DATA PAPER

Early African-American Film Database, 1909–1930

Marika Cifor[1], Hanna Girma[2], Shanya Norman[2] and Miriam Posner[3]

[1] Gender, Sexuality and Women's Studies Program, Bowdoin College, US

[2] Digital Humanities Program, UCLA, US

[3] Department of Information Studies and Digital Humanities Program, UCLA, US

Corresponding author: Miriam Posner, Conceptualization, Data Curation, Investigation, Methodology, Supervision, Visualization, Writing – Original Draft, Writing – Review and Editing (miriam.posner@gmail.com)

This dataset covers silent "race films": films created before the year 1930 featuring African-Americans for primarily African-American audiences. The data was collected through research utilizing a wide variety of primary and secondary sources. It contains records on films, actors, production companies, and other key components of the early race film industry. The data is primarily stored on the research data repository site Zenodo and additionally within a relational database on Airtable. Reuse potential of the data includes pedagogical purposes, social network analysis and as a support for the extant study of silent era and race films. The data will also enable further research in film studies, African-American history, the history of film exhibition, and the digital humanities.

(1) Overview

Context

This dataset was produced over the course of one academic quarter by a research team of five undergraduate students, one graduate student, and a faculty member in the Digital Humanities program of the University of California, Los Angeles (UCLA). Taken as a whole, the dataset contains information on films, actors, production companies, and other aspects of early silent-era African-American race films. The database is intended to allow scholars, researchers, and broader public to learn about this period in film history that is too rarely discussed in popular or academic settings.

The data contains the 303 silent race films we identified in our research, linked to 759 actors and other film personnel and 176 race film companies, each record bolstered with descriptive and archival information. The project's website (http://dhbasecamp.humanities.ucla.edu/afamfilm/) also contains a number of maps and visualizations, designed to show how this data might be used, along with a set of beginner-friendly tutorials to aid users' work with the data set. The data is contained within four linked spreadsheets: People, Film, Companies, and Sources. Each of the fields is described in the data dictionary available in the project's README file.

A "race film," as many scholars have noted, is notoriously difficult to define, since various experts give different criteria for the definition and since a film's connotations often change depending on its exhibition context. Indeed, scholarly consensus on race filmmaking is that its personnel constitute not a rigidly bound club but "a circle — a loose federation of production companies and producers who competed with and depended on each other"[1].[1] After an extensive survey of primary and secondary literature on silent and race films, we settled on a definition of a race film as *a film with African-American cast members, produced by an independent production company and discussed or advertised as a race film in the African-American press.*

We thus started composing the dataset by casting a very wide net, and then gradually pared down our data. While we immediately discarded known blackface films, we recorded most of the films from our time period captured in the secondary literature on African-American film, particularly those contained in Larry Richards's *African-American Films through 1959: A Comprehensive Filmography* and Henry Sampson's *Blacks in Black and White* [2].[2] Sampson's and Richard's filmographies are both more capacious than our own, containing not only race films per se, but also all films featuring African-American actors in prominent roles.

As we refined our definition of race filmmaking, we then considered each film individually, eliminating those films produced by mainstream production houses and those we had learned were blackface comedies designed for white audiences. We also supplemented the data collected from secondary research with additional films we had uncovered in our primary-source research. For a full, annotated list of the archival and secondary sources we employed, please see http://dhbasecamp.humanities.ucla.edu/afamfilm/sources-further-reading/.

(2) Methods

Steps

We began our research by entering data we derived from the clipping and note files contained in the George P. Johnson Negro Film Collection, a collection held within the UCLA Library Special Collections at the Charles E. Young Research Library. George P. Johnson, a key figure in the Lincoln Motion Picture Company, assembled these files both during his time in the race film industry and for five decades afterward [3].[3] Early in our archival research we discovered, however, that while the materials within the Johnson Collection are invaluable, they also contain a number of discrepancies and omissions. Therefore, we quickly expanded our scope to consider other sources, both secondary sources (books and articles) and primary sources (chiefly African-American newspapers from the period).

As we gathered information about films and people from the period, we collaboratively entered the information into the relational database software Airtable (www.airtable.com). In determining whether a film or person met our criteria for inclusion in the dataset, our policy was to err on the side of comprehensiveness, and then to confirm the record's appropriateness for conclusion after the fact, generally by checking primary sources including African-American newspapers to see if the film circulated among African-American audiences as a "race film."

Because of the number of people entering data, we then employed a number of strategies to eliminate duplicates and variant spellings, described in the "Quality Control" section below.

Sampling strategy

There was no sampling process completed with this data. This project aimed to capture the entirety of the race film industry in this period with as great a depth as possible. Therefore, this dataset contains every race film produced prior to 1930 which we were able to discover and to verify, as well as all of the film personnel and film companies involved in the early race film industry we uncovered.

Quality Control

For the purpose of this project, we had to solidify the ambiguity of the term "race film" in order to refine our data. Employing the definition proposed by Pearl Bowser, Jane Gaines, and Charles Musser, in their introduction to *Oscar Micheaux and His Circle* we determined that we would only include silent films created before 1930 for African-American audiences. This definition was the primary factor that informed our decisions to include or exclude pieces of data [4].[4]

Since we began by casting a wide net and then winnowed our database to its current size, we excised a significant number of records from the dataset. So that other scholars can evaluate our decisions, the films and people we discarded from the database are themselves captured in our dataset, as three linked Excel files in a folder labeled "discarded_data."

In the course of our research, we discovered many discrepancies in both primary and secondary sources, some of which have apparently propagated through the literature over decades of scholarship. For example, the actor Ardelle Dabney sometimes appears as Ardella Dabney or Adelle Dabney. Therefore, wherever possible, we have independently verified film titles and personal names that appear in secondary literature, sourcing them to newspapers or other primary materials from the period. Where variant names, titles, or other information exists, we have captured those variations in the "AKA" field (for personal names) or "alternate title" (for films). We have also made notes in the "Notes" field of any discrepancies our research has uncovered.

To normalize the data, we used several clustering methods, available for data cleanup and transformation in OpenRefine, an open source desktop application for such data wrangling activities. A range of methods and algorithms were employed to discover possible redundancies. First, we used a series of key collision methods, including fingerprinting and phonetic fingerprinting. The latter method that gathers together similar sounding words, for example, allowed for us to identify errors that are likely perpetuated as to people misunderstanding or guessing as to the spelling of a word, particularly in this case less common names, after only hearing them spoken aloud. Key collision methods are quick and simple to employ using OpenRefine, however they vacillate frequently between being too strict or too lax in their assessment of how much difference to tolerate between strings analyzed. Therefore, we utilized a series of nearest neighbor methods, including both Levenshtein distance and PPM available in OpenRefine. These latter methods allow for fine tuning of distance thresholds between strings. In every case of a possible redundancy, we examined the records individually, verified that the duplication was an error via searches in the primary literature, and, where appropriate, collapsed the records.

(3) Dataset description

Object name

Race Film Database.

Format names and versions

Four linked CSV tables, including companies.csv, films.csv, people.csv, and sources.csv. Each of these tables contains a number of variables. The tables are linked via production companies, films, people, and sources. In addition, there are three Excel tables consisting of discarded data.

Creation dates

2016-03-01 to 2016-06-08.

Dataset Creators

Berry, Monica. Data curation, investigation, methodology, visualization, writing – original draft, writing – review & editing. Affiliation: Digital Humanities Program, UCLA.

Cifor, Marika. Data curation, investigation, methodology, visualization, writing. Affiliation: Department of Information Studies, UCLA.

Contreras, Karla. Data curation, investigation, methodology, supervision, visualization, writing – original draft,

writing – review & editing. Affiliation: Digital Humanities Program, UCLA.

Girma, Hanna. Data curation, investigation, methodology, visualization, writing – original draft, writing – review & editing. Affiliation: Digital Humanities Program, UCLA.

Lam, William. Data curation, investigation, methodology, visualization. Affiliation: Digital Humanities Program, UCLA.

Norman, Shanya. Project administration, investigation, writing – original draft, writing – review and editing. Affiliation: Digital Humanities Program, UCLA.

Posner, Miriam. Conceptualization, data curation, investigation, methodology, supervision, visualization, writing – original draft, writing – review & editing. Affiliation: Digital Humanities Program, UCLA.

Yoshioka, Aya Grace. Data curation, investigation, methodology, visualization. Affiliation: Digital Humanities Program, UCLA.

Language
English.

License
CC-BY 4.0.

Repository name
Zenodo.

Publication date
2016-06-08.

(4) Reuse potential

These data have reuse potential for scholarly research within studies of silent era and race films and filmmaking, as well as within other scholarship in film studies, African American history, and the digital humanities. All entries contain a citation or a link to the primary and/or secondary source where the data was obtained for anyone needing to validate specific entries or seeking to further their knowledge. The data could be combined with other similar sets of data, or could be mashed up with datasets that draw related variables together. For example, the HoMER network (History of Moviegoing, Exhibition, and Reception) might combine our film titles with its list of film exhibitions, thus helping to shed light on where these important works were actually shown. It also might be combined with information about the locations of existing films, thus allowing film viewers to find copies of the few race films that have survived.

The data also have potential for reuse and augmentation by archivists and librarians who hold collections related to the race film industry or film history more broadly. It may also provide the basis for more complete description of existing archival materials and collections and for building new relationships between archival institutions collecting in these areas.

The data is valuable as a teaching and learning tool for students in digital humanities, African-American, film studies, and other courses related to American history and culture. Each data visualization we have included is accompanied by

the specific data used and step-by-step instructions on how to create similar visualizations. The project's website features a series of tutorials on working with this dataset. These tutorials are aimed at making the data accessible for use in visualizations and analyses regardless of the user's experience level. For example, the dataset does include locations for production companies, and in the tutorials, we provide the basics of mapping using the dataset and refer users to further resources.

A particular strength of the database we created is that it will allow researchers to newly identify and visualize the connections among various entities and central figures in this social world. Social network analysis is thus a fruitful area for potential reuse of the data. On the project website, we include a series of network diagrams we created with the data. For example, a diagram of all of the people included in the dataset ranging from actors to writers to cinematographers exposes how closely connected this community was in spite of the porous boundaries of race filmmaking itself. We discovered through this analysis that the race film network seems to be composed of one main component and a number of smaller components. The large network comprised of people who worked on films produced by Oscar Micheaux, the Lincoln Motion Picture Company, and Foster. Ebony film players are also deeply embedded within this larger network, a factor that is intriguing, given the Ebony company's controversial status as a race film company. A set of tutorials on our website offers guidance to users who wish to conduct their own social network analyses.

While this dataset represents our best effort to capture the people, companies, and films active in the race film industry, no dataset can perfectly capture the full complexity and dynamism of this period. For example, some films we included in our data, such as those produced by the Ebony company, might be considered race films by some researchers and not others. Similarly, ownership and personnel of race film companies changed a great deal over the time period at stake, a dynamism we don't capture in the data. However, this is the first publicly available dataset of the industry, and our hope is that other scholars use it as a starting point.

We also hope that this project increases interest in this part of history that is too rarely discussed. The early African-American film industry gives us insight into multiple facets of history and culture, and it is our hope that this project will prompt others continue to explore its historical and contemporary significance.

Notes

[1] Pearl Bowser, Jane Gaines, and Charles Musser, "Introduction: Oscar Micheaux and Race Movies of the Silent Period," in *Oscar Micheaux and His Circle: African-American Filmmaking and Race Cinema of the Silent Era*, by Pearl Bowser, Jane Gaines, and Charles Musser (Bloomington: Indiana University Press, 2001), xvii–xxx, xx.

[2] Richards L. *African American Films Through 1959: A Comprehensive, Illustrated Filmography*. Jefferson N.C.: McFarland; 1998; 1. Sampson H. *Blacks in Black and*

White: A Source Book on Black Films. 2nd ed. Metuchen N.J.: Scarecrow Press; 1995.

[3] George P. Johnson Negro Film Collection (Collection 1042). UCLA Library Special Collections, Charles E. Young Research Library, University of California, Los Angeles.

[4] Bowser P, Gaines J, Musser C. Introduction: Oscar Micheaux and Race Movies of the Silent Period. In: Oscar Micheaux and His Circle: African-American Filmmaking and Race Cinema of the Silent Era. Bloomington: Indiana University Press; 2001. p. xvii–xxx.

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

Authors Information

Cifor, Marika. Data curation, investigation, methodology, visualization, writing.

Girma, Hanna. Data curation, investigation, methodology, visualization, writing – original draft, writing – review & editing.

Norman, Shanya. Project administration, investigation, writing – original draft, writing – review and editing.

Posner, Miriam. Conceptualization, data curation, investigation, methodology, supervision, visualization, writing – original draft, writing – review & editing.

References

1. **Bowser, P, Gaines, J** and **Musser, C** 2001 *Oscar Micheaux and His Circle: African-American Filmmaking and Race Cinema of the Silent Era.* Bloomington: Indiana University Press.

2. **Sampson, H** 1995 *Blacks in Black and White: A Source Book on Black Films.* 2nd ed. Metuchen N.J.: Scarecrow Press.

3. **George, P** Johnson Negro Film Collection (Collection 1042). UCLA Library Special Collections, Charles E. Young Research Library, University of California, Los Angeles.

4. **Richards, L** 1998 *African American Films Through 1959: A Comprehensive, Illustrated Filmography.* Jefferson N.C.: McFarland.