**Local Cinema History at Scale: Data and Methods for Comparative Exhibition Studies**

**Institutional affiliations:** University of Oregon

**Abstract**

Digital tools and digitized sources have expanded our ability to research and present regional film histories, along with the hope of conducting comparative work across both place and time. Alongside these projects are increasing calls for more deliberate coordination of tools, methods, and sources to create more meaningful comparisons. However, it remains difficult for researchers to know what digital projects exist for comparative work, and the methods, points of comparison, data structure, and sources used all considerably vary. Utilizing research data management principles, we conducted an exploratory survey of local film exhibition digital projects to document the current historiographic landscape, and to assess existing coverage of geography, time, sources, data structures, metadata schema, data accessibility and reproducibility. The dataset from the survey results can be shared by researchers to better discover each other's work, but also to serve as a guide to best practices going forward.

**Author bios**

Michael Aronson is an Associate Professor of Cinema Studies at the University of Oregon. The author of *Nickelodeon City: Pittsburgh at the Movies, 1905-1929* his research focuses on exhibition and industrial practices and has also appeared in a number of journals including, *The Moving Image*, *Film History* and *Cinema Journal*.

Gabriele Hayden is Librarian for Research Data Management and Reproducibility at the University of Oregon. She writes and presents on issues in data sharing and reproducibility in the humanities and social sciences. In her previous career as English professor she studied the role of translation from Spanish in controversies over race and style in US modernist poetry.

Elizabeth Peterson is a Humanities Librarian at the University of Oregon. Her research on Oregon film history can be found in *Film History, The Moving Image*, and *Oregon Historical Quarterly.* She is the author of the book *Tribal Libraries in the United States* (McFarland, 2007).

**Keywords**

film exhibition, New Cinema History, digital humanities, data reproducibility, data management

## Introduction

Our historiographic intentions were pure. We wanted to build a web-based platform in which our undergraduate students could be taught to collaboratively research, write and map the history of moviegoing and exhibition in cities and towns across the state of Oregon. One of the primary goals of the Oregon Theater Project[1] was therefore pedagogical: we built it to give our students hands-on experience with digital humanities principles and tools, and to serve as a public forum for the students' resulting original research. In this regard, we have happily succeeded. However, our other key project objective was to incorporate open data design principles to facilitate an accessible and sustainable database of information about Oregon exhibition history--data that could then be freely utilized by film scholars, ourselves and others, to conduct comparative analysis across both time and space. Here, at least until this point, we have sadly fallen short, and, unfortunately, we're not alone.

> *Our pronouns for this article are "we/us" as this is a deeply collaborative set of projects. But to clarify for our readers, there are two overlapping sets of we/us. The first, the Oregon Theater Project, is an existing and ongoing partnership between a film historian, Michael Aronson, and a Humanities Librarian, Elizabeth Peterson. However, when it came to our goals to better share our Oregon data alongside an analysis of available datasets from similar exhibition and moviegoing projects we realized we were well out of our depth. We needed someone with expertise in DMPs, PIDs, TDRs, etc., all things data that we did not fully understand. So the we/us that encompasses the work of this paper includes our third collaborator, University of Oregon's Research Data Management and Reproducibility Librarian, Gabriele Hayden.*

As is obvious to anyone who has undertaken this kind of work, the need for resources, including expert labor, remains indispensable and widely uneven in their availability. While there are evermore small-d democratic tools for doing digital scholarship and creating digital collections, there are also, we should all acknowledge, forgivable limits to what one can

[1] Oregon Theater Project, accessed August 30, 2021, http://oregontheaterproject.uoregon.edu.

accomplish on one's own or with a research partner. Some film historians are well-versed in the digital tools, datasets and back-end hierarchies, and some film historians are able to collaborate with data scientists, and some of us are largely on our own. Still, these tools in conjunction with a growing set of digitized periodical sources have increasingly expanded our collective ability to research and publicly share local, regional and national film histories and the resulting data that structure these records. Amazing, richly detailed histories of exhibition and moviegoing from individual theaters[2] to entire countries[3] are now available for anyone to access on the "shallow" web. In terms of offering up historical content, it's breathtaking how far we have come in the last couple of decades. However, it remains surprisingly difficult for researchers to know where and how to locate the growing plethora of these online projects--there is no real clearinghouse, directory or repository that focuses exclusively on digital projects--and the methods, points of comparison, data structures and sources used by those that are locatable vary considerably. As Richard Maltby noted, now over a decade ago, "[t]he increasing interest in the specific circumstances in which films were circulated and watched historically has resulted in a proliferation of often local or regional case studies. A key challenge facing this field of research is to integrate the diversity of local microhistorical studies into interpretative frameworks at the meso level."[4] If we agree, ten years later, that easily shareable data is an intrinsic building block required for this goal, it remains a surprisingly elusive one.

As an initial remedy to these challenges, we have produced an exploratory survey of existing digital exhibition projects to document the historiographic landscape, in which we have assessed the collective existing coverage of geography, time, sources, data structures, metadata schema, and data accessibility. Simply put, what is currently out there and how usable is it by others? And because it is bad form to call out the mote in others' eyes while walking around with a beam in our own, we also document here our own better-late-than-never process of implementing accessible data sharing for the Oregon Theater Project. Our goal with these two intertwined efforts is to mark a modest and preliminary path for others to hopefully follow, so that our growing collective potential of dispersed data can be integrated into new, original comparative research on the history of moviegoing and exhibition.

---

[2] Cine ZOOlogie, accessed September 1, 2021, http://www.cinemazoologie.be/

[3] Cinema Context: Film in the Netherlands from 1896, accessed September 1, 2021, https://cinemacontext.nl/

[4] Richard Maltby, "New Cinema Histories," in *Explorations in New Cinema History: Approaches and Case Studies,* eds. Richard Maltby, Daniel Biltereyst and Phillippe Meers (Chichester, UK: Wiley-Blackwell, 2011), 1-40.

The following paper is broken into three sections, a description of our exploratory survey methodology, its findings, and a set of recommendations based on it, as well as our own process of making the data accessible for our Oregon platform.

**Methodology**

When we proposed this paper, our goal was to provide an initial comparative dataset derived from the existing accessible exhibition-focused platforms found on the open web, and to hopefully add our data to the mix. We have arrived at a moment of collective desire for comparative analysis of local cinema histories, even if we're not yet always certain as to what constitutes a meaningful comparison. Where and when is it possible (and useful) to generalize about the unique? Daniel Biltereyst and Philippe Meers[5] recognized the need for "more systematic comparative research" with shared datasets from different areas to enable discovery of "larger patterns" and to test hypotheses within film exhibition and cinemagoing. And some of this valuable work has begun; Pafort-Overduin, Sedgwick, and Van de Vijver[6] in one project focus on the film titles consumed in three different cities in the Netherlands and Great Britain in the mid-1930s, and Thunnis van Oort[7] compares "cinema economies and moviegoing cultures" in the Netherlands and Belgium in the immediate post-WWII period by looking at the industrial organization of film exhibitors in the respective countries. More recently, a large research group used "harmonized data sets" of feature films screened in cinemas in six European cities in 1952 to track and analyze distribution patterns.[8] The CFP for this issue of *Iluminace* prompted us to add our voice to the call for more comparative work along with our recommendation to structure individual projects in ways that make them more comparable. But first we needed to know what work currently exists within this realm. How does the Oregon Theater Project's content and structure compare to other projects within this field?

***Definitions, Criteria for Inclusion***

[5] Daniel Biltereyst and Philippe Meers, "New Cinema History and the Comparative Mode: Reflections on Comparing Historical Cinema Cultures," *Alphaville* 11 (2016), 13–32.
[6] Clara Pafort-Overduin, John Sedgwick, and Lies Van de Vijver, "Identifying Cinema Cultures and Audience Preferences: A Comparative Analysis of Audience Choice and Popularity in Three Medium-sized Northern European Cities in the Mid-1930s," *TMG Journal for Media History* 21, no. 1 (2018), 102-118.
[7] Thunnis van Oort,"Industrial Organization of Film Exhibitors in the Low Countries: Comparing the Netherlands and Belgium, 1945–1960," *Historical Journal of Film, Radio and Television* 37, no. 3 (2016), 475-498.
[8] Thunnis van Oort, and Jessica Leonora Whitehead, "Common Ground: Comparative Histories of Cinema Audiences," *TMG Journal for Media History* 23, no. 1-2 (2020), 1-11.

We are not scientists. That is not stated as an excuse for any perceived lack of rigor, but rather as context for the underlying rationale we used in locating and analyzing the projects in this survey of existing public-facing digital projects on local/regional film exhibition history. First, it is important here to state the obvious; there are many significant works of exhibition history, including both recent and canonical studies, that because of their qualitative-nature or presentational form do not lend themselves for easy inclusion in this survey.[9] Our goal with this project, instead, is to expand access to datasets that are, or should be, easily found online by anyone interested, by those that are explicitly collaborators in shared research, but also scholars (both lay and professional) with more autonomous pursuits. Data should be shared and it should be straightforward to locate. However, traditional means of distributing scholarly work, within mostly closed academic and commercial ecosystems, can sometimes be at odds with discoverability. As a result, it's important to emphasize here that our project is an *exploratory survey* of a particular presentation mode of research and not an *inventory* of all the research taking place within the field. The survey and its results represent a snapshot in time; they should not be considered definitive or final. Even during the relatively short time it took to write and revise this article, several of the projects we surveyed changed in significant ways, and new projects will likely come online by the time this article is published.

This inevitable evolution further underscores the importance of ensuring access to the data behind these digital projects. To that end, we are sharing this survey in a repository[10] with a license that allows for reuse, with the hope that others download our data, add to it, and, ultimately, publish a more comprehensive version of this survey that includes important work--particularly work in other languages and from Asia, Africa, and Eastern Europe--that we have failed to locate.

We started by defining "existing digital projects on local/regional film exhibition history." This preference for local/regional over universal/global is consistent with the broad

[9] See, for instance: Jacqueline Najuma Stewart, *Migrating to the Movies: Cinema and Black Urban Modernity*, (Berkeley: University of California Press, 2005); Charlene Regester, "From the Buzzard's Roost: Black Movie-Going in Durham and Other North Carolina Cities During the Early Period of American Cinema," *Film History* 17, no. 1 (2005), 113-24.

[10] Gabriele Hayden, Elizabeth Peterson, and Michael Aronson, "Replication Data for: Local Cinema History at Scale: Data and Methods for Comparative Exhibition Studies" (Harvard Dataverse), accessed March 14, 2022, https://doi.org/10.7910/DVN/6WOQPO.

"bottom up" goals of New Cinema History.[11] Although no longer very new at this point (and arguably not even 'new' when first labeled as such), New Cinema History, in its decentering of the film as text, and in its multivalent approaches to studying movies as cultural and social phenomena, arguably remains one of the more vibrant subfields in our discipline, in large part for the very methodological and interdisciplinary creativity that fostered many of our surveyed projects.

To be included in our exploratory survey, a project has the following basic criteria:

- includes information about venues that focused primarily on moving pictures, e.g., names of venues, locations, capacity, dates of operation
- be currently publicly visible online (even if it does not yet display the information it promises, or its display is broken)

Projects may also include the following additional elements:

- programming, e.g., titles, dates of screening, and venues where screened
- names of people associated with venues, e.g., owners, managers, staff
- audience demographic information
- maps or other geospatial information
- images
- timelines
- stories about moviegoing, e.g., oral histories, user-submitted memories
- information about sources, including citations, images of or links to primary source material such as newspaper clippings and city directories

Projects may have extensive interpretive text about venues, people, and topics, or they might simply provide the data with minimal narrative framing.

In addition to these content elements, we also wanted to survey the data components of projects. Each web project, almost by definition, typically has a front end--the public website--that users can navigate, along with a back end--the data and source code--that determines what

[11] Daniel Biltereyst, Richard Maltby, and Philippe Meers, Philippe, eds. *The Routledge Companion to New Cinema History* (New York: Routledge, 2019).

users see on the website.[12] If we're hoping to do comparative work, a basic requirement is access to more than the public-facing website.[13] Thus, we designed our survey instrument to document how projects structure their data, whether or not it is freely available as a machine-readable data set that one can download without contacting a person, and, if so, in what formats, the existence of persistent links, and what permissions exist for re-use. We utilized the FAIR principles[14] as a guide, as they provide a simple and understandable framework for data reproducibility. The FAIR guidelines recommend ways to make data **F**indable, **A**ccessible, **I**nteroperable, and **R**eproducible, all crucial elements that must be in place to enable researchers to share and compare data.

With these criteria as a determining structure, our exploratory survey excludes journal articles, books and book chapters, conference presentations, and dissertations unless they also had an online digital project component that met the minimal criteria. Consequently, the survey leaves out many potentially notable and interesting research projects about local film exhibition history.

***Literature Review***

Our first stop to compile the list of projects was The HoMER Network's map and list of research projects.[15] The HoMER Network (History of Moviegoing, Exhibition, and Reception) is the primary international scholarly association for people working within this specialized subfield of cinema studies. Their website includes an interactive map and list of research projects, which anyone can submit to using a provided form. The map allows one to filter results by type of project: database, mapping, oral history, app, digitisation, and N/A. In April 2021 when we started our survey, there were 66 projects listed, but only a handful of those could be characterized as digital projects, and some of those were no longer online. We'll return to the idea of a potential HoMER directory version 2.0 in the Recommendations section of this paper, but nonetheless, HoMER was a good place to start because it pointed us to active scholars, who in turn cited other work that is online but which is not yet on the HoMER list.

---

[12] We are using back end and front end here in their widely accepted vernacular meaning: "back end (data, concepts, research concepts) and front end (interface)," Joanna Drucker and Patrik BO Svensson, "The Why and How of Middleware," *Digital Humanities Quarterly,* 010, no. 2 (May 19, 2016), paragraph 60.

[13] "Nous lançons un appel pour l'accès libre aux données et aux métadonnées. Celles-ci doivent être documentées et interopérables, autant techniquement que conceptuellement." Marin Dacos, "Manifeste des Digital humanities," Billet, *THATCamp Paris* (blog), March 26, 2011, https://tcp.hypotheses.org/318.

[14] "FAIR Principles," GOFAIR, accessed September 1, 2021, https://www.go-fair.org/fair-principles.

[15] "HoMER Projects," HoMER Network, accessed September 1, 2021, https://homernetwork.org/homer-projects.

Another productive source for identifying projects was a November 2020 special issue of *TMG Journal for Media History,*[16] which published a set of articles on the theme of Comparative Histories of Moviegoing. Articles in this issue both described existing digital projects and cited other projects that we were able to locate and add to our list. *Alphaville* provided a good number of leads in Issue 11 on European cinema heritage, as well as a more recent dossier in Issue 21 titled *History of Cinemagoing: Archives and Projects.* We also searched and followed cited references in the journals *Research Data Journal for the Humanities and Social Sciences*, the *Journal of Open Humanities Data*, and *The Arclight Guidebook to Media History and Digital Humanities.*[17] Helpful inventories of European projects included the European Association for Digital Humanities projects,[18] the Cinema City Cultures research network website,[19] "DH Cinema Projects" listed on the Transformations Conference website,[20] and "Mapping Performing Arts Data" by Vincent Baptist.[21] Lastly, we sent an email inquiry to the HoMER listserv to solicit information about projects that fit within the scope of our survey.

At the conclusion of this initial gathering and sifting process, we had 35 titles on our list of online projects to survey, a number that grew to 40 with several last minute additions. Again, finding just over three dozen sites required this set of scholars, fairly well-versed in the field, considerable time to locate across multiple and diverse sources.

***Survey Design***

We used Google Sheets to create the survey instrument. The survey has two worksheets: one to document the public-facing content of a project, and one to document the back-end data components of a project. On the content sheet, we created fields that capture the kinds of information each project includes. These fields align with the core elements of film exhibition research, e.g., venues, dates of operation, capacity, programming, etc., that are most likely to be used in comparative work.

---

[16] *TMG Journal for Media History* 23, no. 1-2 (2020).

[17] Charles Acland and Eric Hoyt, eds., *The Arclight Guidebook to Media History and the Digital Humanities* (Reframe Books, 2016).

[18] "Projects," European Association for Digital Humanities, accessed August 25, 2021, https://eadh.org/projects.

[19] Cinema City Cultures, accessed August 30, 2021, http://cinemacitycultures.org/.

[20] "DH Cinema Projects," Transformations I Cinema and Media Studies Research Meets Digital Humanities Tools, accessed August 29, 2021, https://transformationsconference.net/dh-cinema-projects/.

[21] Vincent Baptist, Julia Noordegraaf, and Thunnis Van Oort, "A Digital Toolkit to Detect Cinema Audiences of the Silent Era: Scalable Perspectives on Film Exhibition and Consumption in Amsterdam Neighbourhoods (1907-1928)," *Studies in European Cinema* 18, no. 3 (July 11, 2021): 252–273. doi:10.1080/17411548.2021.1930974.

We ensured consistency by writing a data dictionary that defined in advance the potential answers to each question. We used Google's data validation feature for many of the fields, creating a dropdown menu with predefined answers. We used the International Organization for Standardization (ISO) Country Codes[22] to indicate the countries covered in a project, but we used free text for the "Region(s)" field, since we wanted a less structured way to capture the varying scales at which projects operated--whether cities, states, provinces, or neighborhoods. Our primary focus for the front-end survey was on simply documenting the existence of content on each project's public website. We did not incorporate more specific descriptions of how that content is displayed or how it functions. For example, if a project incorporates a map of some kind, we would indicate "Yes" on the survey sheet, but not what kind of map, if it is static or interactive, if it incorporates different media, if it has different layers, etc. Additionally, we did not survey the presence of specific digital humanities tools, such as timelines and data visualizations, although many of the projects in the survey take advantage of these tools to present and illuminate research in new and exciting ways that traditional print publications cannot. Other public-facing website components we chose not to examine include usability, accessibility, intended audience, use of images, the presence of interactive tools, or the functionality of tools such as searchable databases or dynamic maps. We did indicate if a project covers a single region or if it aggregates data from multiple locations, but aside from that we chose not to document the spectrum of approaches and methods within the projects we surveyed, such as the use of digitized newspapers to compile programming information, recording oral histories of theaters to document perceptions of audiences, or the use of linked open data. All of these components would be interesting to add to a future survey and analysis of these projects.

As much as one might hope for an easy checklist that would seamlessly work to assess any and all of the projects on our list, there is so much variety in terms of content and presentation that we still had to make many subjective judgments about what to include and how to describe included projects. Some presented a searchable database of venues that included dates of operation, ownership, programs, and lists of sources--Cinema Context, Cinematic

[22] "ISO 3166 Country Codes," International Organization for Standardization, accessed August 30, 2021, https://www.iso.org/iso-3166-country-codes.html.

Brno,[23] The Early Cinema in Scotland Research Project.[24] Others required more deliberation. For example, the U.K.'s Cinema Theatre Association's "Listed Cinemas" is a static list of current and historical cinema venues.[25] The list isn't interactive or searchable, there is no programming information about the venues, nor does the list incorporate any of the digital humanities features we find in other projects. But because the list includes addresses, years of construction, names of architects, and current use--data that could be useful for comparative work--we decided it met the minimum threshold for inclusion in the survey. Another tricky example is the Italian CineRicordi (CineMemories),[26] which provides video oral histories of moviegoing in mid-century Italy. This is a rich collection of primary source documents about audiences that captures what it felt like to go to the movies in that place and time. This kind of qualitative data doesn't lend itself to comparison as tidily as, say, the number of theaters in operation at a given time, but the information is still an important contribution to our understanding of the cinema experience for ordinary people. Participants mention specific theaters and programs in their memory videos, but the project's mission is not to systematically document venues or programs. Thus, we included CineRicordi and a few other oral history projects like it in the survey, but when indicating the presence of data related to venues and programs in these projects in the "Venues" and "Programming" fields of the survey, we entered "No" or "Partially" to try to describe this partial and ambiguous presence of potentially usable data.

We also decided to include projects we defined as "Placeholder" projects. Some of these serve as landing pages for research networks focused on regional film exhibition histories, with information about and links to articles, presentations, dissertations, and digital projects. The five sites we included in the survey are frequently mentioned in the literature, and while they mention places to find data about venues, programming, etc., the sites themselves do not present that data. So while it seemed necessary to include these projects in the survey, most of the content elements we look for appear absent. In some cases, we found through reading articles or presentations links to the data being used in (but not linked to) the front-end projects that these placeholder websites "tease." We marked many of the fields for these projects "Undetermined,"

[23] Cinematic Brno, Department of Film and Audiovisual Culture at Faculty of Philosophy, Masaryk University, Brno, accessed August 31, 2021, https://www.phil.muni.cz/filmovebrno/?id=103&lang=1.
[24] The Early Cinema in Scotland Research Project, accessed September 1, 2021, https://earlycinema.gla.ac.uk/.
[25] "Listed Cinemas," Cinema Theatre Association, accessed August 30, 2021, https://cinema-theatre.org.uk/uk-cinemas/listed-cinemas/.
[26] CineRicordi, accessed August 31, 2021, https://www.cinericordi.it/.

since we could not determine what elements might ultimately be included in the as yet unpublished web display.

Finally, we included a field in the survey to document the presence of cited sources. Most of these projects, including our own, originated within a scholarly framework. A fundamental norm of scholarship is to attribute the sources one uses to both acknowledge others' work and to allow other researchers to consult those sources with an eye to develop their own interpretations and arguments. It is labor-intensive, to be sure, but the practice ensures that these projects preserve a scholarly integrity and authority that aligns with traditional academic publications. When citations and sources are present in these projects, there is enough variation in form and content that we had to include an option in the data sheet for "Partially" and "Undetermined." For example, Lost Cinemas of Leeds has a "Sources" page with a list of the sources used, such as city directories, newspapers, books, and manuscript collections, but the theater profiles on the site do not have specific citations to those sources.[27] We marked this one and others that take this approach as "Partially." In contrast, the Australian Cinemas Map relies on a single publication for its records, the film industry journal *Film Weekly,*[28] yet each theater profile has more specific references to *Film Weekly* that would enable a researcher to find those references readily to confirm the information. For this reason, we gave this project and others with a similar level of rigor a "Yes" for the presence of sources.

The information we collected regarding the data that fuels each project focuses on basic elements that research data management librarians have long argued are essential to the sharing of research data and code. Because our focus is on allowing for comparative work, we focus on data sharing, rather than on elements that humanities scholars have defined as important to reproducing digital project front ends.[29] And within those elements, we focused on those aspects we feel are most essential to allowing cinema studies scholars with limited technical and financial resources to reuse data to advance the field: that the data is shared at all, that it is

---

[27] "Sources," Hiding in Plain Sight: Lost Cinemas of Leeds, accessed August 30, 2021, https://lostcinemas.co.uk/sources/.

[28] Australian Cinemas Map: A Map of Film Weekly Motion Picture Directory Cinema Data, 1948-1971, Flinders Institute for Research in the Humanities, accessed August 31, 2021, https://auscinemas.flinders.edu.au/.

[29] For a good recent overview of challenges associated with that second project, see the introduction to Florian Kräutli, Esther Chen, and Matteo Valleriani, "Linked Data Strategies for Conserving Digital Research Outputs: The Shelf Life of Digital Humanities," in *Information and Knowledge Organisation in Digital Humanities* (Routledge, 2021). For standards in this vein, see C.M. Sperberg-McQueen and David Dubin, "Data Representation," *Digital Humanities Data Curation* (blog), 2017, https://guide.dhcuration.org/contents/data-representation/.

findable (linked from the front-end website and held in a scholarly repository with a DOI), and that it is reusable (data is in an open file format).[30] A complete list of the data fields for the "content" analysis and the "data" analysis is in the appendix.

The survey instrument includes a data dictionary to help define and describe each of the fields in more detail. The full data set (csv), data dictionary, R code for generating graphs, and Readme file are available in the UO Libraries Dataverse Repository.[31]

**Findings**

While our original primary goal with this survey was to locate and analyze the accessibility of datasets associated with online platforms, in documenting the fields as described above we found a number of macro patterns visible across both front-end and back-end elements.

***Digital Project Website***

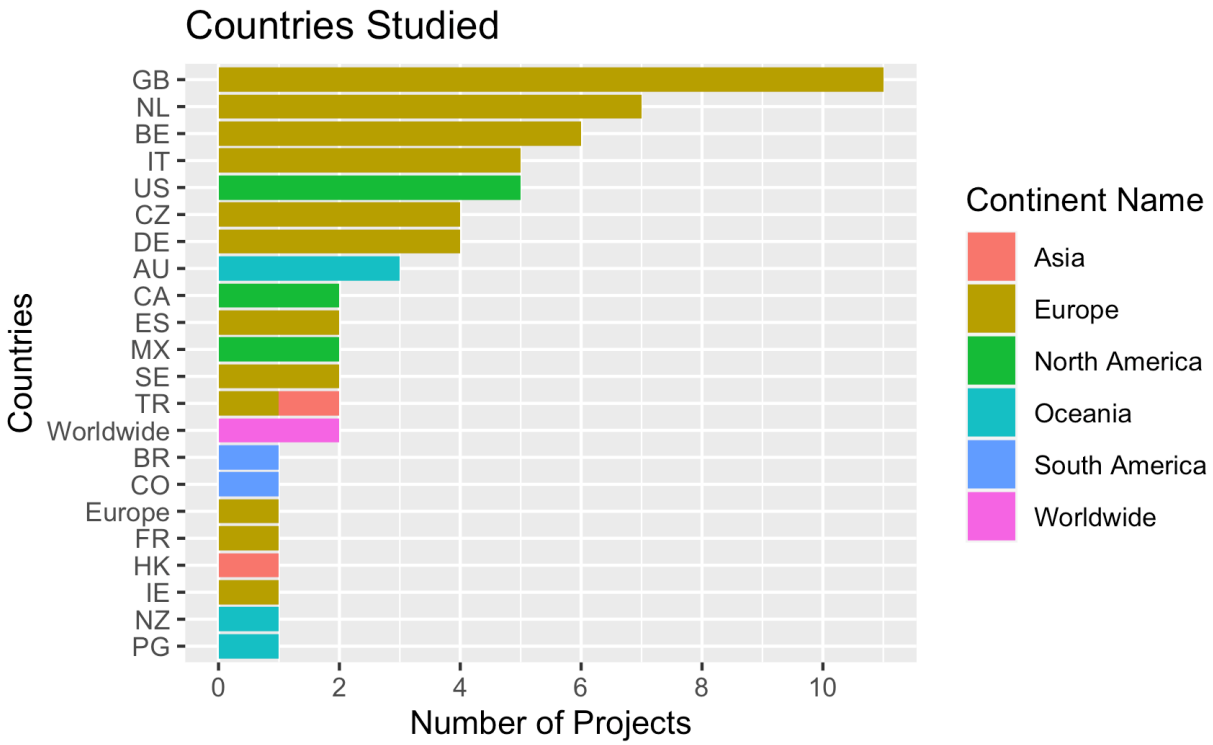
The majority of projects focus on Western Europe, with a smaller number that covered North America and Oceania. The Global South is conspicuously absent, with only one project including any data from Asia or Africa. This lack is probably both a result of our own flawed epistemic perspectives and linguistic capabilities (and limited process of project surveying) and the wider skewed political economy of academic knowledge production, particularly in digital forms, that has historically marginalized analysis of cinematic practices in the Global South including Asia, Africa and Latin America.[32] It is clear that there is a critical need to broaden the spectrum of digitally presented exhibition case studies to better reflect all regions of the world, and the type of publicly-accessible data sharing that we are suggesting here can act as one positive structural element towards that more globalizing goal. (Note that projects surveyed can include more than one country. Count is of each time a country appears in a project--and thus

[30] For a top-level overview of some of the different issues in knowledge organization between institutions and small or individual researchers, see Koraljka Golub, Ahmad M. Kamal, and Johan Vekselius, "Knowledge Organisation for Digital Humanities: An Introduction," in *Information and Knowledge Organisation in Digital Humanities* (Routledge, 2021), 5-6. In the field of cinema studies, we might recognize three categories: the museum or cultural institution, the large, grant-funded research group, and the small group or individual researcher.

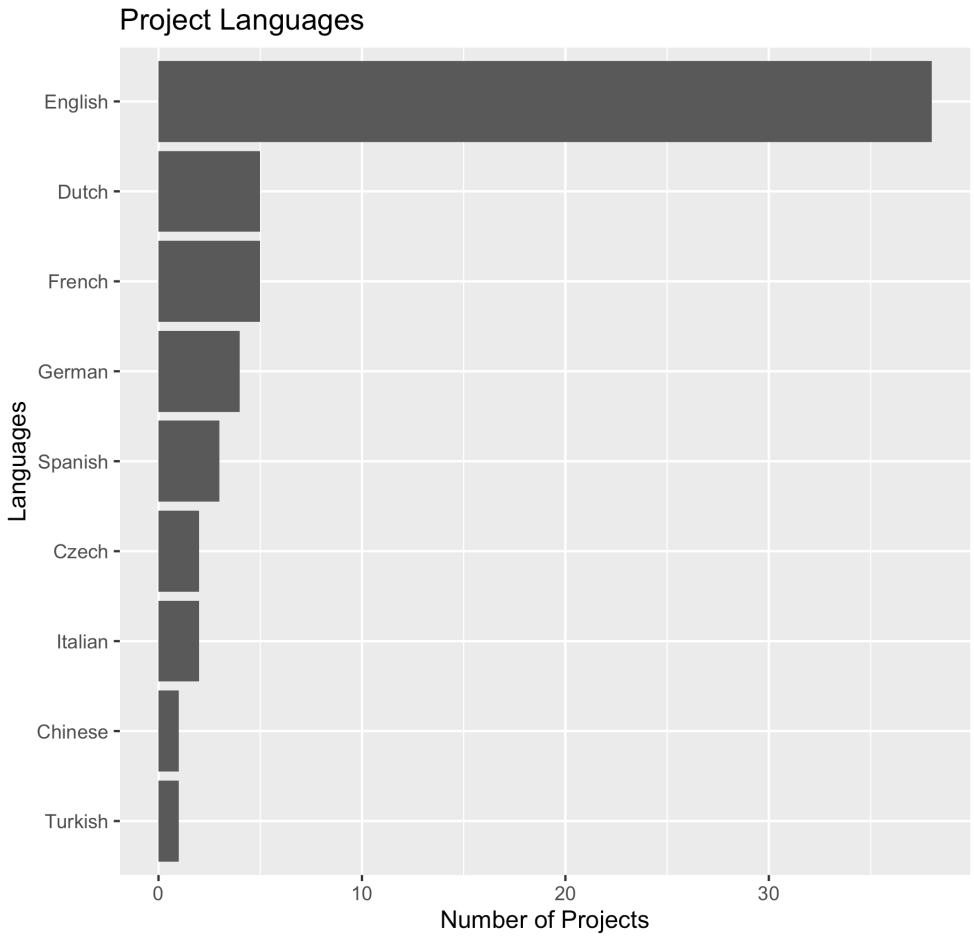
[31] Gabriele Hayden, Elizabeth Peterson, and Michael Aronson, "Replication Data for: Local Cinema History at Scale: Data and Methods for Comparative Exhibition Studies" (Harvard Dataverse), accessed March 14, 2022, https://doi.org/10.7910/DVN/6WOQPO.

[32] Despite their importance for documenting films made in their respectives countries, we chose to exclude the projects Indiancine.ma (https://indiancine.ma/) and Turkishcine.ma (https://turkishcine.ma/) because the sites do not include information about where, how, or when these films were exhibited or how audiences received them--elements that are central to the focus of this exploratory survey.

total adds up to more than the number of projects surveyed.)

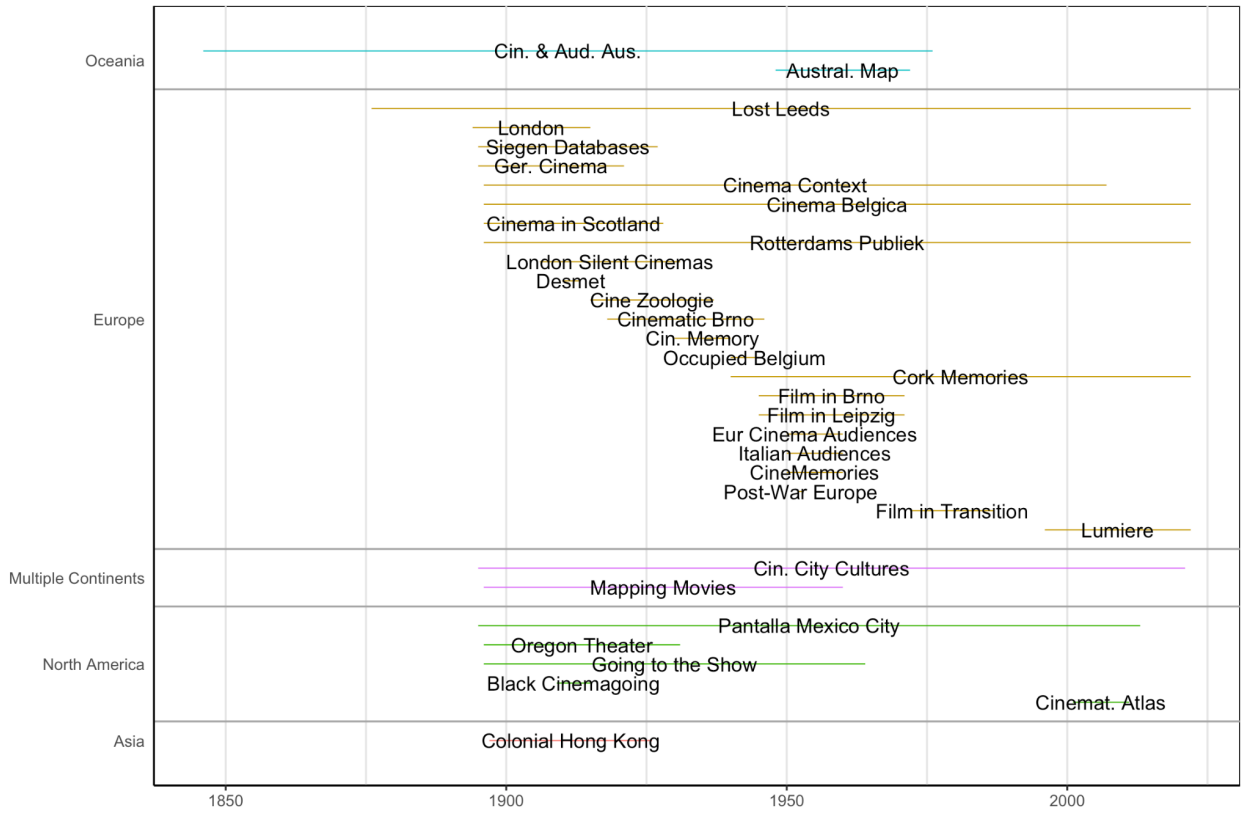



Not surprisingly, the dominance of colonialist languages for project presentation, overwhelmingly English, largely matches this subject focus. (Again, some projects present their information in more than one language. Count is of each time a country appears in a project--and thus total adds up to more than the number of projects surveyed.)




Many projects start with the "birth of cinema," while others are focused on a particular historical moment. One can speculate that the early cinema projects may be benefiting from the growth in digitized primary sources collections from this era, most of which are in the public domain and readily available in public repositories. Projects that focus on a more narrow time period may have more limited sources available, such as oral histories of living people, or proprietary box office data sold by a commercial vendor. A number of projects did not indicate in their documentation the range of time they covered, and so have been omitted from the timeline below. For more information on each project, including exact dates, long titles, links to websites, and projects excluded, see the full survey.[33]

[33] Gabriele Hayden, Elizabeth Peterson, and Michael Aronson, "Replication Data for: Local Cinema History at Scale: Data and Methods for Comparative Exhibition Studies" (Harvard Dataverse), accessed March 14, 2022, https://doi.org/10.7910/DVN/6WOQPO.




Around half of these project websites offered minimal scholarly or narrative framing. Sometimes that narrative was available offline in the form of scholarly articles. The other half of the projects offered either a single narrative introduction, or short contextual narratives describing, for example, individual theaters.

Eighty percent of projects cite sources for some or all of their material, which is in keeping with their (mostly) scholarly origins.[34] A few projects are crowd-sourced, such as the granddaddy of movie theater love, Cinema Treasures,[35] which covers theaters past and present across the US. Other projects are microhistories by local cine-enthusiasts, such as the charmingly illustrated Hiding In Plain Site: The Lost Cinemas of Leeds. The rich relationship, as well as inherent tensions, between scholarly and lay historians is rarely directly acknowledged but a deeply-ingrained aspect of locally-oriented exhibition and moviegoing study.

***Digital Project Data for Reuse***

[34] Projects that reuse data often simply cite to the original data set. This means that the user may need to track down citations in the source data set in order to amass a complete set of citations. These and other examples where sources would require more work to reconstruct were marked "partially."

[35] Cinema Treasures, Cinema Treasures LLC, accessed August 29, 2021, http://cinematreasures.org/.

Our survey shows that the vast majority of the projects we explored--sixty-eight percent--did not openly share their data (or code to re-access the data) in a form that is aggregated, machine readable, and reusable. This came as a surprise to us, especially considering the current discursive enthusiasm in our (sub)field for all things comparative.

Let us be clear, through a specific example, about what we mean by openly sharing data: The exciting new History of Film Exhibition and Reception in Colonial Hong Kong project is openly-searchable, includes extensive metadata, and directly links to the full text sources at Hong Kong Public Libraries for entries in their database.[36] However, because it offers no way for a scholar or amateur enthusiast to access and reuse the entire database in another project, we consider this data not shared. By contrast, the code and data to create the Cinema Context database has been archived in a repository and is licensed for reuse by anyone interested in doing so; we consider that data to be shared.

It is possible that more of these surveyed sites do, in fact, make their data accessible elsewhere but that we were unable to find it. Of the twelve sites that share some or all of their data, three do not include any direct links to their open data; we only discovered that the researchers behind those sites share data through searches in repositories and references in articles.

***Data, When Provided, Available in a Public Repository***

Although it is valuable to share data via links from a project website, without maintenance that website will break over time in ways that disables access.[37] For example, in theory the Cinematographic Atlas project shares its data in open formats via a link directly to the website.[38] However, the data is or was hosted on Geocommons without a stable link (DOI or similar), and GeoCommons appears to be at least partially defunct. As a result, all of the links have rotted, which is a sad end after all the hard work and thought that went into making and sharing this data. Luckily, some of the data was find-able after about thirty minutes of Googling;

---

[36] "A History of Film Exhibition and Reception in Colonial Hong Kong (1897-1925)," Lingnan University (2021), accessed August 31, 2021, https://digital.library.ln.edu.hk/en/projects/flim/intro.

[37] "The Web as scientific platform is full of digital wastelands, caused by the end of research projects." Christine Barats, Valérie Schafer, and Andreas Fickers, "Fading Away... The Challenge of Sustainability in Digital Studies," Digital Humanities Quarterly 014, no. 3 (September 25, 2020).

[38] Daniel Naud, "Canadian Movie Theaters Spatial Distribution," *Cinematographic Atlas* (2012), accessed September 1, 2021, http://atlascine.site44.com/distrib.html.

someone has made it available via GitHub, at least for the moment, and links for it are in our survey.[39]

The failure to archive in a stable repository is widespread. Of the twelve projects that share data, only five do so in an archival repository with a stable link (DOI). GitHub repositories can easily be archived via the EU-backed, open archival repository Zenodo,[40] but none of the projects that share data directly or indirectly via GitHub have done this. The example of the Cinematographic Atlas project shows how fragile the data shared outside of a repository is, and how unlikely it is to continue to be accessible for more than a few years.

***License for Reuse***

Including a license with shared data is essential to allowing data to be reused. Of the thirteen projects that share some or all their data, five do not include a license, and seven include some kind of Creative Commons license. An additional four projects that do not share machine-readable data nevertheless include some kind of license. These include a notice of copyright, Terms of Use legal document, a notice that copyright was retained by the site creators, and a notice that the material shared was under no known copyright restrictions (in reference to the right of the website to display the material).

***File Formats***

Of the twelve projects that share some or all of their data, two projects share data in csv, tsv, or txt formats (comma-separated values, tab-separated values, and plain text). These simple, text-based formats offer, we argue, a gold standard for accessible data sharing of small to medium-sized projects. Character separated text formats such as csv and tsv are not fully standardized–and thus at least one guide to digital humanities data curation aimed at institutional data eschews them in favor of xml formats.[41] However, data management librarians continue to

[39] The GitHub README says, "GeoCommons was a community data sharing site where anyone could upload and attribute data, discover and download in open formats . . . This repository contains a full archive of the public datasets shared at GeoCommons.com." Andrew Turner, "README.Md," *Geoiq/Gc_data* (GitHub repository), February 9, 2022, https://github.com/geoiq/gc_data. While the README author refers to a "repository," GitHub repositories are hosted by a commercial vendor (GitHub, owned by Microsoft) that has no commitment to offering long-term archiving or stable links beyond their current business interests. Cinematographic Atlas-related data is available at http://geocommons.com/datasets?id=67107 and https://github.com/geoiq/gc_data/blob/master/datasets/67107.geojson.

[40] "Making Your Code Citable," GitHub Guides (2016), accessed August 29, 2021, https://guides.github.com/activities/citable-code/.

[41] C.M. Sperberg-McQueen and David Dubin, "Data Representation," *Digital Humanities Data Curation* (blog), 2017, https://guide.dhcuration.org/contents/data-representation/.

recommend their use to individual researchers and small labs because they are widely used and understood across many research domains.[42] They are simple for all users--technical and non-technical alike--to reuse, and are most likely to remain reusable for at least the next few decades. The Cine ZOOlogie project, for example, shares data as csv and tsv files, accompanied by a brief data dictionary in plain text format. The Early Cinema in Scotland project shares data both as an Excel document and a csv document; by making the data available in two formats, they are able to optimize for ease of use (Excel is likely the easiest format for non-technical users in the present day) *and* for longevity (csv).[43] Mapping Desmet[44] shares data only in Microsoft Excel's standard open XML format (xlsx). While the Library of Congress approves this format for archiving, its open specification runs to 6,000 pages, and data management librarians widely recommend character separated text files instead (csv or tab).

Others we surveyed share data in formats that are partially or fully open, but require technical expertise to use. This is not always a criticism–geographic data, for example, simply requires greater technical expertise to use. As noted in a recent report on *FAIR Data Sharing in the Humanities*, appropriate open formats for data reuse depend on the data type, data model, community norms.[45] Two projects we surveyed share geographic data in open formats recommended (shapefile)[46] or accepted (kmz)[47] by the Library of Congress.[48]

---

[42] Willow Dressel, "Research Data Management Instruction for Digital Humanities," Journal of EScience Librarianship 6, no. 2 (December 8, 2017), https://doi.org/10.7191/jeslib.2017.1115.

[43] The Early Cinema in Scotland project shares this data only on their website, making it susceptible to being lost should the website stop being updated. The Cine ZOOlogie project, on the other hand, shares the data in a repository and includes a link from an article on the topic to the repository; however, there is no direct link from the website to the repository, making it difficult for lay users to find and access the data.

[44] Mapping Desmet, accessed August 31, 2021, http://mappingdesmet.humanities.uva.nl/#/.

[45] ALLEA (2020). Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities. Edited by Natalie Harrower, Maciej Maryl, Beat Immenhauser, and Timea Biro. Berlin. doi: 10.7486/DRI.tq582c863

[46] Mapping Movies shares data in a shapefile, which is itself made up of a number of files, some of which can be exported to xml or other open data formats. Jeffrey Klenotic, Mapping Movies, accessed August 30, 2021, https://www.mappingmovies.com/.

[47] Liverpool: City in Film Online shares data in a kmz file, which consists of a file in Google's Keyhole Markup Language format and possibly other files, in one zipped archive. Unzipping gives users a kml file that can be renamed to xml and imported into Excel as an xml data source. "City in Film," The Liverpool School of Architecture, University of Liverpool, accessed August 30, 2021, http://cityinfilm.org/.

[48] As noted in a recent report on archiving geospatial files, the shapefile is an industry standard. Artefactual Systems and the Digital Preservation Coalition, "Preserving GIS" (Digital Preservation Coalition, July 26, 2021), https://doi.org/10.7207/twgn21-16. Because the mapping software company ESRI designed and controls the shapefile's open specification, it is perceived as less "open" than other formats, but because it is a well-documented industry standard, it is nevertheless appropriate for archiving.

| Format | Ease of Use Today | Risk that it won't open in 10 years | Company/Software |
|---|---|---|---|
| txt | easy | low | Open source |
| csv | easy | low | Open source |
| xlsx | easy | medium | Microsoft Excel (open specification) |
| kmz, kml | Somewhat technical | medium | Google (open specification) |
| shapefile (shp) | Somewhat technical | medium | ESRI (open specification) |
| geojson | Very technical | low | Open source |
| RDF/linked open data (ttl, rq) | Very technical | low | Open source |
| mysql | Very technical | low | Oracle (open specification) |

Several large European projects share or plan to share their RDF data via either an API (Cinematic Brno, 1918-1945) or a SPARQL query endpoint. For example, the Cinema Context project shares a MySQL database (via a repository),[49] an RDF version[50] with openly shared code,[51] and a query endpoint that allows users to access the entire database as Linked Open Data via SPARQL queries.[52] While RDF is a best in class open format that can be expected to persist over time, re-using the data shared via either method requires a level of technical expertise not held by many cinema scholars or librarians. Project leaders are aware of these challenges, and working to disseminate this knowledge through workshops, such as the series of workshops held at the 2022 HoMER Network conference. Nevertheless, this data is functionally not accessible to cinema scholars who lack access to major grant funding to hire a data specialist. While this to

[49] Karl Dibbets, Cinema Context. Film in Nederland Vanaf 1896: Een Encyclopedie van de Filmcultuur. DANS (2018), accessed September 2, 2021, https://doi.org/10.17026/dans-z9y-c5g6.
[50] "Cinema Context RDF Documentation," GitLab, accessed September 2, 2021, https://uvacreate.gitlab.io/cinema-context/cinema-context-rdf/.
[51] "Cinema Context RDF," GitLab, accessed September 2, 2021, https://gitlab.com/uvacreate/cinema-context/cinema-context-rdf/.
[52] "SPARQL endpoint for Cinema Context," CREATE, accessed September 2, 2021, https://data.create.humanities.uva.nl/.

some degree reflects the complexities of working with large data sets, we would like to suggest that sharing a snapshot of the data tables in a repository in csv or tsv format would further increase access to the data. For example, our data librarian co-author (Gabriele) worked for several years directly with SQL databases and has studied basic SPARQL. Nevertheless, learning to translate the SPARQL query results or the mysql data dump to something that could be explored in Excel, R, or Python would take substantial work.[53]

***Discoverability***

We found that simply locating many of these projects was a challenge. Although they are all on the web and can be potentially located via a Google search, one must first know the name of the project to search for it. There is no easy way to find all of them, and few link to each other. There does not seem to be a single clearinghouse for digital mapping projects, much less for one specifically about film exhibition history. This may reflect the many documented challenges associated with creating and maintaining digital directories.[54] Nevertheless, finding data sets for the few projects that have them was also difficult. The HoMER Network website is the most likely place to collect links and descriptions of these projects, but in its current form it is best described as partial, and its map provides the location of the researchers themselves rather than the site of the history they're describing. Projects we stumbled on in the initial information gathering stage would be much easier to find with a profile on the HoMER site, such as the Australian Cinemas Map or Black Cinemagoing: Black Women and Cinemas in Harlem.[55] Judgment falls on us on this topic as well in terms of limiting the visibility of our project; adding the Oregon Theater Project to the HoMER map is still on our to-do list. Our recommendations for increasing discoverability for both individual and comparative projects can be found in the final section of the paper.

***Persistence***

Going To The Show is gone. Or rather, like a number of projects in our survey, it is still online but no longer updated, and, more significantly, its maps, the project's core feature, appear

---

[53] Analysis in R, Python, or Excel would require data in csv/tsv format. This could be achieved either by writing code that would "translate" the existing mysql code or RDF triples to a csv format, or by importing the mysql code into a new database and then exporting the data as a csv. None of this is a simple proposition.

[54] See, for example, Quinn Dombrowski, "The Directory Paradox," in *Debates in Digital Humanities: Institutions, Infrastructures at the Interstices*, eds. Anne McGrail, et al. (Minneapolis: University of Minnesota Press, forthcoming).

[55] Agatha Frymus, "Black Women & Cinemas in Harlem," *Black Cinemagoing* (2020), accessed August 31, 2021, https://blackcinemagoing.wordpress.com/.

to no longer exist. Going To The Show, Mapping Moviegoing in North Carolina[56] was begun in 2006 led by Robert C. Allen, a crucial pioneer of empirically-driven film history, and is often cited as greatly influencing the field's move towards geospatially-oriented film studies. So while its impact lives on, the data that drove it remains inaccessible. Going To The Show is by no means alone in this loss, and the reasons for this can be easy enough to identify: people get new jobs, retire, die, they lose their funding, take on new research interests, or the underlying technology is no longer supported or freely available.[57] In the case of Going To The Show, changes to Google's proprietary mapping API, including a shift from a free to pay per views model, likely hastened the demise of its geographic presentation.

Presumably, in many cases, the underlying data used to create the websites and underlying databases still exists, so the data could be recovered, stored in a repository and shared, even if the website itself is mostly moribund.

**Film Scholars Heal Thyselves: Recommendations and Limitations**

While our intentions from the beginning of the project's development were to build a site that was open-source in both interface and database, our lack of extensive IT knowledge required our reliance on a University systems and programming team to build the digital infrastructure on the front and back end. That team was supportive of these goals, but their primary task was to build *us* a working platform. So, although we were lucky to have been awarded an internal grant of approximately $15,000 for the project build, we ultimately ran low on both funds and time as the project neared its initial public launch, and so were forced to prioritize the human-facing and student-focused elements of the website. As the platform was built as an empty framework for our students to fill in with their own research on Oregon's theaters, there was no initial dataset to share with potentially-interested outside scholars. That has changed as students have begun to research and publish their work on the site over multiple iterations of the undergraduate course, producing a growing dataset that was, until now, not even accessible to us, since no way to export the data was configured in our system. From our survey, it appears that many existing local exhibition platforms do not have the resources to openly share their data.

---

[56] "Going to the Show: Mapping Moviegoing in North Carolina," *Documenting the American South*, University of North Carolina at Chapel Hill (2008), accessed August 31, 2021, http://gtts.oasis.unc.edu/.

[57] Christine Barats, Valérie Schafer, and Andreas Fickers, "Fading Away... The Challenge of Sustainability in Digital Studies," *Digital Humanities Quarterly* 014, no. 3 (September 25, 2020).

Lack of technical know-how and funding are real barriers for many of us providing this kind of public-facing digital scholarship, but we should also acknowledge that many of us in the humanities and social-science fields are taught to fear the incomplete and imperfect, especially when publishing. The result, we believe, is a too-often desire to wait until the dataset is "complete" before providing broad access, a desire that is unrealistic, and even historiographically problematic, but which remains unfortunately durable within the field. For us to overcome this hurdle, we were determined in writing this article to follow (at least) a "good-enough" set of practices, and to do it in a prescriptive and transparent manner that could be replicated by many if not all of the existing active platforms which currently do not publicly share their data from their primary websites.

Our first step was to engage someone at our University with the background and knowledge to help facilitate the process, in our case Gabriele Hayden, the University's Research Data Management and Reproducibility Librarian, co-author of this article. If I'm honest (Michael, i.e., Film Historian), I had no idea that such a person existed at our mid-tier state institution, but increasingly such expertise is available to many scholars across higher education academia. Although in our instance Gabriele became an active collaborator on the larger project that our survey and this publication entails, in the case of adding data accessibility to the Oregon Theater Project, we decided to role-play our jobs and responsibilities as a data librarian and an inexperienced, if eager, digital scholar. As a first step we provided her with access to the back end of the Oregon Theater Project platform, which, in our situation, was designed to collect several content types into a Drupal database as students entered information about theaters. Our developers configured a query that produces a partial dataset that can be downloaded as a csv file (comma-separated-values) that includes venues, number of seats, locations, dates of operation, and names of owners. We learned that csv is a highly reusable machine-readable open format common across all computer platforms and well-suited for the types of tabular datasets these exhibition platforms are most likely to produce for comparative use by scholars. Csv is the "lowest common denominator" for open data, and we suggest that all exhibition platforms should (at least) make their data available in this format.

In our initial "meeting," Gabriele walked us through the existing data fields and helped us find and clean up any inconsistencies in the organization or content of the dataset. In our case, reformatting for the international date standard across the set was necessary before downloading

a csv file ready for initial deposit. Extracting a snapshot of the existing data is critical for issues of both archive and access, and much easier to maintain from a technological and labor standpoint. Unfortunately, our current data export process does not allow us to archive crucial elements of our project, in particular, digital images of newspaper documentation. In our process, all of our data is entered directly in the database and not held elsewhere. This means that should institutional support for this project cease, or when the technology upon which it is built becomes so out of date that an expensive migration is necessary, we are in danger of the connection between images on the site and metadata about (descriptions of) those images.[58] We are working with our developers on creating a series of views that will allow for more data to be extracted from the system.

Once we have data, where to put it? For data to be managed, and made accessible in a continuous and sustained way, it should be deposited in a location that ensures its long-term stewardship. Researchers, both creators and users, need to be assured that data sets are retrievable, annotated sufficiently to understand the context of their creation, and assigned license information that specifies the conditions of reuse. As we've seen with some of the platforms in the survey, digital data is fragile, file formats or software/hardware can become obsolete, and/or websites that hold the data can become inaccessible when links break or pages move. Scholars are equally fragile and over time action or inaction can render the data unobtainable or corrupt. "Trustworthy Digital Repositories" are the recommended preservation solution to these issues and while we do not endorse specific repositories, we would suggest utilizing those created and/or supported by well-funded non-profit institutions such as Harvard's Dataverse, Zenodo[59] (CERN, EU) or national institutional centers such as DANS which is maintained with funding from the Dutch government.[60] In our case, we employed Harvard's Dataverse, as it has been adopted for academic use by our own institution.[61] Although each

[58] To be clear, it is possible to archive the "front end" of a digital project using the Internet Archive, or tools such as Webrecorder (https://webrecorder.net/), which, according to its website "provides a suite of open source projects and tools to capture interactive websites and replay them at a later time as accurately as possible." While this is immensely valuable, it does not retain the data that created the website in a format that allows for easy re-use. Instead, each piece of written text would need to be extracted via OCR or hand-transcription and manually entered into a database--a laborious project.

[59] Zenodo, CERN Data Centre & Invenio, accessed August 30, 2021, https://zenodo.org/.

[60] "Dutch National Centre of Expertise and Repository for Research Data," Data Archiving and Networked Services DANS, accessed August 29, 2021, https://dans.knaw.nl/en.

[61] Michael Aronson and Elizabeth Peterson, "Oregon Theater Project Database" (Harvard Dataverse, March 14, 2022), https://doi.org/10.7910/DVN/FGOUZ3.

repository has a different interface, creating a user account, adding and uploading the datasets are straight-forward and do not require technical expertise.

While adding our dataset to a repository was an important first step for accessibility, accessibility alone does not render the data truly usable. For it to be reusable, data must be accompanied by metadata, "data about data," to communicate the content of the set, the purposes under which it was created, and the ways in which it can be reused. While the repository form required to submit your data involves the inclusion of citation metadata and descriptive keywords, our data librarian also suggested that we complete and upload alongside our dataset a readme style metadata sheet[62] with additional types of authorial, technical and administrative metadata. One of the most important pieces of metadata that was created when we added our dataset to the Dataverse is a persistent identifier (PID), a globally unique identifier that creates a stable (persistent) link for objects, in this case our dataset. In our case, the Harvard Dataverse provides a Digital Object Identifier (DOI) at no cost that is now associated with our Oregon dataset. Additionally at the recommendation of our librarian, we added our own ORCiDs[63] (unique author/researcher identifiers) to the metadata. Lastly, but not leastly, we want researchers to reuse our data. To avoid any legal ambiguity that could limit this usage we employed the Creative Commons (CC) "choose a license" system to include a license with as few (non-commercial) restrictions for reuse as possible.[64]

While we learned a lot in the process of walking through the steps with our data librarian, the amount of time and labor required to move from a csv to providing an accessible and persistent dataset for global researchers to deploy was brief--hours not days--and straightforward.

Our recommendations are designed to be neither radical nor disproportionately difficult. If you already have, or plan to build, a digital platform focused on exhibition and moviegoing do these four simple things:

1. Share data early and often. We are not attempting to define "early and often," as each project's chronological trajectory will be different, but it is always better to share partial data than no data at all. For our own Oregon Theater Project, for instance, we are now

[62] "AUTHOR_DATASET_ReadmeTemplate.txt," Cornell University, accessed September 1, 2021, https://cornell.app.box.com/v/ReadmeTemplate.
[63] ORCid, accessed August 30, 2021, https://orcid.org/.
[64] "License Chooser," Creative Commons, accessed September 2, 2021, https://chooser-beta.creativecommons.org/.

committed to uploading a revised dataset immediately after each iteration of our undergraduate course.

2. Share in open formats. As stated above, we are recommending that every site, at a minimum, provide their dataset in csv format. While there are certainly formats with higher degrees of functionality for complex data manipulation and visualization, such as kmz or linked data (RDF), csv offers the easiest road to global data access in the field.
3. Share in a repository with a DOI. Websites break, people get old and tired, DOI's don't. But also, include links to the data with any necessary explanatory context on your website and in any articles you write about the project. This will make it easier for others to find your work!
4. Share with a license that says what people can do with it. Within the obvious and necessary limits of copyright and IP broadly share what you've found and learned.

A great deal of ink has been spilled regarding why scholars don't regularly share their data (it can be a lot of work), wondering about the value of sharing data, and who will use data once it is shared. One recent study found that scholars who shared data along with their publications had up to 25.36% higher average citation impact.[65] But this sounds abstract, and there is the question of whether scholars who share their data also simply have more resources or are better organized, or something else such that data sharing is simply correlated with higher publication rates. What is so exciting about the field of local cinema studies, from the perspective of a data librarian (Gabriele) is that the value of sharing this data is very evident to everyone in the field. The data is small enough to be managed with relatively little technical expertise, and yet its aggregation would clearly add value to the field. This is the premise behind the productive work we're beginning to see in comparative case studies like those presented in the "Comparative Histories of Moviegoing" edition of the *Journal For Media History*. At the same time, as our survey clearly shows, there remains widespread practical barriers to broader access and implementation of significant amounts of movie culture data for reuse, experimentation and presentation. Along with others, Daniel Bilteryest and Phillippe Meers have

[65] Giovanni Colavizza, Iain Hrynaszkiewicz, Isla Staden, Kirstie Whitaker, and Barbara McGillivray, "The Citation Advantage of Linking Publications to Research Data," *PloS one* 15, no. 4 (2020). Another study finds a lower, but still significant citation advantage of 9%: Heather A. Piwowar, and Todd J. Vision. "Data Reuse and the Open Data Citation Advantage," *PeerJ* 1 (2013).

eloquently called for alternatives to the kinds of systematic comparison methodologies of "classical" data comparison that we are suggesting here, and we are wholly supportive of a widely spacious and malleable approach to what constitutes comparative analysis of cinema history and culture.[66] Ultimately, however, our survey shows that there is a strong disconnect between current broadly stated field goals and what is actually being disseminated through the online platforms that collectively provide the most visible representation of the history of movie exhibition and moviegoing. To that end, we'd like to offer one final recommendation, that the HoMER network website return to its original online project-oriented roots, as described by Jeffrey Klenotic in his historiographic mapping of New Cinema History, to serve as a geographic directory of data-accessible movie culture platforms "as well as a place where users could gather to explore and remix data, and make serendipitous discoveries."[67]

There are continuing debates and tensions over the place and power of comparative work in our field, and how best to meaningfully produce it across and among a highly divergent global group of individuals and institutions. However, starting and maintaining these relatively simple processes for data retention and accessibility, alongside an online directory of projects and their data, can offer us a collective strong foundation from which comparative work may evolve as the field itself matures and becomes as richly complex as the individual microhistories from which it is derived.

[66] Biltereyst and Meers, "New Cinema History and the Comparative Mode: Reflections on Comparing Historical Cinema Cultures."
[67] Jeffrey Klenotic, "Mapping Flat, Deep, and Slow: On the 'Spirit of Place' in New Cinema History," *TMG Journal for Media History* 23, no. 1-2 (2020), 1-34.

Bibliography


Acland, Charles and Eric Hoyt, eds. *The Arclight Guidebook to Media History and the Digital Humanities* (Reframe Books, 2016).

ALLEA (2020). Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities. Edited by Natalie Harrower, Maciej Maryl, Beat Immenhauser, and Timea Biro. Berlin. doi: 10.7486/DRI.tq582c863.

Aronson, Michael and Elizabeth Peterson, "Oregon Theater Project Database" (Harvard Dataverse, March 14, 2022), https://doi.org/10.7910/DVN/FGOUZ3.

Artefactual Systems and the Digital Preservation Coalition, "Preserving GIS" (Digital Preservation Coalition, July 26, 2021), https://doi.org/10.7207/twgn21-16.

Australian Cinemas Map: A Map of Film Weekly Motion Picture Directory Cinema Data, 1948-1971, Flinders Institute for Research in the Humanities (2011), accessed August 31, 2021, https://auscinemas.flinders.edu.au/.

"AUTHOR_DATASET_ReadmeTemplate.txt," Cornell University, accessed August 30, 2021, https://cornell.app.box.com/v/ReadmeTemplate.

Baptist, Vincent, Julia Noordegraaf, and Thunnis Van Oort, "A Digital Toolkit to Detect Cinema Audiences of the Silent Era: Scalable Perspectives on Film Exhibition and Consumption in Amsterdam Neighbourhoods (1907-1928)," *Studies in European Cinema* 18, no. 3 (July 11, 2021): 252–273. doi:10.1080/17411548.2021.1930974.

Barats, Christine, Valérie Schafer, and Andreas Fickers, "Fading Away... The Challenge of Sustainability in Digital Studies," *Digital Humanities Quarterly* 014, no. 3 (September 25, 2020).

Biltereyst, Daniel, Richard Maltby, and Philippe Meers, Philippe, eds. *The Routledge Companion to New Cinema History* (New York: Routledge, 2019).

Biltereyst, Daniel and Philippe Meers. "New Cinema History and the Comparative Mode: Reflections on Comparing Historical Cinema Cultures," *Alphaville* 11 (2016), 13–32.

CineRicordi, accessed August 30, 2021, https://www.cinericordi.it/

Cine ZOOlogie, accessed August 30, 2021, http://www.cinemazoologie.be/.

Cinema City Cultures, accessed August 31, 2021, http://cinemacitycultures.org/.

"Cinema Context RDF,"GitLab, accessed September 2, 2021, https://gitlab.com/uvacreate/cinema-context/cinema-context-rdf/.

"Cinema Context RDF Documentation," GitLab, accessed September 2, 2021, https://uvacreate.gitlab.io/cinema-context/cinema-context-rdf/.

Cinema Context: Film in the Netherlands from 1896, accessed August 29, 2021, https://cinemacontext.nl/.

Cinema Treasures, Cinema Treasures LLC, accessed August 29, 2021, http://cinematreasures.org/.

Cinematic Brno, Department of Film and Audiovisual Culture at Faculty of Philosophy, Masaryk University, Brno (2011), accessed March 10, 2021, https://cinematicbrno.phil.muni.cz/ .

"City in Film," The Liverpool School of Architecture, University of Liverpool, accessed August 30, 2021, http://cityinfilm.org/.

Colavizza, Giovanni, Iain Hrynaszkiewicz, Isla Staden, Kirstie Whitaker, and Barbara McGillivray. "The Citation Advantage of Linking Publications to Research Data," *PloS one* 15, no. 4 (2020).

Dacos, Marin, "Manifeste des Digital humanities," Billet, *THATCamp Paris* (blog), March 26, 2011, https://tcp.hypotheses.org/318.

"DH Cinema Projects," Transformations I Cinema and Media Studies Research Meets Digital Humanities Tools, accessed August 29, 2021, https://transformationsconference.net/dh-cinema-projects/.

Dibbets, Karl, Cinema Context. Film in Nederland Vanaf 1896: Een Encyclopedie van de Filmcultuur. DANS (2018), accessed September 2, 2021, https://doi.org/10.17026/dans-z9y-c5g6.

Dombrowski, Quinn, "The Directory Paradox," in *Debates in Digital Humanities: Institutions, Infrastructures at the Interstices*, eds. Anne McGrail, et al. (Minneapolis: University of Minnesota Press, forthcoming).

Dressel, Willow, "Research Data Management Instruction for Digital Humanities," *Journal of EScience Librarianship* 6, no. 2 (December 8, 2017), https://doi.org/10.7191/jeslib.2017.1115.

Drucker, Joanna and Patrik BO Svensson, "The Why and How of Middleware," *Digital Humanities Quarterly,* 010, no. 2 (May 19, 2016), paragraph 60.

"Dutch National Centre of Expertise and Repository for Research Data," Data Archiving and Networked Services DANS, accessed September 1, 2021, https://dans.knaw.nl/en.

The Early Cinema in Scotland Research Project, accessed August 29, 2021, https://earlycinema.gla.ac.uk/.

"FAIR Principles," GOFAIR, accessed August 30, 2021, https://www.go-fair.org/fair-principles.

Frymus, Agatha. "Black Women & Cinemas in Harlem," Black Cinemagoing (2020), accessed August 31, 2021, https://blackcinemagoing.wordpress.com/.

"Going to the Show: Mapping Moviegoing in North Carolina," Documenting the American South, University of North Carolina at Chapel Hill (2008), accessed August 31, 2021, http://gtts.oasis.unc.edu/.

Golub, Koraljka, Ahmad M. Kamal, and Johan Vekselius, "Knowledge Organisation for Digital Humanities: An Introduction," in *Information and Knowledge Organisation in Digital Humanities* (Routledge, 2021), 5-6.

Hayden, Gabriele, Elizabeth Peterson, and Michael Aronson. "Replication Data for: Local Cinema History at Scale: Data and Methods for Comparative Exhibition Studies," *Harvard Dataverse*, The President & Fellows of Harvard College (2022), accessed March 14, 2022. https://doi.org/10.7910/DVN/6WOQPO.

"A History of Film Exhibition and Reception in Colonial Hong Kong (1897-1925)," Lingnan University (2021), accessed August 29, 2021, https://digital.library.ln.edu.hk/en/projects/flim/intro.

"HoMER Projects," HoMER Network, accessed August 29, 2021, https://homernetwork.org/homer-projects.

"ISO 3166 Country Codes," International Organization for Standardization, accessed August 29, 2021, https://www.iso.org/iso-3166-country-codes.html.

Klenotic, Jeffrey. "Mapping Flat, Deep, and Slow: On the 'Spirit of Place' in New Cinema History." *TMG Journal for Media History* 23, no. 1-2 (2020), 1-34.

Klenotic, Jeffrey, Mapping Movies, accessed August 28, 2021, https://www.mappingmovies.com/.

Kräutli, Florian, Esther Chen, and Matteo Valleriani, "Linked Data Strategies for Conserving Digital Research Outputs: The Shelf Life of Digital Humanities," in *Information and Knowledge Organisation in Digital Humanities* (Routledge, 2021).

"License Chooser," Creative Commons, accessed September 2, 2021, https://chooser-beta.creativecommons.org/.

"Listed Cinemas," Cinema Theatre Association (2021), accessed August 29, 2021, https://cinema-theatre.org.uk/uk-cinemas/listed-cinemas/.

Maltby, Richard. "New Cinema Histories," in *Explorations in New Cinema History: Approaches and Case Studies,* eds. Richard Maltby, Daniel Biltereyst and Phillippe Meers (Chichester, UK: Wiley-Blackwell, 2011), 1-40.

"Making Your Code Citable," GitHub Guides (2016), accessed August 30, 2021, https://guides.github.com/activities/citable-code/.

Mapping Desmet, accessed August 31, 2021, http://mappingdesmet.humanities.uva.nl/#/.

Naud, Daniel. "Canadian Movie Theaters Spatial Distribution," *Cinematographic Atlas* (2012), accessed August 29, 2021, http://atlascine.site44.com/distrib.html.

ORCid, accessed August 29, 2021, https://orcid.org/.

Oregon Theater Project, accessed September 2, 2021, http://oregontheaterproject.uoregon.edu.

Pafort-Overduin, Clara, John Sedgwick, and Lies Van de Vijver, "Identifying Cinema Cultures and Audience Preferences: A Comparative Analysis of Audience Choice and Popularity in Three Medium-sized Northern European Cities in the Mid-1930s," *TMG Journal for Media History* 21, no. 1 (2018), 102-118.

Piwowar, Heather A., and Todd J. Vision. "Data Reuse and the Open Data Citation Advantage," *PeerJ* 1 (2013).

"Projects," European Association for Digital Humanities, accessed August 29, 2021, https://eadh.org/projects.

Regester, Charlene, "From the Buzzard's Roost: Black Movie-Going in Durham and Other North Carolina Cities During the Early Period of American Cinema," *Film History* 17, no. 1 (2005), 113-24.

"Sources," Hiding in Plain Sight: Lost Cinemas of Leeds, accessed August 30, 2021, https://lostcinemas.co.uk/sources/.

"SPARQL endpoint for Cinema Context," CREATE, accessed September 2, 2021, https://data.create.humanities.uva.nl/.

Sperberg-McQueen, C.M. and David Dubin, "Data Representation," *Digital Humanities Data Curation* (blog), 2017, https://guide.dhcuration.org/contents/data-representation/.

Stewart, Jacqueline Najuma, *Migrating to the Movies: Cinema and Black Urban Modernity*, (Berkeley: University of California Press, 2005).

*TMG Journal for Media History* 23, no. 1-2 (2020).

Turner, Andrew, "README.Md," *Geoiq/Gc_data* (GitHub repository), February 9, 2022, https://github.com/geoiq/gc_data.

van Oort, Thunnis. "Industrial Organization of Film Exhibitors in the Low Countries: Comparing the Netherlands and Belgium, 1945–1960," *Historical Journal of Film, Radio and Television* 37, no. 3 (2016), 475-498.

van Oort, Thunnis, and Jessica Leonora Whitehead. 2020. "Common Ground: Comparative Histories of Cinema Audiences," *TMG Journal for Media History* 23, no. 1-2 (2020), 1-11.

Zenodo, CERN Data Centre & Invenio, accessed August 29, 2021, https://zenodo.org/.

Appendix A

Film Exhibition Digital Projects Surveyed

| | |
|---|---|
| Cinema Context | https://cinemacontext.nl/ |
| Cinematic Brno, 1918-1945 | CinematicBrno.phil.muni.cz |
| Mapping Movies | https://www.mappingmovies.com/ |
| European Cinema Audiences | https://www.europeancinemaaudiences.org/ |
| Black Cinemagoing: Black Women and Cinemas in Harlem | https://blackcinemagoing.wordpress.com/ |
| Siegen Cinema Databases | http://fk615.221b.de/siegen/start/show/ |
| The London Project | http://londonfilm.bbk.ac.uk/ |
| Italian Cinema Audiences | https://italiancinemaaudiences.org/ |
| Cinema Belgica | https://www.cinemabelgica.be/ |
| The Oregon Theater Project | http://otp.uoregon.edu |
| The German Early Cinema Database | http://earlycinema.dch.phil-fak.uni-koeln.de/ |
| Early Cinema in Scotland, 1896-1927 | https://earlycinema.gla.ac.uk/ |
| London's Silent Cinemas | http://www.londonssilentcinemas.com/ |
| Lumiere: Database on Admission of Films Released in Europe | https://lumiere.obs.coe.int/web/search/ |
| Film Culture in Brno, 1945-1970: The History of Distribution, Exhibition and Reception | https://www.phil.muni.cz/dedur/?&lang=1 |
| Film Culture Leipzig: 1945-1970 | https://www.phil.muni.cz/leipzigcinema/ |
| Going to the Show: Mapping Moviegoing in North Carolina | http://gtts.oasis.unc.edu/ |
| Australian Cinemas Map | https://auscinemas.flinders.edu.au/ |
| Liverpool: City in Film Online | http://cityinfilm.org/ |
| Cinema City Cultures | http://cinemacitycultures.org/ |
| Cinematographic Atlas: Cybercartographic Atlas of Canadien Cinema | http://atlascine.site44.com/projet1.html |
| CineMemories | https://www.cinericordi.it/ |
| Cinema Memory and the Digital Archive | https://www.lancaster.ac.uk/fass/projects/cmda/ |
| Cine Zoologie: 1915–1936: A film programming database | http://www.cinemazoologie.be/ |

| | |
|---|---|
| Cinema Treasures | http://cinematreasures.org/ |
| Prominent Itinerant Cinema Shows, 1896-1908 (Paul S. Moore) | https://psmoore.ca/prominent-itinerant-cinema/ |
| Mapping Film Programming across Post-War Europe (1952) | https://www.dansdatajournal.nl/rdp/showcases1/oort2020b.html |
| Cinema in Occupied Belgium | https://www.cinema-in-occupied-belgium.be/ |
| Hiding in Plain Sight: Discovering the Lost Cinemas of Leeds | https://lostcinemas.co.uk |
| Film Culture in Transition [1972-1986] | https://www.ucm.es/filmcultureintransition-madrid |
| Cinema Theatre Association: Listed Cinemas | https://cinema-theatre.org.uk/uk-cinemas/listed-cinemas/england/ |
| Scottish Cinemas and Theatres | http://www.scottishcinemas.org.uk/index.html |
| Cinema and Theatre Historical Society Association of Australia | http://www.caths.org.au/ |
| Mapping Desmet | http://mappingdesmet.humanities.uva.nl/#/ |
| UK Cinemas | http://www.ukcinemas.org.uk/ |
| Kinomatics | https://kinomatics.com/ |
| Cultura de la Pantalla in Mexico City (1895-2012) | http://cinemacitycultures.org/mexico-city/ |
| Cork Movie Memories | http://corkmoviememories.com/ |
| Rotterdams Publiek | https://rotterdamspubliek.nl |
| A History of Film Exhibition and Reception in Colonial Hong Kong | https://digital.library.ln.edu.hk/en/projects/flim/intro |
| Cinema and Audience Research Project | https://caarp.edu.au/ |

Appendix B

Survey Data Fields

Below is a complete list of the front-end “content” data fields in the survey:

- Unique ID
- Project Title
- URL
- Project Type
- Date started

- Last update
- Region(s)
- Countries
- Time coverage
- Sources cited
- Venues
- Capacity
- Programming
- Ownership/management
- Geospatial
- Demographic information about audiences
- Languages of presentation
- Presentation of content

Below is a complete list of the back-end “data” elements of projects in the survey:

- Unique ID
- Project Title
- URL
- Front-end interface
- Back-end data access
- Back-end link
- Persistent identifier
- Geospatial content
- License for reuse
- File format for data
- Open file format
- Readme provided
- Persistent identifier
- Date in repository