SARAH-MAI DANG

The *Women Film Pioneers Explorer*

What Data Visualizations Can Tell Us about Women in Film History

ABSTRACT In view of the increasing production and use of data in the course of digitalization, the goal of feminist film historians to increase the visibility of women's work has taken on a new urgency. Through the production, processing, and dissemination of data, blind spots in a research field such as feminist film history can be maintained or amplified, but also minimized. Access to data as well as the critical reflection on that data is therefore one of the greatest challenges for humanistic scholars today. Against this backdrop, this article discusses how digital data visualization can enhance and transform research on women in early cinema. Presenting a case study on the *Women Film Pioneers Explorer*, I argue that data visualizations can help us reflect on our own (feminist) film historiographical approaches, epistemological premises, and representative conventions and thus on the "situatedness of knowledges." KEYWORDS data visualization, digital film history, early cinema, feminist film history, situated knowledges, research data management, women workers

INTRODUCTION

The *Women Film Pioneers Explorer* (WFPE) is a digital visualization project based on the data of the *Women Film Pioneers Project* (WFPP). Like the WFPP, the *Explorer* seeks to make research on women in early cinema accessible to a wider audience.[1] With its various visualizations, it makes the global work of women, and the blind spots and research gaps, in early cinema more visible. In doing so, it introduces new perspectives on film history and helps evaluate existing ones, such as the concept of film, canon, and authorship. In addition to stimulating further reflections on digital methods in film history, the WFPE website encourages users to participate in the WFPP as well as create additional visualizations to analyze and share research data.

The *Women Film Pioneers Explorer* is the result of a two-semester project seminar in computer science, conducted in 2020/2021 at Philipps-Universität Marburg. The project represents a collaboration between the

BMBF research group Aesthetics of Access: Visualizing Research Data on Women in Film History (DAVIF), which I led, and the working group Graphics and Multimedia, led by Thorsten Thormählen. The goal of my research group, funded by the German Federal Ministry of Education and Research (BMBF) for four years (2021–2025), is to explore data visualizations as a tool for telling different stories differently.[2] The *Women Film Pioneers Explorer* is the first case study of DAVIF.

THE WOMEN FILM PIONEERS EXPLORER (WFPE)

The *Women Film Pioneers Explorer* (figure 1) allows users to explore the data collection of the *Women Film Pioneers Project* through interactive visualizations.[3] The WFPP is a freely accessible platform for research on women workers in silent film. It was launched in 2013.[4] As of August 2022, it contains more than three hundred career profiles (including images and sometimes film clips) written by scholars, curators, and archivists. It also features a couple of overview essays and other information. The WFPP is an invaluable yet expandable resource for film historical research.[5] In addition to the DFF—Deutsches Filminstitut & Filmmuseum—the WFPP is one of the film historiographical project partners of our research group, both of whom have generously agreed to make their research data available for our explorations and to exchange knowledge and expertise.
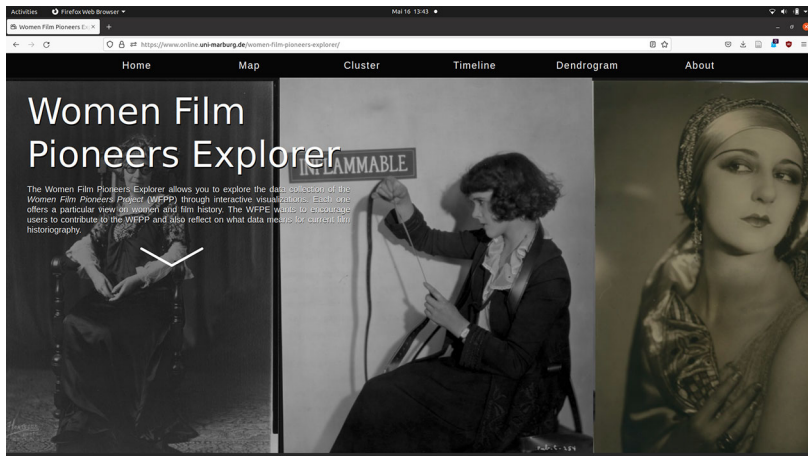

FIGURE 1. The landing page of the *Women Film Pioneer Explorer*, screenshot taken May 16, 2022, www.online.uni-marburg.de/women-film-pioneers-explorer.

For the *Women Film Pioneers Explorer*, five computer-science master's level students, Henri Dickel, Matija Miskovic, Karazm Noori, Christian Schmidt, and Atefeh Soltanifard, tested a number of classic forms of visualization to explore new ways of displaying the WFPP's research and to understand which data the platform actually holds. They documented the creation process on GitHub.[6] The dataset requested for this case study and provided in CSV format includes: Pioneer IDs, names, aliases, permalinks of individual profiles, links of portrait photos, job titles, workplaces, and dates of each individual's birth and death. It is archived by the library of Columbia University.[7] Records of birth and death were not documented by the WFPP, as that project emphasizes women's careers and not their biographies. Importing the WFPP's data into various visualizations, the students created a map, a cluster, a timeline, a dendrogram, a histogram, and statistical charts, which were brought together in the *Women Film Pioneers Explorer*. These visualizations allow the selected data to be viewed in a geographical, proportional, relational, chronological, and hierarchical manner.

The map shows the women's global activities during a selected period of time (figure 2). This example also reveals that the WFPP's database contains information on fifteen women who worked in Germany, among other countries, between 1895 and 1926. These fifteen women had at least a total of twenty-five different jobs and worked in at least twelve different countries (listed under "connections" somewhat confusingly). In the left column, an "example pioneer" is shown on a random basis. In the screenshot, the website
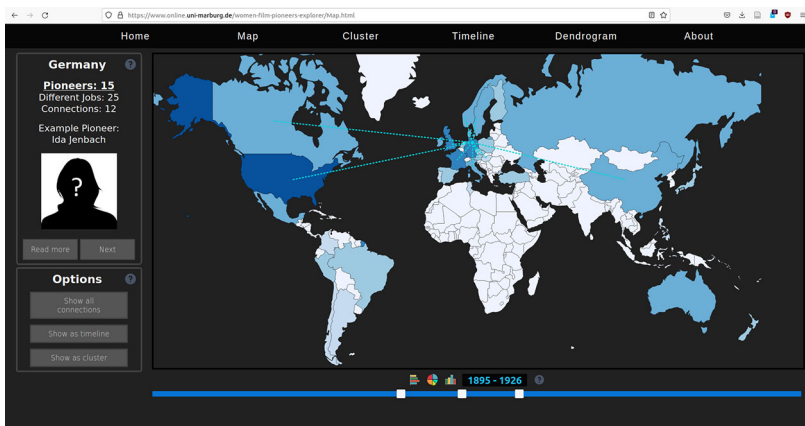

FIGURE 2. A map displays global connections of women workers, screenshot of the *Women Film Pioneer Explorer*, taken May 2, 2022, www.online.uni-marburg. de/women-film-pioneers-explorer/Map.html.

features Ida Jenbach, for whom there is no photo in the database, at least not at the time the visualization was completed in autumn 2021. The "read more" button takes the user to the respective profile on the WFPP website, which contains detailed information on the individual's career and suggestions for further reading. The "show all connections" button displays all known careers the women pursued during the selected period regardless of geographical location, while the "show as timeline" and "show as cluster" buttons take the user to other visualization variants that display the results of the query on a timeline or in a group. The period of time selected for any given search can be varied as desired using the axis located under the map. The graphic symbols next to the axis take the user to statistical graphs, a bar chart, pie chart, and column chart.

The results of this example search alone are revelatory. With fifteen women and twenty-five jobs, the map demonstrates that one woman had more than one job, or that the authors of the WFPP profile pages assigned several different jobs to these women. It is remarkable that these fifteen women worked in at least twelve different countries, including Germany. In order to learn which women worked where, the user can interact with the timeline visualization. As will be explained, if the user accesses the map through the timeline, the WFPE highlights the pioneers' workplaces organized geographically.

While it is fascinating from a film historical point of view to see the numerous connections and individual paths, it is important to keep in mind that this map, like the entire *Explorer*, is based on a project-specific data query. It does *not* represent *all* women in early cinema. Logically, it can only display what has been previously collected by the WFPP. Like any other database, the WFPP cannot hold a complete collection on women in film history. Databases are always the result of certain conditions in certain place and time constellations. They represent situated knowledges.

It bears stressing, in other words, that the WFPP data derives from research results on women in early cinema curated according to editorial criteria. Those criteria, in turn, are continuously changing and expanding. For this reason it is by no means the case, as the map suggests, that only one woman worked in Tunisia's early film industry, let alone in the entire African continent. Instead, a huge gap becomes visible in this visualization, namely, that the WFPP does not hold any data on this continent from this period. Yet. This is understandable insofar as the project started in the United States and is hosted by Columbia University in New York, as is the main initiator,

Jane Gaines. Even though the WFPP never claims to draw a complete picture of women in early cinema, but on the contrary calls for further contributions, the false impression may still arise that developments in early film culture took place far from Africa. In principle, it is therefore important to be aware that the underlying data material must always be analyzed to fully understand visualizations. Thus, I would argue that the *Explorer* encourages users to reflect on what has been included and excluded from film history.

The pie chart (figure 3), which can be accessed via the map, provides further information about the data collection of the WFPP. This example shows that the United States tallies 47.1 percent or 163 entries for the years 1895 to 1926, and thus represents the country with the most profiles. This result does not mean that almost half of the women listed in the database were born in the United States. The numbers instead refer to the countries where the women worked. The United States is followed by Great Britain (9.2 percent, 32 entries), France (5.2 percent, 18 entries), Germany (4.3 percent, 15 entries) and Australia (4.0 percent, 14 entries). About a third of the entries refer to other countries. Again, it is worth noting that some women worked in several countries. As the map demonstrates, women who worked in Germany were particularly likely to work in other countries as well.

In the cluster visualization (figure 4), the pioneers can be grouped according to job title or countries in a proportional manner. The bigger the bubble, the more entries will be found in the WFPP's collection. The featured
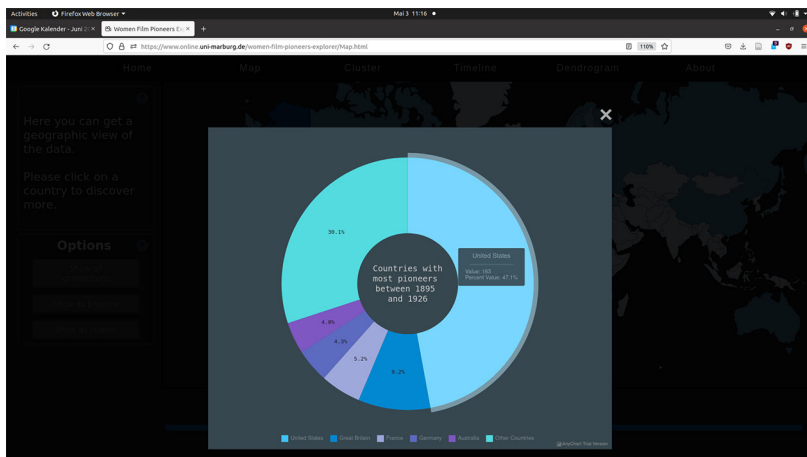

FIGURE 3. A pie chart displays the research foci of the WFPP, screenshot of the *Women Film Pioneer Explorer*, taken May 3, 2022, www.online.uni-marburg.de/women-film-pioneers-explorer/Map.html.

FIGURE 4. A cluster shows the many different occupations that the pioneers practiced, screenshot of the *Women Film Pioneer Explorer*, taken May 2, 2022, www.online.uni-marburg.de/women-film-pioneers-explorer/Cluster.html.

pioneers change continuously on a random basis. A third variant of the cluster displays all three hundred women profiles included in the WFPP at once.[8] As in the map, the individual entities are directly linked to the respective profile pages.

The timeline (figure 5) shows the span of years each pioneer lived, bracketed by when they were born and died. Entries can be sorted chronologically or alphabetically. It is also possible to add other events, such as World War II,

FIGURE 5. A timeline shows when each pioneer lived, screenshot of the *Women Film Pioneer Explorer*, taken May 2, 2022, www.online.uni-marburg.de/women-film-pioneers-explorer/Timeline.html.

and thus carry out an individual historical contextualization. The entries are linked to the *Explorer*'s map, which then shows the geographical location where the selected pioneer worked (which cannot be identified in detail when visiting the map first).

The dendrogram was selected as the sixth form of visualization (figures 6a, 6b). It reveals the myriad professions in early cinema, organized in a hierarchical manner. In doing so, it shifts the focus from the privilege often granted in film history to the individual author/director and reveals a more complex, collective conception of film production. The display of the many diverse professions makes it clear that film is not just a work of art, but also the result
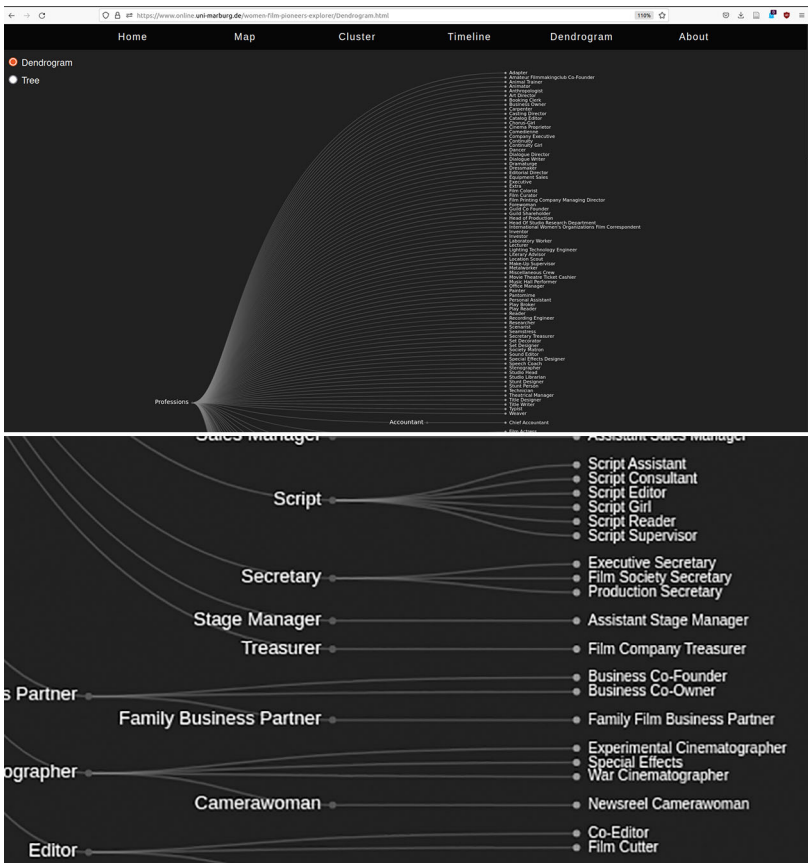

FIGURE 6A, B. The dendrogram shows the many professions in a hierarchical manner, screenshot taken May 3, 2022, www.online.uni-marburg.de/women-film-pioneers-explorer/Dendrogram.html.

of a collaborative effort. It becomes obvious that in addition to directing, script writing, and camera work, a considerable array of expertise is required for a film—not least the secretary's work.[9]

CONCLUSION: SOURCE CRITICISM AND RESEARCH PROCESSES

Visualizations do not only represent information but also can *generate* information and raise further questions. While it is instructive to learn by navigating the *Explorer* that the WFPP holds data on fifteen women who worked in eleven other countries besides Germany and in a total of twenty-five different professions between 1895 and 1926, it would make sense to create further visualizations to explore the data in more depths. For example, it would be helpful to know which pioneer worked at what time, where, and in what capacity. And if film is to be grasped beyond the idea of an *auteur*, it would also be interesting to find out who worked with whom on which film. It should be possible to aggregate this information using the WFPP profiles. In addition, it would certainly be helpful to implement a search field for individual queries. To be able to do this, however, the WFPP data would first have to be specifically modeled and organized to ensure meaningful search results.

These and other suggestions for improvement (for example, linking to other databases such as Wikidata according to Linked Open Data principles, adding birthplaces to show the women's migration movements, or defining job titles in the dendrogram) were discussed in various contexts, most recently in a workshop of the research group at the Women and the Silent Screen XI Conference. If the visualization were to be the central outcome of our project, these suggestions should be implemented in a next version. Within the framework of our research, however, the *Explorer* serves primarily as an analytical approach for the systematic reflection of data-based methods in the context of feminist film history. Even if the *Explorer* may appear quite conventional from a visual point of view and not particularly elaborate compared to artistically designed data visualizations, it provides many new insights—in terms of data visualizations and the specific data sources as well as the entire research process.

Evaluating the *Women Film Pioneers Project*, it soon became clear that a comprehensive investigation of the data provenance was necessary in order to fully understand the result of the case study. For this reason, after the launch of the website our research group conducted expert interviews with

the WFPP project manager, Kate Saccone, and project leader and founder, Jane Gaines, on the genesis of the long-standing endeavor and workflows.[10] It was particularly revealing to learn that the diverse job titles are the result of individual discussions and not standardized data models. In intensive debates between the editors (mostly Gaines or Saccone, but also other editors assigned to particular countries) and the authors of the profile pages, job titles are defined and assigned to the women pioneers. This means that the metadata in the CSV file is based on the research of film historians and not on a taxonomy or an ontology. This explains why such unusual job titles like "society matron" appear in the dendrogram. Assuming that film history itself is shaped by coincidences, contradictions, and contingencies, the WFPP emphasizes epistemological uncertainties by allowing the authors to categorize the pioneers based on their research. Therefore, a conscious decision was made against a classification system.

The work of the *Women Film Pioneers Project* is mainly shaped by specific research interests for the purpose of a plural film historiography and less by curatorial considerations for the purpose of an interoperable reuse.[11] The latter determines, for example, the approach of the Deutsches Filminstitut & Filmmuseum (DFF), which models its data according to the European standard EN15907.[12] In contrast to the film historiographical database of the DFF, the WFPP focuses more on the pioneers than the films. However, what it means to define a person as a "Woman Film Pioneer" is a question the editor-curators themselves keep asking. What do the terms *woman*, *film*, and *pioneer* mean in the light of current film and gender theoretical discourses? What are the underlying premises and assumptions of these designations? For example, does the concept of the "pioneer" imply the very premises of a master narrative that feminist scholars generally criticize because it obscures the collaborative work necessary for film production and even bears imperialist connotations?[13]

With the focus on the data provenance, the project partner itself became the object of our research. While the initial aim was to use the WFPP's biographical data to make research on women in early cinema more visible and to explore data visualizations, the complex production of the source material gradually came to the fore of the study. Analyzing the data in addition to the visualization, various premises and work processes of the WFPP could be recognized that had hardly been noticed before. As already emphasized, the *Explorer* does not and cannot represent a global overview of all the women at the beginning of film history. Instead, it visualizes the

database of the WFPP, and strictly speaking it visualizes an excerpt, namely, the queried metadata. The *Explorer* shows what data is in the collection and where—so far—there seems to be little information. In this respect, the visualizations are also a corpus analysis of the WFPP. They allow users to engage in-depth with the project's politics of representation and thus with a specific approach to feminist film historiography. Since the data on which the *Explorer* is based are not updated, it is a kind of snapshot of film historiographical work. Thus, this case study not only helps us reflect on what data visualizations can tell us about women in film history but also what they can tell us about *doing* film history. ■

SARAH-MAI DANG is Principal Investigator of the BMBF research group Aesthetics of Access: Visualizing Research Data on Women in Film History (DAVIF) (2021–2025) at the Institute of Media Studies at Philipps-Universität Marburg. She also founded the DFG research network New Directions in Film Historiography (2019–2022). In addition, she initiated the Open Media Studies blog and the publishing project oa books. She received a doctoral degree in Film Studies from Freie Universität Berlin and a Master of Arts from the University of Michigan and has published widely on the effects of digitization in the humanities. (https://orcid.org/0000-0002-1960-247X, @SarahMaiDang@fedihum.org,)

ACKNOWLEDGMENTS

NOTES

1. This article is based on a translation of the author's book chapter, "Forschung explorieren. Zu den Möglichkeiten digitaler Datenvisualisierungen für die feministische Filmgeschichtsschreibung," in *Handbuch Digitale Medien und Methoden*, ed. Laura Niebling, Felix Raczkowski, and Sven Stollfuß (Wiesbaden: Springer, 2023).

2. For a project description, see www.uni-marburg.de/en/fb09/institutes/media-studies/research/research-projects/davif.

3. See Henri Dickel, Matija Miskovic, Kharazm Noori, Christian Schmidt, Atefeh Soltanifard, Sarah-Mai Dang, and Thorsten Thormählen, "Women Film Pioneers Explorer," www.online.uni-marburg.de/women-film-pioneers-explorer/, 2020. Last accessed July 25, 2022.

4. See Jane Gaines, Radha Vatsal, and Monica Dall'Asta, "About the Project—Women Film Pioneers Project," https://wfpp.columbia.edu/about. Last accessed May 3, 2022.

5. See Sarah-Mai Dang, "Unknowable Facts and Digital Databases: Reflections on the Women Film Pioneers Project and Women in Film History," *Digital Humanities Quarterly* 14, no. 4 (2020). www.digitalhumanities.org/dhq/vol/14/4/000528/000528.html.

6. https://github.com/MatijaMi/data-vis-women-in-film.

7. See Jane Gaines and Columbia University Libraries, "Women Film Pioneers Project Biographical Data," 2020. https://doi.org/10.7916/m4dc-n768.

8. The WFPP manages several hundred entries on pioneers of the silent film era. While there are now more than three hundred profile pages, as of July 2022, it lists 641 "unhistoricized" pioneers (names that are still awaiting a more detailed description in the WFPP).

9. See Erin Hill, *Never Done: A History of Women's Work in Media Production* (New Brunswick, NJ: Rutgers University Press, 2016), 22–24. However, it is important to note that the term "secretary" can have very different meanings. It is thus important to take a closer look at each job title and the respective job profile. Both can change over time and vary from country to country. This is especially interesting when it comes to the translation of terms.

10. A catalog of questions was drawn up on the basis of the first preliminary talks. A structured interview has already been conducted with Kate Saccone. An interview with Jane Gaines will follow. The interviews will be edited and published.

11. In the context of data modeling, a distinction is generally made between curation-driven and research-driven. While curation-driven modeling focuses primarily on potential subsequent use in different contexts and, as a result, seeks to ensure interoperability through standards, research-driven modeling primarily aims at a concrete research interest that stems from a specific disciplinary context. See Fotis Jannidis and Julia Flanders, "A Gentle Introduction to Data Modeling," in *The Shape of Data in the Digital Humanities: Modeling Texts and Text-Based Resources*, ed. Julia Flanders and Fotis Jannidis (London: Routledge, 2019), 26–94. However, these definitions cannot always be clearly separated from each other, as pointed out in another book chapter. See Sarah-Mai Dang, "o.J.—Recherchepraktiken, Datenquellen und Modellierungen," in *Doing Research. Wissenschaftspraktiken zwischen Positionierung und Suchanfrage*, ed. Sandra Hofhues and Konstanze Schütze (Bielefeld: Transcript, 2022), 330–37, 333.

12. Deutsches Filminstitut (DIF) e.V, "EN 15907: Filmstandards.org," http://filmstandards.org/fsc/index.php?title=EN_15907 (2011). Last accessed June 20, 2022.

13. See Kiki Loveday, "The Pioneer Paradigm," *Feminist Media Histories* 8, no. 1 (2022): 165–80.