Oregon Theater Project: A Dataset of Oregon Cinemas from the Silent Era

MICHAEL ARONSON (iD)

ELIZABETH PETERSON (iD)

GABRIELE HAYDEN (iD)

*Author affiliations can be found in the back matter of this article

ABSTRACT

The *Oregon Theater Project* (OTP) dataset is part of an ongoing collaborative research project by undergraduate students enrolled in successive iterations of "Exhibition & Audiences," a Cinema Studies course at the University of Oregon. It will be updated with additional data each time the course is taught. The data set comprises geo/historical data about movie theaters (cinemas) and exhibition in the state from approximately 1894 to 1929. The data is presented on a public website (https://oregontheaterproject.uoregon.edu/) which includes maps and individual theater profiles produced by the students. All profiles, and the underlying data, are reviewed by the course instructors and edited as needed for clarification or accuracy. Profiles include, where available, the theater name, address, city, state, latitude, longitude, number of seats, owner/manager names, and a narrative description. The underlying data, shared as Excel documents and tab-delimited spreadsheets, invites historical comparative analysis of film exhibition practices across time and locale, both local and global.

CORRESPONDING AUTHOR:

Dr. Michael Aronson

Cinema Studies, University of Oregon, Eugene, OR, USA

aronson@uoregon.edu

(1) OVERVIEW

REPOSITORY LOCATION

Harvard Dataverse: https://doi.org/10.7910/DVN/FGOUZ3

Front end interface: https://oregontheaterproject.uoregon.edu/

Context

The *Oregon Theater Project* (OTP) is one of an increasing number of digital projects documenting and sharing the history of movie theaters (cinemas), film programming, and film reception. Most of these projects do not make their data publicly available in a usable format, even though the value of these data projects is greatly increased if they allow data to be aggregated (Aronson et al., 2022a). This data paper contributes to building open data in regional cinema history; it describes the preliminary version of a data set that will be updated regularly.

The *Oregon Theater Project* (OTP) is a collaboration between faculty in Cinema Studies and the University of Oregon Libraries, with a goal of integrating information literacy skills and concepts, as well as digital humanities tools, into the historical research course "Exhibition & Audiences". Students, guided by faculty mentors, come away from this course with a broad knowledge of film exhibition theory and history, along with a firm grasp of research methods. Students learn how to identify appropriate sources for their information need; to select appropriate research tools from a variety of options; to search efficiently within online databases and digital collections, as well as traditional print-based media; to evaluate sources for credibility and authority; to analyse and interpret primary sources; to use information ethically; to cite their sources appropriately; and to publish their finished work online using a selection of digital humanities presentation tools. Each time the course is taught, students build on and improve the research conducted by students in previous years. A new, improved data set based on this work will be published following each course iteration.

(2) METHOD

In the OTP, undergraduate students learn cinema studies research methods within a context of film exhibition history and audiences course content. Students conduct original research in primary sources to compile data and to compose short narratives about Oregon movie theaters during the period of study (1894–1929). Primary sources include newspapers, industry trade journals, city and county directories, business directories, maps, and photographs. Students in the course use a shared Google Drive with a hierarchical folder and file system to manage their research materials.

STEPS

Students enter data directly into a structured website platform built on a Drupal content management system. Figure 1 shows a screenshot of part of the page students use to enter information about a new theatre. Data is updated directly in the platform every time a class is taught. The Drupal database includes images taken from newspapers that are the source of most of the information contained in the database. These images are taken informally as screen shots and published on our website under "fair use" terms.

Because we do not have copyright documentation or permissions for each image, we are not including the images as part of this data set. However, we include several data columns that reference these files to create more contextual information. First, we include a column, 'works_cited', that offers unstructured text citations to sources. Second, we include both plain text and full html versions of text from the website (column names are 'body', 'body_html'; 'additional_facts', 'additional_facts_html'; 'works_cited'; 'works_cited_html'). The html versions include relative links to images as they are embedded in the text. Finally, we include a variable that lists image file names for images highlighted in a special section on the page ('gallery_images'). In theory, this should allow users to create links back to the images for the lifetime of the website.

Figure 1 A partial screenshot of the Drupal form for entering information about theaters in the Oregon Theater Project website.

QUALITY CONTROL

The course instructors serve as editors for the course data and content. They review every entry for accuracy, citations, and correct formatting. Students follow a file-naming convention that embeds source citation information within file names to ensure proper attribution during data entry and writing. This method also allows the course instructors to easily consult the research materials to verify facts as presented in the theater data and narratives. After the class is finished, the course instructors remediate any data entry errors that affect data completeness (such as missing geospatial coordinates) in the Drupal database. However, because when we began this project proofreading was focused on the human-readable website and not on creating machine-readable data, we have not systematically corrected differences in formatting in string variables such as addresses. Missing data may be blank or listed as 'unknown' or 'Unknown' and there may be extra spaces, periods, or other irregularities. We hope in future versions to remediate these issues.

Data is exported as a csv file from several SQL views in the Drupal database, cleaned using an R script, and saved as new spreadsheets. As documented in the Readme file and the R script included with the data set, we trim white space from some columns, split out some variables, and join several spreadsheets to create final versions we think may be most useful to future users. Blanks have been left as they are rather than converted to NAs. To make this data widely accessible, we share results in tab-delimited form and as Excel files; we also share the original files downloaded from Drupal and the R script used to process them. In future versions of this data set, we hope to also include links to theater urls in the front-end database and shapefiles corresponding to theater locations.

DATA STRUCTURE

While the data readme will include complete, up-to-date documentation of data variables as the data set grows and evolves, here we highlight import elements of the processed data that we expect will remain stable over time. The tabular data contained in the files 'theaters_[date]. tab' and 'theaters_excel_[date].xlsx' includes the following important variables:

id (integer) – Unique ID assigned to each theater "entry" in Drupal. A theater with the same name will sometimes be listed more than once (and thus will have more than one theater id). Sometimes this means that the theater has moved, and sometimes it means that two unrelated theaters with the same name appear in two locations.

theater_name (character) – Theater refers to a physical building, sometimes called a "cinema" or "cineplex." We are defining a theater as anywhere where a film was displayed to a public audience. Theater names are not unique.

address (character) – Full address (if known) or intersection. We hope in future to standardize entries in this column.

city, state, city_state (character) – City in Oregon, state (OR), or "City, OR".

latitude, longitude (double/float) – in degrees.

start_date_of_operation, end_date_of_operation (date) – In "yyyy-mm-dd" format. Theaters for which no closing date was entered were coded by the Drupal database as "ongoing" or "still open." This may mean they are in fact still open, or it may mean that the closing date is unknown. In either case, the data export records their closing date as the date the data was last downloaded. These theaters will have the most recent "end_date" entries and are recognizable as many will "end" on the same recent day.

start_year, end_year (integer) – in "yyyy" format.

number_of_seats (character) – venue capacity. This is sometimes an integer, but sometimes it includes more extensive notes or estimates.

owner_and_manager_names (character) – If individual names were created as separate entries in the Drupal database, then each name is separated by a semicolon in this column. However, some entries were created as just one entry separated by commas or have complex annotations. We hope in future to standardize this field to allow exploration of who owned more than one theater.

body, additional_facts, body_html, additional_facts_html (character) – Descriptions of the movie theater written by a student or group of students. "html" versions include all html formatting that creates the page, including links to embedded images. IMPORTANT NOTE: in the 'theaters_excel_[date].xlsx' version of the data set, 'body_html' is replaced by 'body_html_length', which is an integer value listing the number of characters in the 'body_html' column. Because some columns exceed the maximum cell length in Excel, 'body_html' is omitted from the Excel files.

gallery_images (character) – list of 0 to many relative links to images used in the "gallery" section of a blog post, separated by semicolons.

The 'owners_[date].tab' and 'owners_excel_[date].xlsx' files repeat information found in the theaters spreadsheets but create a new row for each owner/manager of a particular theater that was broken out (separated by a semicolon) in the original data. "owner_and_manager_ name's" (character) is the only column containing unique values in this spreadsheet.

The 'articles_[date].tab' and 'articles_excel_[date].xlsx' spreadsheets include a list of articles (blog posts) that are not entries for a specific theater. The articles data have a unique integer id assigned by Drupal, 'gallery_images', 'body', and either 'body_html' or 'body_html_length' columns with the same specifications as the theaters data sets. Columns unique to this data set include 'authored_by' (character), which is the name of the Drupal user who uploaded the article (sometimes but not always the article author), and 'categories' (character), a list of 0 to many topic tags assigned in Drupal and separated by semicolons.

Data users could link articles to theaters spreadsheets via the 'related_cities_and_theaters' column in the articles data, which sometimes indicates that the article is describing a theater set in a particular city. Any such join would be incomplete, since the column takes between 0 and many cities or theaters, separated by a semicolon. The column would need to be divided into multiple columns and parsed to identify cities vs theaters. In future we plan to parse this column for users. Cities are listed in the format "City, OR" and could be joined via the 'city_ state' column in the theaters spreadsheet. Theaters should be listed using the same name used in the 'theater_name' column in the 'theaters' spreadsheet, but there may be errors. Since the combination of 'theater_name' and 'city_state' is likely to be unique, articles could be imperfectly joined to theaters using both columns as keys.

(3) DATASET DESCRIPTION

Object name – Oregon Theater Project Database. See 'OR_Theater_Project_Readme_2022-08. txt' for complete list of filenames.

Format names and versions – tab, txt, xlsx, R, PDF

Creation dates – 2020-01-01 to 2022-08-26

DATASET CREATORS

Michael Aronson and Elizabeth Peterson (University of Oregon) were responsible for conceptualization, funding acquisition, project administration, supervision, dataset creation and editing. John Zhao and Gabriele Hayden (University of Oregon) designed the data export views, and Gabriele Hayden cleaned and curated the dataset.

The following University of Oregon students contributed research and writing to create this dataset: Lauren Adzima, Khalil Afariogun, Andrew Arachikavitz, Malia Balzer, Jacob Beeson, Sylas Bosman, Kyra Brennan, Ezra Brothers, Christian Cancilla, Katy Cannon, Eliza Castillo-Salazar, Jourdan Cerillo, Tom Chamberlain, Shelby Chapman, Cody Churchill, Jude Corwin, Heath Cotter, Julian D'Ambra, Megan Deck, Patrick Dunham, Chloe Duryea, Leah Durkee, Morgan Egbert, Maggie Elias, Jack Elliot, Joseph Endler, Emily Fine, Kyle Fleming, Alex Fox, Javier Fregoso, Sammie Garcia, Hayden Garrett, Ireland Gill, Austin Griggs, Tayte Hansen, Isabella Harrington, Kara Hilton, Ashli Horrell, Amanda James, Zach Jones, Ethan Laarman-Hughes, Addie Lacewell, Abby Lewis, Jimmy Lieu, Kaden Lipkin, Joie Littleton, Wanfang Long, Peter Lovejoy, Shelby Marthaller, Cassie McCready, Carly McDaniel, Brittany McDowell, Brendan McMahon, Eric McMichael, Maddie Miner, Maryam Moghaddami, Jack Moran, Parker Morgan, Nicholas Mundorff, Alexis Neal, Michael O'Ryan, Kelsey Parker, Dre Parkinson, Reese Patanjo, Katherine Pelch, Ben Pettis, Sienna Pigg, Shelby Platt, Ellie Reis, Bailey Rierden, Manuel Rios, Jayna Rogers, Anthoni Rosas, Emily Ruthruff, Payton Schiffer, Becca Schomer, Huntley Sims, Bella Smith, Megan Snyder, Britnee Spelce-Will, Malley Stanovsek, Connor Templeman, Weston Tengan, Jess Thompson, Sarah Tidwell, Evan Vacek, Dylan Wakelin, Jalon Watts, Joe Weber, Makaal Williams, Veronica Wilson, Charlie Winn, David Young, and Sam Zepeda.

Language – English

License – CC-BY

Repository name – Harvard Dataverse

Publication date – 2022-10-31

(4) REUSE POTENTIAL

This data is likely to be of interest to scholars in the humanities and social sciences. It could be used to create new visualizations or digital exhibitions; re-creating a map of these venues, for example, could be a project for an advanced digital humanities course. It could be aggregated with other regional, national, or international cinema history projects, such as that shared on the Mapping Movies site, or could be modified to fit the data model used by Cinema Context or the European Cinema Audiences project[1] to allow for the comparative study of cinema venues (Klenotic, 2022; CREATE, 2022). However, this would require standardizing many of the freeform columns in our data. The information contained in this data set would map onto the Venue, Address, Person, Company, Publication, and Archive tables in the original Cinema Context SQL database (van Oort & Noordegraaf, 2020). This data could also be used in social science research, for example to track the relationship between the opening and closing of theaters and larger socioeconomic trends across Oregon.

One of our anonymous reviewers offered several specific, inspiring suggestions for how our data set, aggregated with others, could be useful in tracking historical questions. For example, the data on theater owners and managers could be cleaned and aggregated with other data sets to map female business ownership during the years leading up to the passage of the 19th amendment granting women's suffrage in the US in 1920. Theater openings and closings might offer insights—particularly when aggregated with other historical business data in Oregon or data on other theaters across the US—into how businesses adapted to economic shocks such as World War I, the 1918 flu pandemic, or the white supremacist terrorism of the Red Summer of 1919.

Scholars seeking to pursue the kinds of data aggregation that would allow for such work must do a great deal of sophisticated data processing to normalize data across differences of data definition and structure. We have done our best to document how our data is defined and

---

1   https://www.europeancinemaaudiences.org/research/, last accessed date: 8 November 2022.

structured to allow for others to build on our work. However, as we discuss in Aronson et al. (2022a), the first challenge scholars face is simply gaining access to the data itself. The data set from that paper includes links to the minority of projects surveyed that do share data as of 2022 and may form a starting point for scholars seeking to do comparative work (Aronson et al., 2022b). We are inspired to share our own small, imperfect data set to model for colleagues what we hope they will do as well: share data early and often, updating as the extent and quality of the data improves over time.

ACKNOWLEDGEMENTS

FUNDING INFORMATION

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

Michael Aronson: Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing

Elizabeth Peterson: Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing

Gabriele Hayden: Data Curation, Writing

AUTHOR AFFILIATIONS

Dr. Michael Aronson   orcid.org/0000-0003-1790-7816
Cinema Studies, University of Oregon, Eugene, OR, USA
Elizabeth Peterson   orcid.org/0000-0003-1258-4122
Digital Scholarship Services, University of Oregon Libraries, Eugene, OR, USA
Dr. Gabriele Hayden   orcid.org/0000-0003-4740-4187
Data Services, University of Oregon Libraries, Eugene, OR, USA

REFERENCES

Aronson, A., Peterson, E., & **Hayden, G.** (2022a). Local cinema history at scale: Data and methods for comparative exhibition studies. (forthcoming). *Iluminace: Journal for Film Theory, History, and Aesthetics, 34*(2). Preprint. DOI: https://doi.org/10.7264/t0ky-0q37

Aronson, A., Peterson, E., & **Hayden, G.** (2022b). "Replication Data for: Local Cinema History at Scale: Data and Methods for Comparative Exhibition Studies". *Harvard Dataverse, V1*. UNF:6:/qdV535CScvkd2ODC/ DAkQ== [fileUNF]. DOI: https://doi.org/10.7910/DVN/6WOQPO

CREATE. (2022). Cinema Context RDF Documentation. Retrieved from https://uvacreate.gitlab.io/cinema-context/cinema-context-rdf/ (last accessed date: 8 November 2022).

Klenotic, J. (2022). Mapping movies. Retrieved from http://mappingmovies.unh.edu/ (last accessed date: 8 November 2022).

van Oort, T., & **Noordegraaf, J.** (2020). The Cinema Context Database on film exhibition and distribution in the Netherlands: A critical guide: arts and media. *Research Data Journal for the Humanities and Social Sciences, 5*(2), 91–108. DOI: https://doi.org/10.1163/24523666-00502008