



PUBLICACIONES DE LA
ASOCIACIÓN ARGENTINA
DE HUMANIDADES DIGITALES



HOYT, Eric y PIERCE, David. *Media History Digital Library (MHDL)*, 2011.

<https://mediahistoryproject.org/>

Reseña realizada por:

Pamela GIONCO

pamela.gionco@gmail.com

En principio, como internautas casuales, podemos pensar que *Media History Project* o *Media History Digital Library (MHDL)* se trata de un sitio web, que nos muestra una vitrina de tapas de revistas sobre cine y medios. O bien, que es un proyecto de digitalización masiva, que permite el acceso remoto a publicaciones de dominio público dispersas en distintas instituciones, una colección de millones de páginas recuperables a partir de búsquedas por palabra. Sin embargo, un análisis más detallado nos va a permitir comprender y dimensionar este proyecto digital de historia de los medios.

Media History Project no es una simple página web. Se trata de una plataforma que integra diversos recursos y herramientas digitales que pueden convertirse en fundamentales para los estudios de Cine y Audiovisuales, Comunicación e Información, entre otros campos posibles de estudio. Esta iniciativa, liderada actualmente por Eric Hoyt, se desarrolla desde hace más de diez años, construyendo un sistema de información que se retroalimenta y amplía progresivamente.

Los recursos y herramientas digitales que presenta este sistema se integran en forma virtuosa. El proyecto afirma sus cimientos en la *Media History Digital Library (MHDL)*, biblioteca digital fundada en 2009 por David Pierce y que se basa en la digitalización masiva y extensiva de publicaciones periódicas en dominio público vinculadas al cine, las artes audiovisuales, la radiodifusión y la grabación de sonido.

Desde sus inicios¹, el proyecto cuenta con el apoyo de Internet Archive², que dispone su infraestructura para permitir y garantizar el acceso y visualización, el almacenamiento y la preservación digital de los objetos digitales generados en el proceso. A lo largo de su existencia, ha contado con diversas contribuciones (monetarias y/o materiales) de instituciones, tales como



Library of Congress, University of Wisconsin-Madison o Mary Pickford Foundation, y de personas de algún modo conectadas con las publicaciones digitalizadas: coleccionistas, investigadores, personal especialista de museos y bibliotecas (curadores/as, referencistas, preservadores/as), aficionados y aficionadas³.

Las distintas publicaciones digitalizadas se organizan en colecciones temáticas tales como *Fan Magazine Collection (1911-1963)*, *Early Cinema Collection (1903-1928)* o *Technical Journals Collection (1916-1965)*, donde se reúnen varios títulos, indicando además el período en el que cada revista fue publicada. Cada una de las colecciones presenta una breve descripción general, que da cuenta del discurso curatorial y archivístico que determina estas agrupaciones, así como la mención sobre la procedencia de los materiales impresos originales que fueron digitalizados. Cada colección dispone también un listado de cada una de las publicaciones que integran el núcleo temático. Por tanto, es posible acceder a las colecciones temáticas y luego a cada publicación seriada. Los objetos digitales, que se alojan directamente en *Internet Archive*, pueden leerse en línea o descargarse. Se presenta entonces un extenso y colorido escaparate de la cultura impresa del siglo XX, cuyas portadas se exhiben digitalmente en la plataforma.

Nos interesa destacar la diversidad de países de origen y el multilingüismo en la selección de publicaciones digitalizadas. De interés para quienes investigan Cine Latinoamericano pueden ser dos publicaciones editadas en español, para distribuirse en países hispanohablantes: *Cine-Mundial (1916-1946)* y *Mensajero Paramount (1927-1933)*, ambas incluidas en la colección temática *Global Cinema Collection (1904-1957)*, donde encontramos además revistas chinas, francesas, alemanas, entre otros orígenes.

Al intentar acceder a la colección de todas las revistas de la plataforma, se nos propone explorar los repertorios por tema, o bien realizar búsquedas indistintamente por medio de *Lantern* o *Arclight*, herramientas que explicaremos más adelante. Por su parte, la colección *All Books (1837-1973)* presenta tres agrupamientos iniciales (libros más populares, sobre radio y sobre cine de principios del siglo XX) que se muestran, a modo de estantes virtuales, en bandas de imágenes, las portadas de cada publicación. A continuación, se encuentra una tabla de datos bibliográficos básicos (título, autor, fecha) que se presentan como puntos de acceso a las publicaciones, mediante enlaces (desde el título) a los objetos digitales alojados en *Internet Archive*.

Media History Project no es solamente un proyecto de digitalización masiva, que crece progresivamente (en 2011, se digitalizaron 200.000 páginas; en 2017, ya contaba con 2 millones de páginas digitalizadas). Ante esta inmensa cantidad de fuentes primarias digitalizadas, un equipo coordinado por Eric Hoyt desarrolló en 2013 el motor de búsqueda *Lantern*⁴ como una funcionalidad agregada a MHDL.

¹ *Media History Project*. MHDL: Project history and impact. Accesible desde: <https://bit.ly/3x73zYH>.

² *Internet Archive*. MHDL. Accesible desde: <https://archive.org/details/mediahistory>.

³ La lista completa de sponsors y colaboradores/as puede encontrarse en *Media History Project*, About us. Accesible desde: <https://mediahistoryproject.org/about.html>.

⁴ *Lantern: Search, Visualize & Explore the Media History Digital Library*. Accesible desde: <http://lantern.mediahist.org/>.

La búsqueda por medio de dicha herramienta se realiza en todo el texto de (casi) todas las publicaciones que integran la MHDl, que fueron procesadas con reconocimiento óptico de caracteres (OCR, por sus siglas en inglés, *Optical Character Recognition*). La recuperación de la información es listada por página, pudiendo visualizar 10, 20, 50 o 100 recursos a la vez, y ordenar por relevancia o por fecha (del más antiguo al más actual, y viceversa). También es posible filtrar la búsqueda por rangos de fecha (*Date*), título de la publicación (*Title*), colección (*Collection*), tipo documental de la publicación (*Format*) o idioma (*Language*). Cada página recuperada en la búsqueda nos indica los metadatos completos de la publicación, además de interesantes opciones de lectura y descarga: es posible leer la página en el contexto de su publicación, descargar la página en pequeño o gran tamaño (en formato JPG), descargar el volumen completo (en PDF), descargar los archivos asociados al objeto digital (en XML, JPS, TXT) o visitar el sitio *Internet Archive* que aloja la obra digitalizada.

Al momento de poner en funcionamiento *Lantern*, Eric Hoyt y miembros del equipo (Hoyt, Hagenmaier y Hagenmaier, 2013) dieron cuenta de los desafíos del diseño de la plataforma en función del acceso y usabilidad de los recursos digitales. Hoyt (2013) aporta también cinco lecciones sobre el motor de búsqueda desarrollado: la importancia (para la recuperación de información) de la curación de colecciones, la comprensión de la búsqueda por índices conformados por datos y metadatos (no por contenido real), el diseño de algoritmos de búsqueda en escala (y no en profundidad), la relevancia (e imperfección) del OCR para la búsqueda y la tendencia hacia la visualidad de la experiencia de búsqueda en entornos digitales. Este mismo autor, líder del equipo de desarrollo de *Lantern* y actual responsable del *Media History Project*, reconoce que aproximadamente el 50% de las publicaciones sobre cine de la primera mitad del siglo XX no han sido utilizadas aún como fuentes primarias de investigación, dado que se privilegian, en los Estudios sobre Cine, publicaciones comerciales de gran tirada y amplia difusión como *Variety* (Hoyt, 2014).

A este punto, podemos mencionar brevemente producciones científicas que han utilizado MHDl y *Lantern* para acceder a fuentes de investigación. David Bordwell (2013), profesor de la *University of Wisconsin-Madison*, es uno de los primeros investigadores en probar el motor de búsqueda, a instancias del mismo equipo de desarrollo. Tamar Jeffers McDonald, convocada por el blog *Film Studies for Free* (Grant, 2013) realiza una búsqueda exploratoria en MHDl, para descubrir fuentes que nunca había visto sobre la figura de Doris Day, a quien había investigado extensamente en bibliotecas y archivos. Katherine Nagels (2012) reconoce en su trabajo el acceso a publicaciones tempranas de cine digitalizadas por *Media History Project*. Tim Snelson (2017) analiza las relaciones complejas entre exhibidores y audiencias a partir de una serie de fuentes vinculadas a la producción, distribución y exhibición de cine estadounidense de la década del 30, particularmente *Variety*, *Film Daily* y *Motion Picture Herald*, consultadas en MHDl. Lila Silva Foster (2017) rastrea los catálogos de la filmoteca Pathé-Baby para estudiar las temáticas propuestas por esta compañía para consumo doméstico en Brasil, y Castaldo Lunden (2018) publica un dossier sobre las relaciones entre moda, cine y medios. Ambas autoras acceden a materiales digitalizados por *Media History Project*.

En 2014, el equipo que desarrolla *Lantern* recibe el premio *Anne Friedberg Award for Innovative Scholarship*, otorgado por la *Society for Cinema and Media Studies* (SCMS). Por primera vez el premio no es entregado a un texto impreso sino a un proyecto digital. En el último congreso de la SCMS, realizado este mismo año (2021), gran parte de los trabajos presentados reconocían en redes sociales el uso de la plataforma *Media History Project*.

Un detalle no menor a tener en cuenta: si bien las fuentes digitalizadas son multilingües, el motor de búsqueda no lo es, por ejemplo, la búsqueda del término *Méliès* recupera 57 recursos, mientras que *melies* recupera 3637⁵. Esta limitación puede saldarse si entendemos las *lecciones* propuestas por Hoyt (2013): los índices de datos son más efectivos y eficientes en texto plano. Los acentos pueden volverse un obstáculo para la recuperación de la información.

Media History Project no es solo una colección de millones de páginas de revistas de cine recuperables a partir de búsquedas por palabra. Como hemos visto, lo que se encuentra detrás de esta posibilidad es la sistemática *datificación* de esos materiales impresos digitalizados. La elaboración de *datasets* que organizan la información digital construida a partir de los procesos anteriores permite realizar estudios analíticos de datos. En 2014, Eric Hoyt y Charles Acland reciben un subsidio de *Digging into Data* para desarrollar una herramienta llamada *ArLight*⁶, que permite visualizar tendencias de palabras en publicaciones digitalizadas y procesadas por *Media History Project*. Tal como lo conciben y describen sus autores, aplican “la analítica de Twitter a la historia del cine y los medios”⁷, es decir, la idea es analizar mediante *data mining* y visualizar los *trending topics* en las publicaciones digitalizadas por MHDL. Acland y Hoyt (2016) editan una guía de uso y experiencias de aplicación de *ArLight*, como una combinación virtuosa entre la Historia de los Medios y las *Digital Humanities*.

Así como la búsqueda a través de *Lantern* no es multilingüe, los datos que se pueden analizar mediante el entorno *ArLight* también resultan limitados, ya que, en este caso, la búsqueda se realiza solo en algunas de las colecciones (en inglés) curadas por MHDL. Lo cierto es que, aun así, es posible rastrear en estas publicaciones tópicos relevantes para la Historia del Cine y los Medios en países no angloparlantes, ideando estrategias de búsqueda que conozcan estas limitaciones. Hoyt (2019) expresa que a partir de *Media Digital Project* es posible no solo acceder a fuentes primarias que habían sido dejadas de lado, sino también estudiarlas en una escala completamente diferente. *Mapping Movie Magazines*, libro de reciente publicación, editado por Daniel Biltereyst y Lies Van de Vijver (2020) presenta un uso intensivo (y extensivo) de la integración de las herramientas desarrolladas por *Media History Project*.

Podemos decir que estos tres componentes principales (MHDL, *Lantern* y *ArLight*) del sistema de información *Media History Project* apuntan a personas usuarias distintas del otro lado de la pantalla. Contamos con imágenes digitales de las revistas, disponibles en línea a través de la MHDL,

⁵ Esta búsqueda es realizada en 2013 por Eric Hoyt. “A search for ‘Méliès’ returns 57 hits. And if you want to cast a much wider net, the search ‘melies’ returns 2,680 results!” (comentarios a Hoyt, 2013).

⁶ Accesible desde: <https://projectarlight.org/>.

⁷ Eric Hoyt. Digital Projects. Accesible desde: <http://erichoyt.org/projects.html>.

que habilita la lectura casual y el consumo despreocupado por parte de un público general, que no necesariamente tiene objetivos académicos. Luego, quienes accedan a estos materiales digitalizados como fuentes primarias de investigación encontrarán el recorrido página a página como tedioso, similar al realizado presencialmente en bibliotecas y archivos, al recorrer una publicación en soporte papel. La búsqueda por palabra entre millones de imágenes de esas páginas a partir del OCR, con las advertencias del caso, permite la recuperación de recursos de información que pueden ser filtrados, ordenados y descargados para su análisis. Es quizás esta funcionalidad la más utilizada por quienes realizan investigación en Cine y Artes Audiovisuales. Por último, la visualización de la tendencia de palabras en las publicaciones presenta una nueva perspectiva para quienes entienden lo que es posible hacer con los datos, desde la analítica cultural.

La plataforma cuenta además con una sección de reciente aparición (*Teaching*)⁸ vinculada al aprendizaje y la enseñanza. Este apartado se presenta como un espacio que pone a disposición recursos tanto para estudiantes como para docentes. Por una parte, los recursos para estudiantes son en realidad tutoriales y consejos de uso de los tres componentes principales de *Media History Project*. De ninguna manera debe desmerecerse esta sección, ya que se trata de una introducción clara y exhaustiva sobre cada uno de estos entornos, describiendo sus posibilidades de acceso y sus interfaces gráficas (tanto de la plataforma del *Media History Project* como de *Internet Archive*), así como de posibles estrategias de búsqueda y recuperación. Por otra parte, en el segmento de *recursos para docentes* se publican materiales didácticos para utilizar en el aula, al tiempo que invitan a docentes que quieran compartir el diseño de actividades en las cuales se utilice operativamente la plataforma. El envío de la propuesta se puede realizar por correo electrónico para ser incorporado a la página. Al momento, los recursos disponibles son para nivel universitario (de instituciones norteamericanas) y se basan en búsquedas de información en la plataforma, como estrategia de aprendizaje de usos de fuentes primarias y secundarias en la investigación. De esta forma, se pueden dar unos primeros pasos en la integración de las Humanidades Digitales en las prácticas docentes. Cada propuesta presenta: un título descriptivo; la mención de quien la diseñó (nombre, cargo e institución); el título y nivel educativo del curso en el que se aplicó; una breve descripción de la tarea asignada al grupo de estudiantes; los resultados de aprendizaje; y un archivo descargable con el diseño de la actividad.

Con todo, *Media History Project* es una posible respuesta a la pregunta latente en cualquier proyecto de digitalización: qué podemos hacer con nuestras colecciones digitales. La conversión digital como tal resuelve parcialmente los problemas de acceso a las publicaciones analógicas, pero en entornos digitales ya no alcanza con acceder visualmente a fuentes digitalizadas como imágenes: surge la imperiosa necesidad de que esas fuentes sean procesadas para ser recuperables, computables y reutilizables. El equipo de trabajo detrás de este proyecto promueve y facilita nuevas estrategias de investigación de Humanidades Digitales/*Digital Humanities* y las acerca a las

⁸ MHDL: *Teaching resources*. Accesible desde: <https://mediahistoryproject.org/teaching/index.html>.

prácticas de la *New Cinema History*, corriente teórica dentro del ámbito académico angloparlante, que propone renovar los Estudios de Cine y Artes Audiovisuales tradicionales. Esta combinación metodológica y epistémica tiene impacto transformativo en este campo, en tanto permite no solo plantearse nuevas preguntas, sino también recorrer caminos digitales innovadores para elaborar renovadas perspectivas de análisis, también desde Latinoamérica.

REFERENCIAS BIBLIOGRÁFICAS

- Acland, C. R. y Hoyt, E. (Eds.). (2016). *The Arclight Guidebook to Media History and the Digital Humanities*. REFRAME/Project Arclight. <https://projectarclight.org/book/>
- Biltreyst, D. y Van de Vijver, L. (Eds.). (2020). *Mapping Movie Magazines. Digitization, Periodicals and Cinema History*. Palgrave Macmillan. <https://doi.org/10.1007/978-3-030-33277-8>
- Bordwell, D. (12 de agosto de 2013). Magic, this Lantern. *Observation on film art*. [Entrada de blog]. <http://www.davidbordwell.net/blog/2013/08/12/magic-this-lantern/>
- Castaldo Lunden, E. (2018). Introduction: Exploring the Intersections of Fashion, Film, and Media. *Networking Knowledge. Journal of the MeCCSA Postgraduate Network*, 11(1), 1-6. <https://doi.org/10.31165/nk.2018.111.527>
- Foster, L. (2017). Modernidade à brasileira: Ideário moderno e cinema amador na revista ilustrada Cinearte. *Vivomatografías. Revista de estudios sobre precine y cine silente en Latinoamérica*, 3(3). <https://bit.ly/3kaGArm>
- Grant, C. (28 de octubre de 2013). Studying Movie Magazines and Fan Culture! Online Research and Methodology Resources. And LANTERN! *Film Studies for Free*. [Entrada de blog]. <https://filmstudiesforfree.blogspot.com/2013/10/studying-movie-magazines-and-fan.html>
- Hoyt, E. (14 de agosto de 2013). Let's talk about search: Some lessons from building Lantern. *Antenna*. [Entrada de blog]. <http://blog.commarts.wisc.edu/2013/08/14/lets-talk-about-search-some-lessons-from-building-lantern/>
- Hoyt, E. (2014). Lenses for Lantern: Data Mining, Visualization, and Excavating Film History's Neglected Sources. *Film History*, 26(2). <https://doi.org/10.2979/filmhistory.26.2.146>
- Hoyt, E. (2019). Arclight and Zoom Lenses. Searching for Influential Exhibition in Film History's Big Data. En D. Biltreyst, R. Maltby y P. Meers (Eds.), *The Routledge Companion to New Cinema History*. Routledge.
- Hoyt, E., Hagenmaier, W. y Hagenmaier, C. (2013). Media + History + Digital + Library: An Experiment in Synthesis. *e-Media Studies*, 3(1). <https://doi.org/10.1349/ps1.1938-6060.a.430>
- Nagels, K. (2012). "Those funny subtitles": Silent Film Intertitles in Exhibition and Discourse. *Early Popular Visual Culture*, 10(4). <https://doi.org/10.1080/17460654.2012.724570>
- Snelson, T. (2017). "They'll Be Dancing in the Aisles!": Youth Audiences, Cinema Exhibition and The Mid-1930s Swing Boom. *Historical Journal of Film, Radio and Television*, 37(3). <https://doi.org/10.1080/01439685.2016.1187850>