategically with regard to digital books — contributing their digitized content toward a shared good and being collectively thoughtful about what to prioritize for digitization. Collaboration between libraries is essential to building shared and accessible data, bibliographic metadata, and rights-related information — and to co-investing in robust shared technology solutions, avoiding the pitfalls of fragmented efforts and informal standards.

The Project LEND team emphasizes that the shift toward more open architecture and interoperable systems poses no necessary risk or tradeoff for publishers and copyright holders. This approach need not jeopardize business interests, revenue, or intellectual property. In this context, “open” refers to systems that seamlessly connect with one another, rather than to exclusively openly licensed content. This model aligns with maintaining robust security practices, safeguarding titles with digital rights management (DRM) software, and ensuring fair compensation for publishers and authors.

Interoperability will empower libraries to acquire more content from diverse sources, creating new market opportunities for authors, publishers, and distributors. Notably, major trade publishers, including the “Big Five,”¹ are already leveraging Open Publication Distribution System² (OPDS)-based systems to enable interoperability.

¹ That is, the five largest mainstream publishers in the United States market: Penguin Random House, HarperCollins, Hachette Livre, Macmillan Publishers, and Simon & Schuster.

² The Open Publication Distribution System (OPDS) is a syndication format for describing electronic publications, originally based on RDF and now available in a serialized format. OPDS catalogs enable the aggregation, distribution, discovery, and acquisition of electronic publications. OPDS catalogs use existing or emergent open standards and conventions, with a priority on simplicity.

Academic publishers and authors can similarly benefit by ensuring their titles are accessible within these interoperable frameworks.

If libraries, distributors, and publishers come together to pursue this vision — anchored by ideals of access, ease of use, and interoperability — we can build a future where users have more information and tools at their fingertips than ever before. We can enable new forms of inquiry, accelerate research, and expand knowledge.

Project LEND has mapped a path to a brighter future. Now it is up to us — libraries, distributors, and publishers — to take the next steps along this journey and lead the way toward a better, more integrated future for libraries.

Research Methodology

In 2023-2024, Project LEND engaged directly with users in higher education across the United States to gather data on the ideal digital book experience for supporting research and educational outcomes. We then reviewed the needs and wants of users to identify what was in scope for early phases of the development and maturation of a proposed service.

There were several recurring themes which users surfaced, including ease of discovery (including across providers), ease of access (including across providers and formats), and interoperability more generally. These themes led to a proposed architecture that would support goals of making it easier for users to interact with books from a variety of sources while supporting a wide range of use cases.³

In parallel with the user research, Project LEND examined key legal questions, potential technological pathways, and collections issues in

order to bring together these three areas with our user research findings and develop this proposed service plan for how to meet the needs and expectations of users.

Service Plan

Guiding principles

A number of principles guided our decisions in developing this service plan.

Our primary guiding principles are to **make the experience as easy as possible** for users⁴ by minimizing the number of systems and interfaces they need to interact with and then **embracing accessibility** as a basic tenet for those environments. These principles are balanced against (among other things) the technical challenges of interoperability between systems and aggregating content from diverse sources.

Another principle is to **provide an abundance of content** to users from many providers, respecting copyright and the bounds of fair use. This is balanced against the potential to overwhelm users, the historical difficulties of implementing effective federated search, the privacy implications for such personalization that may enhance the discovery experience, and the costs of acquiring and stewarding content.

Giving users **a choice of reading platforms** that align with their needs (web, e-reader apps, etc.) is another key principle. The heterogeneity of digital book formats and digital rights management (DRM) across providers makes this challenging, and in our research we found that users often connected specific technologies to their preferred use of content. (For example, users seemed to cite the PDF format because of its consistency and perceived interoperability rather than because it was the best technology.)

³ Our service plan segregates the needs of users who engage with corpora of texts in computational or other non-expressive ways (e.g., digital humanities); these are discussed later, as they have some different user experience needs.

⁴ Project LEND focused on the needs of users in higher education, although it's reasonable to conclude that users more broadly would benefit from an improved user experience.

This service plan also aims to **support commonly needed user functions**, such as user annotation, search within books, and saving portions of digital books, with user-generated metadata persisting beyond any business practices (e.g., a defined loan or usage period). The limited number of options for interoperable annotation tools, potential legal restrictions (copyright and contractual) on saving content, and the diversity of user research workflows complicate this.

Finally, we want to **respect the law and nurture a healthy publishing ecosystem**. Users, copyright owners, and publishers achieve their best respective successes when they succeed together.

With these principles at the foundation, the proposed service plan aims to balance an easy, abundant, choice-filled and lawful experience for users with the technical and policy constraints of working across varied repositories and formats. The architecture described below would allow users to discover and read digital books from many sources through flexible front-end interfaces.

Overall design

Our understanding of a next generation digital book discovery and acquisition environment is founded on our users' expectations of a seamless experience where they can easily browse, search, borrow, read, listen, and manipulate texts with their choice of reading application, with rich accessibility to support the broadest opportunity for research and learning. For this vision to be brought to fruition, libraries should be 1) able to acquire or reference content from their choice of repositories and platforms, 2) provide metadata-driven discovery across all of these sources, and 3) support different reading applications independent of the content sources.

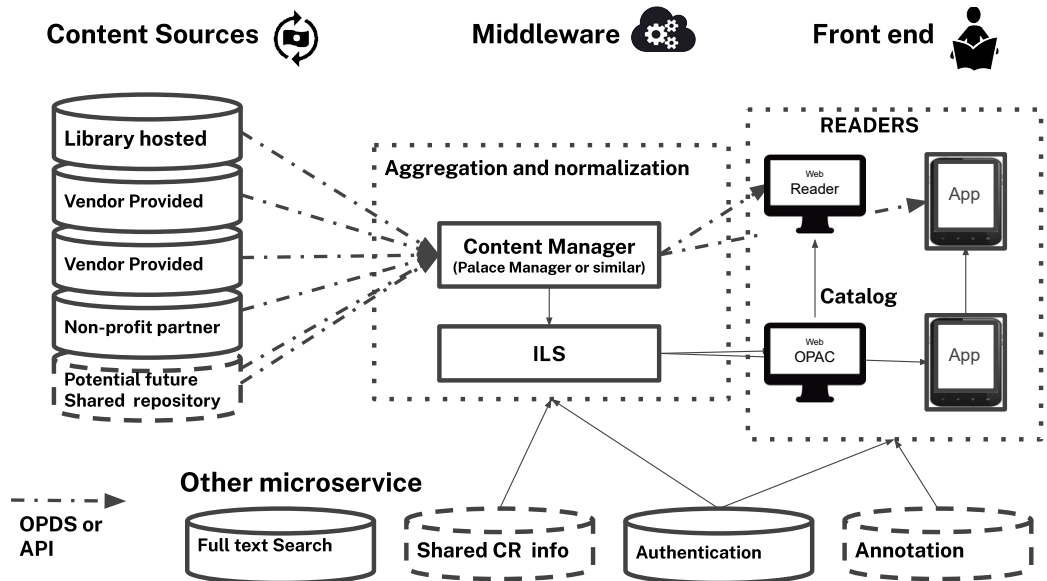
This proposed technology stack presumes multiple content sources⁵ (e.g., HathiTrust, the Internet Archive, plus distributors of licensed titles), a middleware⁶ layer for content aggregation and management (including the application of content protection where appropriate), and support for multiple end-user interfaces for discovery and access. This set of interlocking components can be brought together by utilizing standards such as OPDS for discovery, and EDRLab's Radium Lightweight Content Protection⁷ (LCP), for DRM.

⁵ This proposed architecture would allow content to be added to a service over time as technological, legal, and other obstacles are overcome.

⁶ Middleware is a type of computer software program that provides services to software applications beyond those available from the operating system. In this case, middleware would help disparate content repositories all be able to provide discovery and access to disparate reading platforms and devices.

⁷ Lightweight Content Protection refers to technologies that enforce copyright and/or licensing restrictions without hindering or preventing normal (and lawful) user behavior and without impinging on the privacy of the user.

This architecture can be represented in a diagram:



As represented in the diagram, principle components of the system include:

- Content sources: digital book repositories that host content and associated metadata
- Middleware: software that unites metadata and content from multiple repositories, and manage authentication and content protection, such as DRM
- Front end, comprised of two segments:
 - Systems that support content discovery, such as a shared catalog for discovery, with full-text search capabilities, and support for local curation to provide tailored user experience;
 - Multiple user-facing interfaces, preferably including web-based and mobile applications for discovery, annotation, analysis, citation, and reading.

For example, in the fullness of time, leveraging this architecture we imagine this kind of ease of searching and use: Max (a graduate student) is asked by their thesis advisor to look at a seminal work in their field of study. Max pulls up their favorite app for searching for content and, not remembering the exact title, searches for the concepts they remember from the conversation. The app uses full-text availability and “fuzzy searching” to provide Max with a ranked list of potential options. Max selects the correct item and adds it to their personal workspace online, flagging it for offline reading and annotating availability. On the commute home, Max makes notes alongside the text of the book, which are saved in their personal workspace.

Currently there are no production systems which integrate all of these features, and the work demanded to cobble together different existing applications faces a bramble of frequently un-interoperable programming interfaces. The best approximate market solution at the time of writing is

probably Lyrasis' Palace Project,⁸ which manages both ebook metadata and their discovery by readers; Palace also integrates with Lyrasis' Biblioboard repository for content hosting. That said, Palace stems from work in a public library environment and is at present predominantly a mobile reading system. However, Palace supports OPDS catalogs and LCP for DRM, and both EPUB and PDF format ebooks; additionally a limited web reader is available for Biblioboard hosted content, with enhancements for protected content underway.

Access to digital books

Even a modest interpretation of what users need for their research objectives (i.e., access to “all content available”⁹) requires **aggregation or federated search**. OPDS is currently the best mechanism to gather sufficient descriptive metadata from different sources for discovery and access. An OPDS-based system separates discovery and access layers from content hosts and reading systems, meaning that users can find and use digital books from many sources in one place.

This also means that content should be made available in **interoperable formats**. EPUB seems to hold the most promise to meet the articulated desires of users, but either EPUB or PDF with a text layer are better than proprietary formats or digitized texts that lack an accurate OCR or text layer.

Project LEND recommends that library systems and content platforms aspire to meet the [Nielsen Norman usability heuristics for user interface design](#) and the W3C's [Web Content Accessibility Guidelines](#) as much as possible.

Search experience

Project LEND envisions **streamlining of discovery** — to be able to search in one place and find all content available to the user, even if that content is itself decentralized. The architecture described above would allow a system to access various content repositories via interoperable APIs or, better, OPDS. Additional content repositories could be added over time, increasing the value of the service to users and making the content increasingly discoverable.

The ability to evaluate search results can be vastly improved by **enriching metadata**:

- When abstracts, tables of contents, and other commonly-desired metadata are not already available in provided metadata for evaluation of search results, the metadata should be enhanced by leveraging the full-text of the content to create or extract such metadata. For digitized content, “full-text” includes having a high-quality OCR available.
- All policy-based metadata (e.g., terms of access) should be provided with content, so that it can be conveyed to users at the point of access.¹⁰

Improving the ability to **search for content in multiple languages** has many potential components to it. Building a centralized search engine that handles indexing and relevance ranking of the hundreds of languages that may be included in an academic database, as well as content from many different sources, would benefit from access to the full-text. Displaying metadata in the user's native language might be possible, including the use of modern

⁸ The Palace Project seeks to create an easy-to-use system for digital book access for public libraries. Some academic libraries are currently investigating Palace's utility to play a similar role in higher education. <https://thepalaceproject.org>

⁹ As noted by users themselves, there are a vast number of books not yet available digitally, many of which retain actual or potential research value; opportunities for libraries to address this are described below. Ideally, any service would also alert users of other options to obtain content that isn't yet directly available, such as interlibrary loan.

¹⁰ The statements and icons designed by RightsStatements.org offer some options, in addition to the Creative Commons licenses that cover open access material.

technologies (like artificial intelligence) to generate key words and abstracts. If this metadata generation were done prior to indexing, it could also help to power search in multiple languages.

A way to **re-discover something seen earlier** in the research process should also be included; examples of “items looked at earlier in this search session” are abundant in other web interfaces.

Some **normal usability functions** are absent or inconsistently present in book delivery interfaces today, including things as basic as intuitive search functions. Again, the discovery process would need access to full-text of content to enable some search functions desired by users. Other functionality asked for by users are similarly very standard for modern search environments (e.g., the ability to refine search results; ranking results by relevance).

Use of digital books

A reading (and usage more generally) environment should offer a basic suite of features that address near-universal desires of users:

- Flexibility in manipulating content, including reading settings, adjustable text sizes and backgrounds, and adjustable parameters like font size and line spacing – these would be standard system requirements, as well as being important for accessibility
- Searching inside digital books, including doing so without opening up the document; again this requires access to the full-text of the content
- Options for generating consistent and accurate citations for content, including generating such metadata based on factual information about the content as needed

It should also address issues related to **durability of access**, supporting:

- A need to continue using a digital book (overriding or extending any imposed time limits via a seamless “renewal”)
- A need to resume using a digital book (renewing any “expiration” of access)

Some commercial ebook providers are already interrogating how the loaning (and “forced sharing,” e.g., recalls) of books have been mapped onto the ebook market — for example, ProQuest’s concept of Non-Linear Loans.¹¹ Digital books offer entirely new opportunities to rethink how we provide content to users, and an adherence to the business logic of trafficking physical books may be counterproductive.

A reading environment should also include features that let users **manage their access to content**. Users talked about wanting to be able to download content — and especially to download content in such a way that it doesn’t later “expire” before they are done using it. That said, it’s not clear that users strictly need to download the content to achieve their stated goals.

- One goal that seemed important to users was that the content was saved, secure, and durably available. These particular goals might be achievable with some kind of online “bookshelf” managed by the user, for example.
- Another goal that users associated with downloading was to be able to access content during periods of time when the user didn’t have a stable internet connection. Again, if that online “bookshelf” had a “use when offline” option, as modern utilities such as Netflix and Google Drive do, this might not require the kind of downloading that the users were envisioning.
- Finally, downloading was mentioned in the context of wanting to share materials between

¹¹ ProQuest’s Non-Linear Lending model allows simultaneous usage by a fungible number of users, totalling up days-of-use until a cap is reached. (For example, a single person using a book for three days counts the same as three separate people each using the book for one day, even on the same day.)

instructors and students or between colleagues (including colleagues at other institutions). Again, there may be ways to enable these kinds of collaborations without requiring the kind of downloading that the users were envisioning.

Finally, a reading environment should let users **organize the content** they are using — and offer suggestions for organization. Users should be able to add (upload or link) other kinds of content (notes, citations, lectures, slides, readings). The system should offer to rename files (i.e., to match the title of the article/book) as well as deduplicate, categorize, and organize them by concept.

Opportunities for stakeholders (libraries, publishers, distributors)

Users readily confirmed that they are not interested in how libraries (and others) distinguish between commercially available ebooks and the digital books that are created by digitizing physical books. This distinction is often entirely lost on them, and users are confused about why content is kept in so many different places and in so many different formats. While Project LEND began from a place of wanting to understand how libraries might more effectively use books they have digitized, we were agnostic in asking users how they might best advance their research and educational agendas, instead allowing users to speak freely about both commercially available ebooks and digitized content (e.g., from print formats) interchangeably; this blurring best reflects their actual behaviors.

As an example of this, we found that while libraries and publishers have a deep understanding of the different legal and technical contexts surrounding digitized books and ebooks, users not only did not understand this distinction, they did not understand how the technical and legal contexts led to such considerable differences in terms of their experience.

Users clearly would prefer to have libraries and publishers integrate their efforts much more deeply, and expressed this preference at every stage of the research process (discovery, access, usage).

While implementing the technical architecture of the service plan is currently feasible, and incorporation of some content streams is already feasible, it is clear that more research, technical work, policy and legal analysis and collaboration across content producers, libraries and users is needed. This section outlines a number of potential areas of work and collaboration that are needed alongside the implementation of the technical platform to realize the use cases in this document.

Libraries: Expanding digitized content with intentionality

One domain for library action is to look holistically at digitization efforts — mass digitization efforts as well as digitization (largely or entirely uncoordinated) by various libraries acting on their own — with an eye toward building an understanding of how to prioritize such efforts, share the benefits of our investments, and avoid duplication of effort where possible.

This invites libraries (especially those engaged in digitization at scale) to consider how best to manage the prioritization of digitization efforts (e.g., to focus on books not already available as ebooks) as well as the potential for digitization-on-demand in ways that respect copyright and the publishing ecosystem. Even this does not address content that is held by very few libraries in the first place or by zero libraries. By working together, libraries could more systematically build collections that are accessible to users within the boundaries of copyright.

Libraries and legal scholars should also continue to work toward a shared understanding of what is permissible for libraries to do in terms of

digitization and the usage of digital books. One common set of circumstances is that of a book where the publisher cannot discern a return-on-investment of re-publishing it as a commercial ebook, but the print book remains clearly held in copyright, leading to reluctance to digitize it by a library or repository that is uncertain of their right to do so and/or their rights to use a digital copy for any particular purpose. Even for books that fall outside of these particular circumstances, the appetite and capacity of libraries (and other entities) to continue mass digitization efforts and/or support on-demand digitization is not well known beyond inferring from current activities.

It also seems worth exploring how the corpora that have been generated from mass digitization efforts might be leveraged to satisfy document delivery and interlibrary loan fulfillment. Even a modest effort to satisfy requests for content that is clearly in the public domain would be an improvement.

Libraries / publishers: Metadata enrichment

There are metadata implications, as well, for libraries that are stewarding digitized content. Given that factual information about the content is generally lawful to create and use, the metadata for digitized content should be dramatically enriched by leveraging emerging technologies (such as artificial intelligence). Much of the content that has been digitized through mass digitization efforts was acquired by libraries under earlier assumptions about how to optimize metadata,¹² and today's users are frustrated and baffled by how under-described these things are.

Related to this, the OCR for digitized content needs to be of consistently high quality to enable a variety of discovery and use cases.

Publishers / distributors: Enabling usage

Publishers and ebook distributors have made some system and policy design choices that prevent users from engaging with content in the ways that they want, suggesting that a much deeper partnership between publishers and libraries is needed.

Some opportunities for optimizing the user experience include:

- In some cases, publishers or copyright owners allowing or enabling content to participate in an open digital book environment in the first place (e.g., making metadata available in OPDS)
- Providing consistent access to full-text to enable discovery and usage (and to as much text as legally permissible when full-text is disallowed)
- Transitioning from proprietary formats to open and accessible format standards (e.g., EPUB)
- Rethinking when protective measures (such as DRM) are needed to allow some kinds of actions that are fair use (such as the user being able to leverage their preferred translation utility on a section of text)
- Exploring ways to increase the durability of digital access so that content can be used again at a later point in time, including durability for citation and use by peer reviewers, etc., as well as enabling use when internet access is unavailable
- Looking creatively at options for use and "loan periods" that go beyond the recapitulation of how physical items are trafficked by libraries
- Providing metadata about access limitations so it can be displayed and/or conveyed in some way to users

¹² Conciseness was strongly preferred both in the era when the metadata had to fit onto a catalog card and in the subsequent era where digital storage was incredibly expensive. The metadata for material processed during these eras remains largely unchanged, making it hard to assess content during the discovery process.

Libraries: Educating users

There is value in considering anew how best to help users understand their own rights. It was consistent across focus groups that users misunderstood copyright law, which is understandable in light of the law's intricacies and dynamism. For example, several users engaged in text and data mining seemed unaware that there is an exemption from otherwise applicable anti-circumvention law that permits the circumvention of technological protection measures "to deploy text and data mining techniques on a corpus of literary works for the purpose of scholarly research and teaching."¹³ These users are surely also unaware that the Librarian of Congress has recently enlarged that exemption to permit the sharing of access to a corpus compiled for text and data mining purposes with "researchers affiliated with other nonprofit institutions of higher education."¹⁴

Education might also help users better understand the relevant technology and its capabilities. For example, users seem to focus on PDF as a format even while citing affordances they need that PDFs do not provide. These users could benefit from learning about the EPUB format and its benefits.

Finally, there are ongoing opportunities (well known to libraries already) to improve the search literacy of users, especially students who are still maturing their understanding of the research process.

Libraries: Building coalitions

Increasingly, academic libraries are banding together in coalitions to foster the development of increased interoperability in reading systems; improve repository integration; and help to drive widely used standards such as OPDS through formal standards adoption processes in concert with NISO and other international organizations. These (admittedly nascent) activities increase the options available to libraries and readers; provide greater exposure and access to both commercial and not-for-profit content platforms; and ultimately enable a springboard for the development of new functionality in reading and analytical systems. Increasing the number and heterogeneity of academic libraries participating in such efforts over time would help these efforts.

For Future Investigation: Use of Corpora of Digital Books

Project LEND convened dedicated research sessions with users who engage with corpora of texts, using data science and digital humanist research methods. That said, these sessions didn't yield a clear vision of the user types.

These users have needs that go beyond those listed above, and their needs are not always as easily brought together in the above themes, given the sometimes idiosyncratic nature of their work and that many of these users are self-empowering, often building or adapting utilities to carry out their work. Indeed, in many cases the processes they engage in are bespoke.

¹³ 37 CFR part 201.40(b)(5).

¹⁴ Id. As the Register of Copyrights explained in her recommendation, the revised exemption is intended to allow access to the corpus hosted by the originating institution, not to allow downloading or other distribution of copies. See [Section 1201 Rulemaking: Ninth Triennial Proceeding Recommendation of the Register of Copyrights](#) 85-86, 89, 97-98; see also 37 CFR part 201.40(b)(6) (permitting nonprofit institution of higher education to provide access to "researchers affiliated with other nonprofit institutions of higher education, with all access provided only through secure connections and on the condition of authenticated credentials, solely for purposes of text and data mining research or teaching").

While we can pull out certain themes, some other comments were made by only one or two people and it isn't possible to extrapolate to a broadly-shared need without input from more of these researchers. Additionally, the number of users engaging in these research modalities right now is a small percentage of the total users; we would want to verify the actual and potential¹⁵ appetite for improvements before deciding how best to invest resources.

Some themes duplicate needs described above — e.g., streamlined access to as much content as is lawfully available to the user; full-text search capabilities.

Users want both an environment that lets them use a corpus (or a subset thereof) as datasets for computational use, and then to be able to cite what they have built or worked with in non-expressive contexts. Project LEND believes there is value in enabling dataset citation, especially given academia's present reproducibility problem, so that published works of scholarship can be peer reviewed¹⁶ appropriately and all upstream work of a research paper (or other research output) can persist in a way that lets scholarly communication continue to interact with both the findings and the underlying data and methodologies.¹⁷ Further conversations with experts working in this area would be needed to determine how best to meet their various needs for citation and durability.

Users mentioned several other idealized workflows that would implicate the bulk downloading of texts regardless of copyright status — as described by users, this may not be fair use and/or would likely violate contractual terms for commercially available ebooks.¹⁸ This may prove to be an area where users are only able to conceptualize based on what has worked well for them to date. The approach of the HathiTrust Research Center's service design, which limited direct access to in-copyright content, was frustrating to many of these users, leading them quite naturally to ask for environments that were completely free of barriers. Further collaboration with these users to tease out what is needed would be a next step toward a system design that honored copyright while enabling research.

We also note that there are many fewer of these kinds of users, and it is hard to discern whether that is because of barriers to participation (technological barriers, uncertainty about copyright law), because this is still an emerging modality of research and education, or if the modality may remain a small percentage of how research is conducted. This itself would be an important question to take up in considering how much effort to invest in optimizing support for this research.

¹⁵ We acknowledge that we don't know what the appetite could be to work in these modalities if they required less technical facility to engage in and/or seemed to carry less legal risk for the researcher.

¹⁶ That is, available to peer reviewers as well as other parties in the editorial process.

¹⁷ Put another way, it's going to be difficult or even impossible to prove or disprove a paper's findings at a later point in time if the dataset on which the paper is based cannot be saved in some way for later examination.

¹⁸ Similarly, idealized workflows that enable the stripping of DRM/TPM (which only recently became allowable under a 1201 exemption) may be constrained or eliminated in future rulemaking.

Concluding Thoughts

The period of time when this research took place saw tremendous changes. *Hachette v. Internet Archive* was an active case moving through the courts, with questions about the legality of certain digital sharing practices at its heart. The Second Circuit's ruling and its applicability (or not) to higher educational contexts added complexity and uncertainty to our legal analysis. The U.S. Federal Government has also introduced considerable uncertainty in the concluding months of this project — including its support of higher education, the potential to require compliance with federal guidelines to receive support, and so forth. Our project team has not had the capacity to consider all possible future scenarios.

It's critical to focus on who has the most to lose if we take no action: people at less well-funded institutions, students who are economically disadvantaged and do not have the resources to purchase (or rent) all the content they would need to earn a degree, and of course those with reading disabilities. Our focus group participants came from a wide range of institutions yet agreed on a number of key concepts around access and accessibility.

With all of that in mind, we propose three future areas of activity to advance these conversations and continue advocacy:

1. A renewed focus on accessibility
2. A focus on community building and collaborative action
3. Applied legal analysis (taking a discrete set of works and apply legal analysis to determine if those materials could be made available for some kinds of use)

Acknowledgments

Project LEND would not have been possible without the generous support of the Mellon Foundation through a two-year grant (2023-2024), and we thank them for their partnership on this project.

Project LEND drew on the expertise of over 30 librarians, faculty, and other experts working across four research domains over three years of planning and research. We'd like to thank every single one of them for their insights and their contributions. Our colleagues across the UC Libraries and especially at the UC Davis Library also agreed to participate in an early testing stage of the Palace Project in part to contribute insights to Project LEND, and we thank them as well.

Our user research benefited from initial interviews with a number of digital platform managers and designers who generously contributed what they knew about the current state of the user experience to help us build out an environmental scan and foundational use cases. The American Foundation for the Blind was extremely helpful to us in thinking through how best to engage with blind and low vision users.

Finally, we'd like to thank the users who participated in focus groups, interviews, and surveys to help us understand their challenges and frustrations.