Media + History + Digital + Library: An Experiment in Synthesis

Eric Hoyt, Wendy Hagenmaier, and Carl Hagenmaier

It's a big exercise in the synthesis of things. We kept coming back to this statement as we considered how to describe our process in developing the Media History Digital Library's (MHDL) website and Lantern, the project's new search engine. No, it's not the most eloquent expression (though it's a lot nicer than some of the phrases we've shouted at our laptops!). The statement succeeds, however, in capturing the project's blend of different communities, collections, skill sets, and open source software. Even reflecting on the statement in the light of day, away from our error-laden terminal screens, we still think of our work as "a big exercise in the synthesis of things."

In this essay, we consider some of the particulars of this synthesis. We describe the background of the Media History Digital Library, a nonprofit initiative to digitize public domain media periodicals for broad access. However, we focus primarily on our roles as designers, developers, and digital humanists (a synthesis that can both exhilarate and, at times, leave you feeling not particularly scholarly, computer savvy, or, for that matter, human). We analyze the three goals that drive our design work: access, usability, and impact. And we discuss the ongoing challenges to achieving them.

The Project

For decades, scholars have relied on old trade papers and fan magazines in writing the histories of film, media, and broadcasting. Industry trade papers, such as *Film Daily*, offer extensive documentation about the development of the media industries—chronicling the paths taken, the promises never fulfilled. For historians of silent cinema, reading through several years' worth of *Moving Picture World* on microfilm became a rite of passage. You put in an interlibrary loan request for several microfilm reels. Once they arrive, you reserve a machine, carefully load a reel, and slowly crank, looking page by page through this important industry publication in search of articles and advertisements pertinent to your research topic.

Capital Kalem Feature

"Under a Flag of Truce," One of the Best of that Company's War Pictures.

WHEN the Kalem company of players, under the direction of Kenean Buell, established its headquarters at Jacksonville, Florida, it was with the avowed purpose of producing a series of war pictures, to be based for the most part upon historic incidents of the late War of the Rebellion. To further strengthen the work of its players, the Kalem Company contracted for the exclusive services of Mr. Emmett Campbell Hall, a scenario writer who had become unusually proficient in the art of composing picture plays and who is not unknown to the readers of the Moving Picture World. Realizing that a knowledge of the players would enable him to prepare his plays to better advantage, Mr. Hall spent several weeks with the Florida company. The result of this combination has been a number of exceptionally good pictured war stories, notable among which are "A Spartan Mother," "War's Havoc" and "Fighting Dan McCool," each of which are remarkable in point of dramatic action and effects.

While all these plays deserve to be classed as feature subjects, a later addition to the list entitled, "Under a Flag of Truce," shows, if anything, a marked improvement on

Scene from "Under a Flag of Truce" (Kalem).

previous productions. This subject is dated for release on Friday, May 24th. The points of excellence are found in the better grouping of the players and the more spirited action in the ensemble scenes.

There is a slight love story running through the picture which gives a touch of heart interest and a motif for the players. Tom Dean is the hero, who is in love with Rose Neville, daughter of a Southern planter. Tom is a Union sympathizer and is not in favor with the old planter. News of the fall of Sumter brings matters to a crisis. The old planter raises a company and goes to the front. Tom refuses to join and is cast off by his sweetheart. He goes North and enlists in the cause of the Union and some time later returns at the head of a troop, ordered to destroy the old plantation.

The old planter has been wounded in a battle and has found refuge at his home. The approach of the Union forces compels him to hide in one of the negro cabins. During the scrimmage between the opposing forces on the plantation, a worthless character named Bill Wheeler attempts to take liberties with Rose, but is caught by the soldiers and severely flogged. Smarting from his punishment he swears vengeance and organizes a party of bushwhackers with the intention of looting the plantation. Tom Dean, now a captain, overhears the plot of the bushwhackers and goes to the rescue, joining the old planter and his family in the negro cabin in their defence against the marauders.

When the fight is at its height, Rose's sister, Nan, manages to escape from the cabin and reaches the Confederate and Union forces and advises them of the danger that threatens the old Colonel and Captain Dean. Uniting under a flag of truce, the Confederates and Unionists go to the rescue and drive off the bushwhackers, relieving the defenders of the cabin just as they have abandoned hope.

The final scene of this picture shows another truce between the Confederate Colonel and the Union Captain who clasp hands in the light of the burning cabin. This final scene is beautifully toned and constitutes one of the prettiest effects in the picture. Especial music has been written for this picture by W. C. Simon, and a large variety of posters and advertising matter has been prepared. It is a worthy subject and one which the exhibitor may feature to the limit.

The cast of the principal characters is as follows: Tom Dean, Guy Coombs; Rose Neville, Anna Q. Nilsson; Col. Neville, Henry Hallan; Mrs. Neville, Helen Lindroth; Bill Wheeler, Hal Clements.

Mayor Gaynor Wants Action

His Honor Asks New York Aldermen to Get Busy on Picture Theater Ordinance.

Mayor W. J. Gaynor, of New York, went on record last week as a staunch advocate for moving pictures. It has been known for some time that he considers the moving pictures a source of great educational value and in the past has not hesitated to give expression to these views. In no public utterance, however, has the Mayor of New York expressed himself so favorably as he did in a letter to the New York Board of Aldermen when he wrote a letter to that body telling them in effect to "get a move on" and do something with the ordinance which has been buried in committee for nearly six months with no report from the committee as to what they intend to do about it.

The ordinance in question, known as the Folk ordinance, has been printed in full in the Moving Picture World and from all indications the trade is satisfied with it as applied to moving picture shows in the five boroughs of Greater New York.

Apparently the Mayor has no sympathy with the so-called censorship ordinance recently introduced into the Aldermanic Chamber by Mr. Nicoll, for he in no way refers to it in his present letter to the Aldermen.

Mayor Gaynor's letter in full to the Board of Aldermen is as follows:

City of New York, Office of the Mayor, April 29, 1912.
To the Honorable the Board of Aldermen:

Sirs:—At my suggestion the Commissioner of Accounts made a careful study of the conditions of the moving picture shows of the City of New York. This was brought about by many complaints which I had received. He submitted his report on March 22, 1911. I appointed a committee of citizens to examine carefully into the whole subject and to report a proposed ordinance for the government of these shows. They completed their work and reported the ordinance, which was duly submitted to your honorable body. The matter is one of such importance that I have thought it proper to address to you this special communication thereon. It seems to me that this ordinance is well thought out and should receive favorable consideration and be passed by you. I do not undertake to say that it is perfect. I dare say that you may find that it needs certain changes or amendments. All that I urge is that the ordinance be considered and passed as it is, or with such wise amendments as you think proper. These shows are a great solace and the source of much entertainment and education to the whole community. They are attended by all kinds of people, but especially by those who cannot afford to attend more expensive places of amusement. The pictures shown are moral and instructive. The great outcry of certain uninformed persons aginst them which existed not long ago has subsided. Our part is to safeguard these shows in all respects, physical, mental and moral, by a good ordinance.

Respectfully,
W. J. GAYNOR, Mayor.

Not long ago an indignant Brooklynite wanted the Mayor to refuse to sanction the licensing of a place for moving pictures in that city, but the executive promptly dismissed the property owner with the remark that moving pictures were a good thing and that there was no harm in having such shows near any person's residence.

CHARLES URBAN ILL.

Mr. Charles Urban, widely known as the pioneer and perfector of the now famous Kinemacolor motion picture, is reported to be seriously ill at his home in Ashley Gardens, London. Mr. Urban was suddenly ill while at his studios in Wardour Street, London. Physicians were summoned and a diagnosis proved that he was suffering from gastric ulcer. An operation was performed immediately and, while the patient is not out of danger, it is believed that he will make a speedy recovery.

Figures 1 and 2. Scans of *Moving Picture World* from the print original are compared to microfilm printout.

The microfilm process has its benefits. It allows the discovery of fascinating stories that otherwise you would never have known existed. It also gives the scholar a rich sense of context—something far deeper than a series of keyword searches can produce. However, the process suffers from inefficiencies and lost opportunities. It is agonizingly slow work, and even the most systematic researcher misses important articles toward the end of an

eight-hour shift at the microfilm machine. The process promotes research questions

focusing on a tight time frame—for example, the nickelodeon boom, cinema during

World War I, or the transition to sound. And because scholars gravitate toward the

publications that have been transferred to microfilm, countless other histories that could

arise from rarer publications are left unwritten.

We all knew there were ways of using digital technology to improve this research

process, but film historian David Pierce was one of the first to act on this knowledge. He

knew the breadth of publications that existed, and he knew the institutions and private

collectors who possessed original copies (not simply the microfilm). He synthesized his

historical knowledge and personal relationships with his twenty-five years of experience

investigating the copyright status of books and films. Whereas most Hollywood feature

films from the 1920s through 1950s are still under copyright protection, most of the trade

papers and fan magazines of this period are in the public domain. As the adage goes,

"Yesterday's news wraps today's fish." The Hollywood studios applied for renewals for

their copyrighted content, but the news publishers who covered the studios did not.

Unlike digital projects organized around a particular institution's collection, the Media

History Digital Library was conceived from the beginning as a network model. MHDL

knits together the collections of many different individuals and institutions. The project is

supported by owners who lend material and donors who fund the scanning, which is

carried out by the Internet Archive (IA). Our goal is to use the affordances of digital

technology to build a library of media history publications more comprehensive than

what any single institution holds. No single collector or institution, for instance,

possesses a complete original print run of *Moving Picture World*. By scanning volumes

from multiple contributors, however, we are working toward digitally constructing that

complete set.

Figure 3. This January 1927 issue of *Photoplay* was one of the first magazines digitized by the Media History Digital Library.

In 2009, the MHDL coordinated the digitization of ten volumes of the fan magazine *Photoplay* from the collection of the Pacific Film Archive. In 2011, the project dramatically scaled up its scanning activities, digitizing over 200,000 pages of text from several private collections. The collectors, who generously loaned materials to the project, expressed excitement about obtaining digital copies of their bound volumes that were searchable, shareable, and portable. David Pierce and Eric Hoyt, who were ideally situated on the East and West Coasts, respectively, were able to carry out their facilitating roles in person—packing up volumes from collectors' homes, delivering them to the nearest IA scanning center, and ultimately returning the bound volumes to the spaces waiting for them on their owners' shelves.

The Design Challenge

The project's boom in digitized pages opened up a new design challenge—how to present the materials? The volumes are hosted on the [Internet Archive website](), but IA's interface is oriented toward books more than periodicals. Additionally, the massive scale of IA's collection—now exceeding 3 million digitized books!—dwarfed our own modest subcollection. We heard from users who felt lost in the stacks, and not in the good way.

Essentially, we had a digital equivalent of microfilm; yes, from original copies with great quality, and somewhat searchable by volume, but the primary advantage was that the material could be read from a laptop. We did not have a way for users to easily find

information within the collection. Legacy approaches to finding materials—Library of Congress catalog numbers, subject indexing—were insufficient.

We needed to create a new website. A centerpiece of the site had to be a search engine that could search across the full text of the entire collection (a project that is nearing completion as we write this). But for an initial site, we could allow the features to be more limited. We needed a website that represented MHDL's identity, allowed for easy browsability of the collections, and enabled users to quickly find what they were looking for.

We greatly benefited from the framework the Internet Archive had already put into place. Thanks to IA's elegant [BookReader](#), users could read magazines online and search within individual volumes; link to individual page images using hard URLs; and also embed the BookReader in their own webpages. And because IA hosts the files and metadata (and takes great steps to preserve them), we had the opportunity to focus our efforts entirely on the user experience.

Access

In our design of both the initial MHDL website and the new search engine, Lantern, we strive for three goals: access, usability, and impact. One of the most significant powers of the digital sphere is the opportunity it offers to create virtual landscapes—to construct aggregations and juxtapositions of materials and ideas that will never exist together in the real world; places that are open to everyone with an internet connection; spaces in which

the research process accelerates, creating room for deeper discoveries. Digitization does not automatically equal accessibility or the expansion of new avenues for research. Digital library projects prove most successful when they knit together virtual collections that broaden the territory in which users can search and question. With a user-friendly online home and a powerful search portal, MHDL attempts to break new ground in the domains of access and audience. We hope it serves as a tool just as suited to dissertation research on the rise of color film technology as to quick questions about the history of a neighborhood movie theater in Kansas or the scheduled television programming in Dallas the hour JFK was assassinated.

As researchers know all too well, many materials that exist in the public domain and are theoretically free to view may, in reality, be unavailable or unaffordable. If you have to finance a trip to archives across the country to piece together access to public domain materials, or if you are unaware whether certain public domain materials still exist, then the public domain withers both as a concept and a practical resource. Not only can MHDL reinvigorate engagement with public domain materials; we also hope that by offering a variety of formats for download and online viewing, it will enable users to enter and explore the collection from multiple angles. In the future, we hope to give users the ability to toggle between viewing entire bound volumes and examining page- or even article-level segments. Some users should be able to see the collection as tens of thousands of advertisements, while others as thousands of reviews. Users will be able to isolate the particular while maintaining access to the provenance and the broader context in which the fragment exists.

The hosting of the collections within the Internet Archive's robust preservation infrastructure will ensure that this full access to MHDL materials will persist and remain consistent into the future. In our work, we have prioritized keeping the workflow of the site sustainable and its interface extensible as the collections grow, in order to ensure continuity and expansion of access. Maintaining this sustainability and extensibility will continue to be one of our most important challenges.

Usability

We believe that any effective interface for exploration must be fundamentally intuitive and user-friendly. The first step in achieving these goals, however, is to step back and consider exactly who are the users of the collections. Our primary user base consists of scholars, fans, and students. The initial MHDL website, built in WordPress, is our portal to the collection. It is most intuitive for scholars and fans who are already familiar with the publications featured on the site—the same folks, in other words, who just last year were looking at *Moving Picture World* on microfilm. These knowledgeable historians and fans already know the difference between *Film Daily* (a trade paper) and *Photoplay* (a fan magazine). They also have the years of film releases burned into their brains. Here's a quick test: when you type "Mr. Smith Goes to Washington," do you think the title looks incomplete without "(1939)" printed next to it? If so, we're talking about you. Expert film scholars and fans can swiftly navigate their way through the website's collections pages and expandable lists of periodicals to find what they're looking for (see Figure 4).
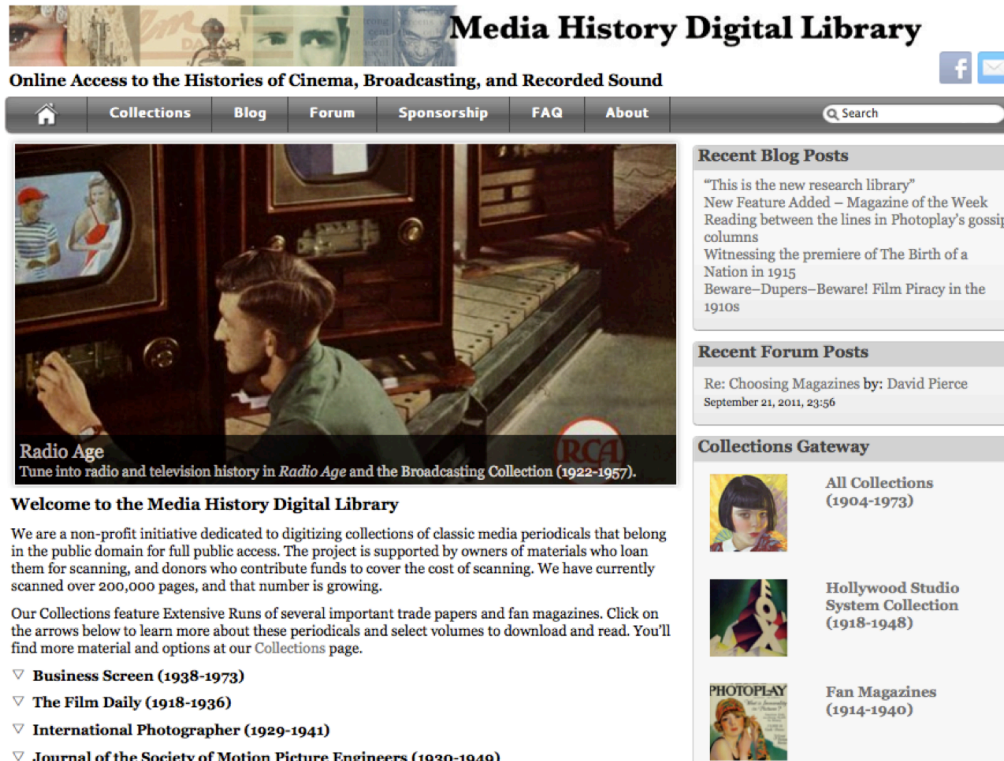
Figure 4. Screenshot shows the initial Media History Digital Library website, launched in September 2011 and built in WordPress.

But there is a larger group of users who are passionate about film yet lack experience researching historic industry periodicals. For this group of users—which includes most undergraduate film students and the fan base of Turner Classic Movies—the initial MHDL site was challenging to navigate. You knew you wanted to see historic news coverage about Cary Grant, but you didn't know where to start looking. Should you start browsing in the Hollywood Studio System Collection or the Fan Magazine Collection? And what years would be the best to look at? Why could you perform keyword searches within an individual volume, but not across multiple volumes at once?

To better serve expert and nonexpert users alike, we are developing Lantern, a customized search tool for media history research. Lantern is powered by the open source

Solr search engine and builds on earlier innovations by IA's outstanding programmers—especially the work of Mike Ang and Raj Kumar on the BookReader and Edward Betts and George Oates on Open Library.

How much about how technology works do you want to know? In the spirit of providing control to our users, we offer you the choice of whether to open the following subsection that gets into the technical details of Lantern's development and our programming challenges. To read the technical section, click here and the section will become visible. Otherwise, the subsection remains hidden and you can simply continue reading down the page.

We have spent months working on Lantern's index and interface. Creating the index has been time-consuming, though more straightforward than developing the interface. We began the indexing process by writing a schema that defines the metadata fields and their data types. To fully harness the power of Solr, we had to create certain fields that look similar but that the computer interprets differently. For instance, "date" is a string, "year" is an integer, and "date-start" and "date-end" are Solr date fields (stored in the YYYY-MM-DDT23:23:59Z format). Each data type enables a different user function—"year" can be faceted, "date-start" and "date-end" are sortable, and the string "date" has the flexibility to display whatever notation best suits the particular work (such as "Jan-Jun 1940" for a six-month volume of a journal or "March 7, 1923" for a single issue).

Once we had the schema in place, we began running the Internet Archive's XML metadata fields through our own XSLT file to create new XML documents that Solr could properly interpret. We also added new metadata fields and imported each item's OCR full text into a new field called <body>. Much to our chagrin, we could not fully automate the process of creating XML documents for each work in the index. The standard tactic of libraries, importing MARC documents through SolrMARC, simply did not meet our needs. MARC records are useful for books but unhelpful for periodicals; a single MARC record describes a periodical's entire run, which might go up to eighty years. We had to roll up our sleeves and do some of the indexing and metadata entries ourselves, and we are grateful for the hard work of Andy Myers, Joseph Pomp, and Derek Long in helping us build up our index.

As laborious as the indexing process has been, the interface has proven the most difficult to develop up to our standards. In November 2011, we built an initial interface using a combination of XSLT, PHP, JavaScript, and CSS (see Figure 5). We had complete control over the interface's appearance and operability. When a user ran a query, we displayed the core metadata fields and up to three full-text snippets (see Figure 6). We provided a link for users to click through to the BookReader, which passed the same query to the BookReader and generated bookmarks pointing to the matching pages).
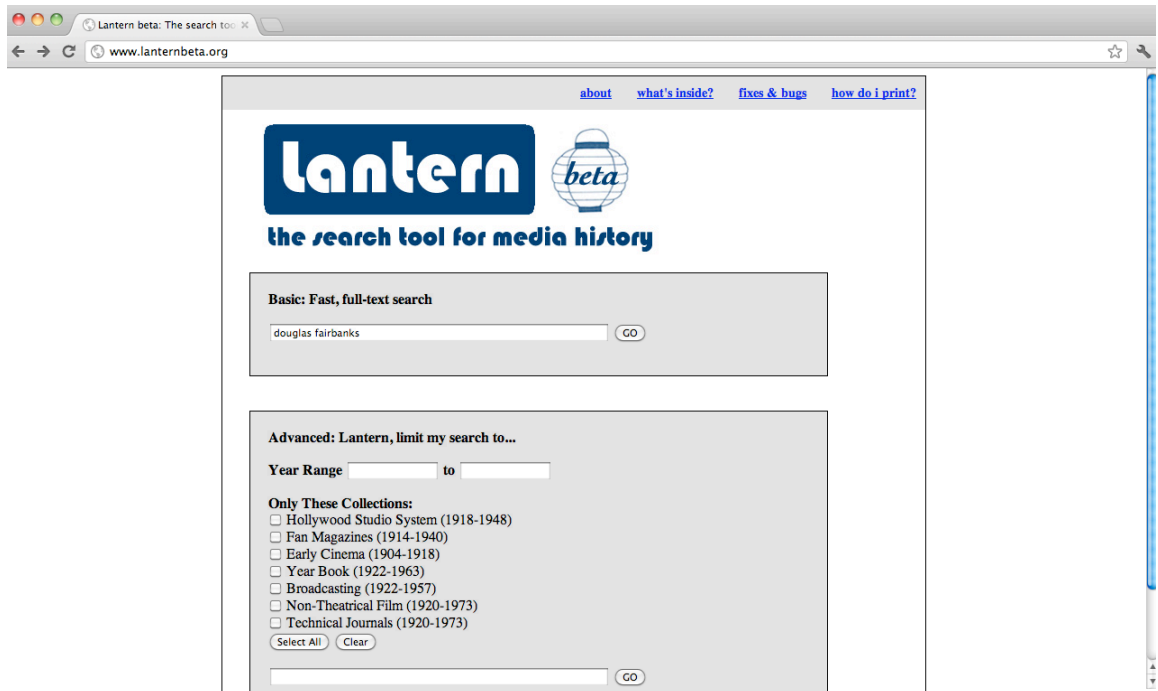
Figure 5. The first Lantern interface was created in November 2011 using a combination of XSLT, PHP, JavaScript, and CSS.
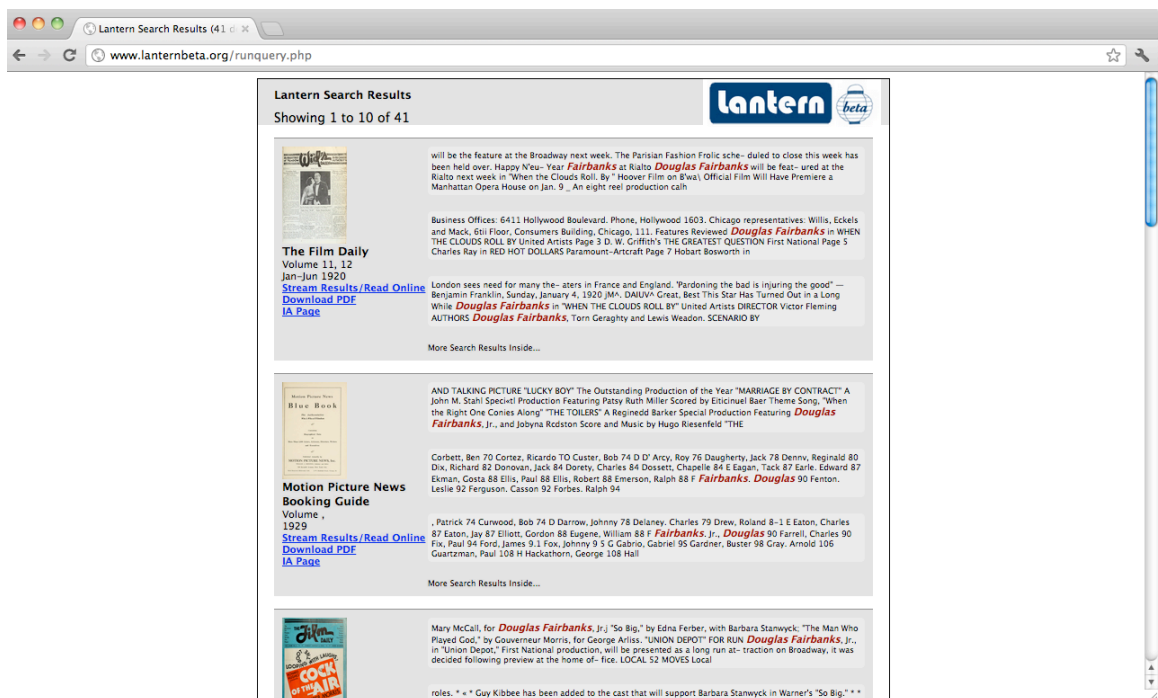
Figure 6. The first Lantern interface yielded snippets of text and slow search speeds.

Ultimately, though, we realized our best option was to abandon this interface. We grew weary of the programming burden of hand-coding every single feature that users have come to take for granted (such as the "Next 10" and "Previous 10" options at the bottom of the results page). We also grew frustrated by the relatively slow search speed of our searches—particularly when they were passed on to the BookReader, which took up to a minute to generate hits for a 1,000-page bound volume of magazines and frequently produced results different from the snippets (the default Boolean operator in the BookReader is "OR" whereas our default operator is "AND").

In April 2012, we began developing a new interface (see Figure 7). We decided to move forward by adapting Blacklight, a remarkable open source application created by the University of Virginia Library. Blacklight is specifically designed for libraries using Solr indexes. The decision to use Blacklight forced Carl Hagenmaier and Eric Hoyt to learn the basics of Ruby on Rails (a language new to them both), but the tremendous out-of-the-box functionality of Blacklight made the decision worthwhile.

Our main challenge, as we write this, is in customizing Blacklight to add the enhanced functions that our users want—particularly, the ability to run a search and move from a snippet in the results page to its corresponding page in a particular work. We already can deliver all of these features, but the process runs too slowly. Lantern is competing for users' attention in a world where any search

that takes more than three seconds is perceived as slow. A search that takes forty-five seconds is agonizingly slow. Fortunately, we have found collaborators at Eric's new institution, the University of Wisconsin–Madison, who are helping us optimize the search speed.

Like Fred and Ginger dancing, Lantern's performance should appear effortless to the average user. The entry point for search consists of a single query box, following the Google model, which searches full text and a few core metadata fields. The advanced search form offers the same faceted search options users have come to expect when running queries in academic search portals or narrowing their Google searches according to category or time line. By relying on familiarity as our guiding principle, we hope to create a website that first-time users can navigate intuitively and comprehend immediately. If the navigation of the site itself remains simple and expected (thus becoming almost unnoticeable), users can focus on the content and on accomplishing the objectives that drew them to the site in the first place. First-time users will thus, ideally, become repeat users.
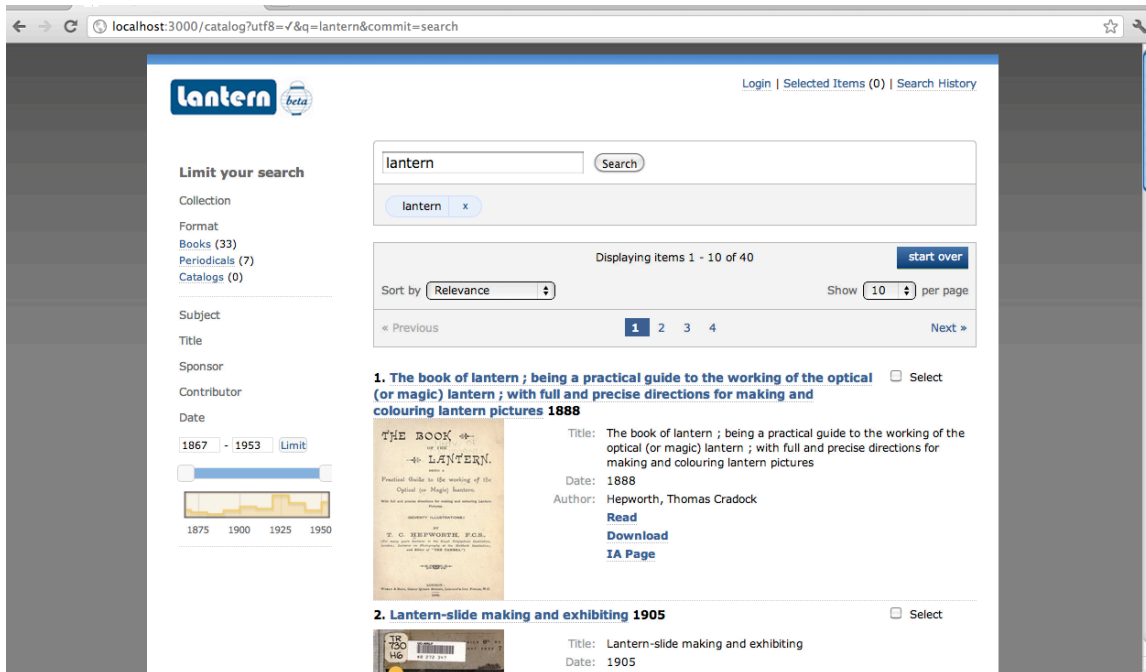
Figure 7. The new Lantern interface, built through customizing the Blacklight Solr interface, which supports faceted searches, was launched in fall 2012.

Lantern is a work in progress and will continue to improve over time. We also anticipate developing more user-oriented applications for the site in the future. As we do, we will adhere to our same core usability strategies: keep it familiar, simplify the navigation, and foreground the content.

Impact

We hope that MHDL's accessibility and usability will position the site to make a significant impact on the field of media studies and in the lives of everyday fans. If users return to the site, if they are able to discover answers to their existing questions as well as new ways of asking questions they've never even considered, then we will have succeeded. Eventually we hope to create innovative tools for mapping and visualizing the

collections over time—to connect MHDL's textual content with video clips and geolocate the evolution of television across the United States, perhaps, or to visualize programming at historic neighborhood theaters across the decades—but in this initial stage, we measure success in terms of the user: the user's identity, needs, and experience. The "what if" element of digital innovation is enticing, but ultimately every "what if" must be based on a "why" if those innovations are to find continued use among a wide audience.

Already, users as diverse as Luke McKernan, lead curator for the moving image at the British Library, *New Yorker* movies editor Richard Brody, and fashion editors from the Italian edition of *Vogue* have remarked on the potential impact of MHDL, but it will be our responsibility to ensure that users return to the site and feel that they gain something valuable from every visit. This is particularly challenging—and important—with a blended user base of experts and more casual fans. Appealing to visitors with varied interests requires us to rely on our instinct toward synthesis—the idea with which we began this exploration.

Synthesis

Ultimately, this theme of synthesis pervades the experience of constructing MHDL and hints at a larger significance for the project. David Pierce's initial implementation plan for the project required a synthesis of film history knowledge with copyright expertise and technical savvy. The scanning process demanded a coordinated, combined effort by the MHDL team, the Internet Archive, and the donors. Creation of the website mandated

a fusion of technological components, from WordPress and scripting languages to XML and Solr.

Further, the content of MHDL itself represents a sort of synthesis of library and archival material. The periodicals are published material and thus, by definition, were not originally unique or archival. As the number of extant copies of them has dwindled, however, what was once considered ubiquitous library material has become rare, nearly archival material, the provenance of which has been disrupted by the absence of complete runs in any one physical place. Through MHDL, the rare, non-circulating material of the archives is transformed into the collections of a lending library, free to be downloaded, annotated, and remixed. MHDL owes a great deal to collectors who had the foresight to gather and preserve these periodicals in the first place, maintaining them through the years when most libraries and other official repositories saw microfilm copies as preferred substitutes for the printed copies.

Overall, the Media History Digital Library represents a synthesis of existing tools and innovative uses. It relies on established cyberinfrastructure (Internet Archive, XML) but modifies these standard structures to create a specialized interface for research. We hope we can offer a collaborative model for other projects that aim to customize existing digital landscapes, to bolster the overall cyberinfrastructure while establishing their own realms—their own virtual stacks—for discovery and delight.

About the Authors

Eric Hoyt is assistant professor of communication arts at the University of Wisconsin–Madison. He is co-director of the Media History Digital Library and project lead of Lantern. His articles on media, law, and culture have appeared in *Cinema Journal*, *Film History*, *Jump Cut*, *World Policy Journal*, and the *International Journal of Learning and Media*. He is currently working on a book manuscript about the content libraries owned by the Hollywood studios.

Wendy Hagenmaier is digital archivist of the Media History Digital Library. She recently received her master's of science in information studies with a specialization in digital archives and preservation from the University of Texas–Austin School of Information, and is currently digital collections archivist at the Georgia Institute of Technology Library.

Carl Hagenmaier volunteers as Lantern's engineer. He works as senior director of engineering, hardware at Chargepoint, Inc., which is the developer of the Chargepoint network of electric vehicle charging stations.