

Sarah-Mai Dang

Managing the Past: Research Data and Film History

Introduction

Given the growing number of data-driven research projects, research data management (RDM) is becoming increasingly important across all disciplines. As I have elaborated elsewhere, it has also become a relevant field in film and media studies, a field which is very likely to grow in the near future.¹ RDM describes the practices of collecting, selecting, modeling, organizing, preserving, and sharing of data in order to facilitate their access and reuse. This involves a sequence of many different steps and phases that may be repeated and vary depending on the research project and discipline. Ideally, RDM involves a management plan that outlines these steps in a research project including data sources, formats, documentation, storage, and access. Today, there are numerous tutorials and guidelines that describe the complex workflows, commonly defined and explained as data lifecycles.²

However, even though there are now many initiatives that seek to address the specific humanities requirements of RDM,³ there is nevertheless much to explore and understand about particular disciplinary challenges. While there are numerous resources that provide general advice, when it comes to specific pro-

1 Sarah-Mai Dang, “Forschungsdatenmanagement in der Filmwissenschaft. Daten, Praktiken und Erkenntnisprozesse,” *montage* AV 29, no. 1 (2020): 119–140.

2 For example, for the German-speaking community the website forschungsdaten.info has become a central reference: *Forschungsdaten.info*, “English Pages. Forschungsdaten und Forschungsdatenmanagement,” accessed February 3, 2023, <https://forschungsdaten.info/english-pages/>. In the Netherlands, the idea of “Data Scopes” (<https://data-scopes.github.io/Data-Scopes/>) has been discussed as part of broader critical tool discussions. Data Scopes explores how research data may (or may not) be linked and shared to ensure greater transparency and to understand how data shape research. Rik Hoekstra and Marijn Koolen, “Data Scopes for Digital History Research,” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 52, no. 2 (2019): 79–94. <https://doi.org/10.1080/01615440.2018.1484676>. I thank Christian Gosvig Olesen for pointing out the “Data Scopes” project in his feedback on an earlier version of this chapter.

3 For instance, specially established digital research centers at universities now offer comprehensive counseling on research data management. Additionally, the recently launched National Research Data Infrastructures (NFDIs) in Germany aim to identify central questions in this regard across disciplines, promote cross-project exchange, and provide concrete assistance in the event of problems.

ject issues such as terms of use for film historical databases or subject-specific storage options, there are still few examples of best practice in film and media studies. Thus, until now learning by doing has been the most effective approach. In addition to practical and legal challenges that must be addressed, we also need to pay more attention to the cultural and political implications of RDM. Until recently, RDM has often been treated as a purely organizational or technical side task, required by funding organizations or the respective institution in terms of reusability and good scholarly practice. Humanities scholars tend to view it as a nuisance because it is not considered to be part of humanities research.⁴ Yet, as more and more scholars are becoming aware, the way RDM is conducted deeply affects our understanding of film culture as well as the scholarship that explores it. For example, what credits we retrieve or include in a database to identify or describe a film does not only determine how we can search for specific criteria, but also reflects what we consider relevant for analyzing film culture. If we list only the title of a film, the year and country of production, and the director, as is common in scholarly works, we cannot analyze the involvement of other professions such as the editor or screenwriter. This demonstrates that ascribing specific metadata to artifacts is not a neutral procedure, but an act of interpretation based on theoretical premises and cultural assumptions. In this respect, interpretation is part of all RDM practices and does not only take place afterwards.⁵ Thus, different from what one might associate RDM with, it is neither pure technical or organizational, nor universally valid. It is shaped by intellectual conventions and institutional frameworks, and vice versa; it shapes our concepts and ideas of what we perceive as film or authorship.⁶ We should not therefore submit a data management plan just to meet formal funding requirements or institutional practices, but, on the contrary, use it to reflect and help shape research.

In this chapter, I reflect on RDM and its impact on digital scholarship regarding film history. Rather than providing a guideline for implementing RDM in film and media studies, I want to draw attention to the theoretical and political implications of RDM that can serve to conduct more informed data practices. In doing

⁴ Sophie G. Einwächter, “Forschungsdaten (in) der Film- und Medienwissenschaft – Sophie G. Einwächter über vorurteilsbehaftete Begrifflichkeiten und fruchtbare Momente in der Lehre,” *Open Media Studies Blog* (2019), accessed March 21, 2023, <https://mediastudies.hypotheses.org/1314>.

⁵ Lisa Gitelman and Virginia Jackson, “Introduction,” in *“Raw Data” Is an Oxymoron*, ed. Lisa Gitelman (Cambridge, MA: MIT Press, 2013), 3.

⁶ See also Julia Noordegraaf, Kathleen Lotze, and Jaap Boter, “Writing Cinema Histories with Digital Databases: The Case of Cinema Context,” *TMG Journal for Media History* 21, no. 2 (2018): 106–126, <https://doi.org/10.18146/2213-7653.2018.369>.

this, I will discuss questions such as: How is the data organized? By whom and for what purpose? What information has been extracted? Which sources have been ignored? How do current premises and ideas inform digital curation practices? How does RDM shape our understanding of film culture?

While data is not only numbers but can also be texts, images, audio, or video,⁷ this chapter focuses on film historical metadata. Drawing on critical approaches to data-driven projects and infrastructures, as well as my own experience with film historical databases, I seek to outline a framework that allows us to systematically scrutinize RDM. To this end, I will compare two digital data initiatives: the Women Film Pioneers Project (WFPP) and filmportal.de of the DFF – Deutsches Filminstitut & Filmmuseum. Given their different designs and objectives, a comparison promises to provide interesting insights into the particularities of RDM and thereby emphasize its relevance. Both initiatives are project partners of my current research group “DAVIF – Visualizing Research Data on Women in Film History” (2021–2025). They provide the research data for exploratory data visualizations in addition to their curatorial knowledge and archival expertise.⁸ This means I have not been involved in the production and curation of the data in any of the projects. Instead, I look at the data as a humanities scholar in retrospect to understand the objects at the heart of my research.

While the initial aim of DAVIF was to make research data on women in early cinema more visible by the means of data visualizations, analyzing how the research was produced in the form of data has unexpectedly become a significant aspect of my research. Conducting the first case study, it soon became clear that, in order to make sense of the data visualizations we created in our project, I needed to understand the underlying source material as well. My focus shifted from exploring the presentation and reuse of research data to critically scrutinizing the digital source. As emphasized by scholars in the digital humanities, when working with data, the research process becomes a fluent, interconnected series of different steps that depart from the more linear structure of traditional methods.⁹ During the course of my study, it became apparent that data visualizations not only enhance existing research, but can also provide valuable insights into data and serve

⁷ Dang, “Forschungsdatenmanagement in der Filmwissenschaft,” 121–122.

⁸ The project is funded by the German Federal Ministry of Education and Research (BMBF) for four years (2021–2025), <https://www.uni-marburg.de/en/fb09/institutes/media-studies/research/research-projects/davif>.

⁹ Andreas Fickers, Juliane Tatarinov, and Tim van der Heijden, “Digital History and Hermeneutics – Between Theory and Practice: An Introduction,” in *Digital History and Hermeneutics: Between Theory and Practice*, ed. Andreas Fickers and Juliane Tatarinov (Berlin, Boston: De Gruyter Oldenbourg, 2022), 9–10.

as a method for the analysis of the data corpus. To better understand the underlying premises and practices, we also conducted semi-structured interviews with individuals involved in the management of the databases. Through these interviews, I was able to understand essential features of the sources and workflows that generated long-standing initiatives.¹⁰

Analyzing Film Historical Metadata

Although the data lifecycle model does not imply that RDM is purely technical, but is intended to simplify the complexities involved, I approach the topic from a different angle in order to emphasize the entangled practices, assumptions, and external factors that one might not necessarily associate with what is commonly referred to as “management.” Based on the interviews with our project partners, the data visualizations conducted in our research group, and other critical approaches, I have identified three focus areas as a starting point for further examining RDM practices and their particularities in order to understand film historical metadata: (1) content and context; (2) data modeling and categorizations; and (3) access and reuse.

1 Content and Context

What is the subject area of the database? What time period does the database cover? What are the geographical foci? What is the primary goal? Where is the database affiliated? Who is involved in the work? Where does the funding come from?

RDM varies depending on the goal of an initiative, where it takes place, and the people involved. If databases are ideally preserved beyond a project’s lifetime, they become part of a digital research infrastructure. As such, they are often perceived as neutral and far removed from personal influence, although decisions are made by individuals who naturally have a particular point of view that shapes the goal of a project and how it is conducted. Thus, to understand the situatedness of databases and their particular contexts, we must also, as Julia Flan-

¹⁰ Sarah-Mai Dang, “The Women Film Pioneers Explorer. What Data Visualizations Can Tell Us about Women in Film History,” *Feminist Media Histories* 9, no. 1 (2023): 76–86. The interviews will be published once they have been edited.

ders states, take into account “digital humanities practitioners” – may it be scholars, designers, or archivists – as responsible parties.”¹¹

The Women Film Pioneers Project (WFPP) is an online platform for research on women in the silent film era, launched in 2013.¹² With more than three hundred career profiles, including filmographic and bibliographic information as well as a collection of film historical resources, its goal is to make women’s global work more visible and facilitate further investigation.¹³ It is run by co-founder Jane M. Gaines, who is based at Columbia University.¹⁴ Kate Saccone serves as the project manager, who also edits and curates the profiles.¹⁵ In addition, many people from various institutions have contributed to the website, both when it was planned as a printed encyclopedia and since it was converted to a digital format – graduate students, volunteers, web designers, external curators, and editors.¹⁶

In order to better understand the data collected and curated by the WFPP, the DAVIF research group conducted interviews with Jane M. Gaines and Kate Saccone about the genesis and approach of the initiative. We also conducted interviews with David Kleingers, who is head of the digital department and strategic development at DFF, and Bianca Sedmak, who manages the filmographic data editing there.

The filmportal.de was launched by the DFF – Deutsches Filminstitut & Filmmuseum in 2005.¹⁷ With information on more than 150,000 films and 250,000 people, according to their own statement, it is the most comprehensive publicly published filmography of Europe. Its goal is to document all film productions in Germany from its beginnings to the present day in order to facilitate both academic and popular inquiries. The cinematographic works documented on film-

¹¹ Julia Flanders, “Building Otherwise,” in *Bodies of Information. Intersectional Feminism and Digital Humanities*, ed. Elisabeth Losh and Jacqueline Wernimont (Minneapolis: University of Minnesota Press, 2018), 289–304.

¹² Jane M. Gaines, Radha Vatsal, and Monica Dall’Asta, *Women Film Pioneers Project* (New York: Columbia University Libraries, 2013), accessed May 2, 2022, <https://wfpp.columbia.edu/>.

¹³ Women Film Pioneers Project, “About the Project – Women Film Pioneers Project,” accessed May 3, 2022, <https://wfpp.columbia.edu/about/>.

¹⁴ See <https://wfpp.columbia.edu/>; on the historical context of the WFPP see also Kate Saccone’s chapter in this volume, “(Re)Visioning Women’s Film History: The Women Film Pioneers Project and Digital Curatorial-Editorial Labor.”

¹⁵ See Saccone, “(Re)Visioning Women’s Film History,” this volume. Saccone reflects on her role and the many different tasks her work as project manager involves.

¹⁶ Women Film Pioneers Project, “Editorial Team and Acknowledgments – Women Film Pioneers Project,” accessed February 17, 2023, <https://wfpp.columbia.edu/editorial-team-and-acknowledgments/>. See also Saccone, “(Re)Visioning Women’s Film History,” this volume.

¹⁷ Deutsches Filminstitut & Filmmuseum, “Filmportal.de. Alles Zum Deutschen Film,” accessed March 21, 2023, <https://www.filmportal.de/>.

portal.de have mostly been screened publicly as a theatrical release or at a festival, for example.

Relevant sources for new entries include the official release lists of the top organization of the film industry, Spitenorganisation der Filmwirtschaft (SPIO), decisions of the Voluntary Self-Regulation Body (FSK), and information from national and regional film funding bodies. Other important sources are the catalog publications of national as well as international festivals, on the basis of which new film work records are also created.¹⁸ The focus is on theatrical releases, but over the years more and more television films have been included as well. Although the majority of the films included in the database are German productions and co-productions, the collection does not exclusively focus on national cinematography. It also includes, as Sedmak pointed out in our interview, historical works the DFF considers relevant for film history, such as *BATTLESHIP POTEMKIN*.

The portal relies on permanent public funding by the Federal Government Commissioner for Culture and Media (Beauftragte der Bundesregierung für Kultur und Medien – BKM) and the Hessian Ministry for Science and Art (Hessische Ministerium für Wissenschaft und Kunst – HMWK). Additional funders are or were at some point the German Federal Film Board (Filmförderungsanstalt – FFA), the Federal Ministry of Economics and Technology (Bundesministerium für Wirtschaft und Technologie), the Initiative Culture & Creative Industries of the Federal Government (Initiative Kultur- und Kreativwirtschaft der Bundesregierung), and the Friedrich Wilhelm Murnau Foundation.¹⁹

While film scholars, historians, PhD students, film curators, and archivists contribute to the WFPP and there are different editors responsible for individual countries, the pioneer profiles are currently primarily edited and curated by film scholars Jane M. Gaines and Kate Saccone. The content of filmportal.de, on the other hand, is collected by special editorial teams who focus on filmographic data, text, and images. The scope of work is certainly large in both projects, but it is organized differently. Within the WFPP, there is a mixture of official positions and volunteer work. While the manager, director, and research assistant positions are official staff positions, the contributors' research and writing is conducted "on the side" like any other publishing labor. In contrast, for filmportal.de, all work is conducted as part of official positions at the DFF.

¹⁸ David Kleingers, interview by Sarah-Mai Dang, Pauline Junginger, and M. Leonie Biebricher, May 5, 2022.

¹⁹ David Kleingers, "Filmportal.de. Die zentrale Internetplattform zum deutschen Film," in *Handbuch Kulturportale: Online-Angebote aus Kultur und Wissenschaft*, ed. Ellen Euler et al. (Berlin, Munich, Boston: De Gruyter Saur, 2015), 204, <https://doi.org/10.1515/9783110405774-018>.

Although both initiatives aim to facilitate further research on film and film history, their goals differ in terms of theoretical, conceptual, and political aspects. The WFPP seeks to foster the visibility of women workers in early cinema and, in doing so, critically reflect the contingencies of film history and “reconfigure film knowledge,”²⁰ whereas the DFF intends to set standards for a systematic registration of film works.²¹ For this purpose, the DFF has created an internal filmographic database (Zentrale Filmographische Datenbank – ZDB), which is the core of filmportal.de. The ZDB was originally the product of a merger of the previous database by the DFF (former DIF) and CineGraph – Hamburgisches Centrum für Filmforschung e.V. Both initiatives collect film historiographical and personal data. However, while the WFPP focuses on individuals and their careers, the DFF pays greater attention to cinematographic works. This is important to keep in mind when searching for specific aspects of film history in these databases, as they reflect these foci in terms of completeness and nuance. For instance, the WFPP provides more details about women’s careers, whereas the DFF offers more comprehensive filmographies.

In principle, we must not forget that if a data collection is missing certain information, another may be able to provide it. As the Women Film Pioneers Explorer, a case study of our project, has shown, the United States is the country with the most WFPP profiles for the years 1895 to 1926, with 47.1% (163 entries).²² It is followed by Great Britain (9.2%, 32 entries), France (5.2%, 18 entries), Germany (4.3%, 15 entries), and Australia (4.0%, 14 entries). There is almost no data on women who worked in Africa, the Middle East, Central Asia, or India. As I have elaborated elsewhere, these figures do not mean that women in the United States were the most active in early cinema.²³ They rather provide information about the content of the WFPP database, and reveal something of its history. These results are understandable insofar as the project started in the United States and is based in New York, as is the main initiator, Jane M. Gaines. In addition, the creation of profiles, from submission to editing to the final posting of new entries, reflects the fact that it requires a great deal of time and coordination work. Yet, a huge gap becomes visible. It is therefore important to be aware of the situatedness of data.²⁴ One must not lightly assume that there is no information on a specific country anywhere or that persons or films did not exist there at

20 Women Film Pioneers Project, “About the Project.”

21 Kleingers, “Filmportal.de,” 206.

22 Henri Dickel et al., “Women Film Pioneers Explorer,” 2021, accessed March 21, 2023, <https://www.online.uni-marburg.de/women-film-pioneers-explorer/>.

23 Dang, “The Women Film Pioneers Explorer.”

24 Dang, “The Women Film Pioneers Explorer.”

all. Instead, we need to remind ourselves that no database can ever be complete. This might seem an obvious fact but can easily be forgotten. Databases are always the result of specific conditions in certain place and time constellations; they are part of particular RDM practices. Like scholarly publications, they represent only partial perspectives, which are nevertheless key to knowledge production.

2 Data Modeling and Categorizations

Which kind of work is involved in data modeling? How is the data organized and structured? What categories are considered relevant? How are political categories such as class, race, and gender addressed?

While it is important to have an interdisciplinary exchange about handling data, it is necessary to first understand RDM from within one's own discipline or in order to understand the implications of data-driven research. Looking at databases in the field of film history, it is clear that, even within a discipline, or even a particular area of a discipline, data can vary to a great extent.

A comparison of the WFPP and the DFF databases reveals the specific approaches of these initiatives. While the WFPP's data modeling can be defined as *research-driven* since it addresses a specific research interest, namely the study of women's work in early cinema, the DFF focuses primarily on the potential reuse of data and seeks to ensure interoperability through standards. Its approach can therefore be described as *curation-driven*.²⁵ This is certainly due to the institutional situatedness of the two initiatives. As Tim van der Heijden points out, because the WFPP is a research project that is institutionally embedded within a university and a university library, it maintains different data management protocols than the DFF. The latter is an institute within a museum that houses some comprehensive archives and collections of material on all aspects of film, and the *filmportal.de* website is only one of the DFF's many projects. Because of its funding and cultural policy position, the DFF works on an institutional level and is

²⁵ In the context of data modeling, a distinction is generally made between curation-driven and research-driven. However, these definitions cannot always be clearly separated from each other, as I have already pointed out. 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), 330–337; 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, New York: Routledge/Taylor & Francis, 2019), 26–94.

more committed to a wider applicability or reuse of data than the WFPP, which is driven primarily by project-specific research questions on a micro level.²⁶

As I have explained elsewhere, it was particularly interesting to learn during the interview with Kate Saccone that the diverse job titles collected in the database are the result of the research done by the authors.²⁷ As a result the database also contains job titles that one would not normally associate with film production, such as “society matron,” “metalworker,” or “carpenter.” A taxonomy emerged from the first set of pioneer profiles and the authors’ archival research, with final decisions made during conversations between the WFPP editors and the contributors.

As Saccone explains in this volume with regard to questions about standardization and interoperability, the taxonomy can be updated as needed.²⁸ She proof-reads and copy edits the profiles and inserts the biographical and occupational metadata suggested by the authors into the website’s content management system, WordPress.²⁹ It is important to note that Saccone’s continuous “(re)visioning,” as she conceptualizes it regarding editorial labor in the digital era, is not a linear process of RDM with clearly defined steps. Instead, it is complex work that involves, as she defines it, “digital curatorial processes of creation, presentation, preservation, and the ongoing management of digital (textual) materials and film historical knowledge.”³⁰ Even though digital tools and infrastructures offer great support and make workflows much easier, working with and on data is mostly manual work and does not happen automatically. RDM demands intensive, painstaking work that requires specific knowledge and technical skills in addition to a wide range of expertise and responsibilities, including making decisions, organizing tasks, communicating with colleagues and institutions, and monitoring processes. This is why the DFF has a permanent filmographic editorial staff of two full-time employees who deal primarily with the new entry of film works and the maintenance of the existing database.³¹ As Kleingers points out, data about a film changes frequently from the first official production announcement to the theatrical release, and its filmography is never complete. The editors are constantly returning to existing entries to add and correct information.³²

26 I thank Tim van der Heijden for these important remarks about an earlier version of this chapter.

27 Dang, “The Women Film Pioneers Explorer.”

28 Saccone, “(Re)Visioning Women’s Film History,” this volume.

29 The final terms are always chosen in collaboration with the authors and based on the existing taxonomy. For a detailed workflow see Saccone, “(Re)Visioning Women’s Film History,” this volume.

30 Saccone, “(Re)Visioning Women’s Film History,” this volume.

31 Kleingers, interview, May 5, 2022.

32 Kleingers, interview, May 5, 2022.

Based on the assumption that film history is itself characterized by coincidences, contradictions, and contingencies, the WFPP emphasizes epistemological uncertainties by allowing individual categorizations, even encouraging the authors to take their own approach.³³ This principle, which stems from theoretical and political considerations, reveals the numerous diverse professions that women held in early cinema, and thus their manifold impact in film history.³⁴ As it turns out, a decade after the launch of the WFPP, Jane M. Gaines contends that this approach has helped to identify an “incredible range of names and titles and types of occupations we never dreamed existed when we first began.”³⁵

Unlike the WFPP, the DFF models its data in an SQL database, a relational database with structured tables, according to the European standard EN15907. This standard defines a set of metadata to describe cinematographic works, including their variants and manifestations. EN15907 is not a data model itself but a scheme that offers a standardized approach for developing one.³⁶ As computer science students I work with on data visualization have noted, this standard includes a comprehensive terminology (although terms to be used in the DFF database are not specified) but, interestingly, no details on informational procedures are provided (e.g., that years must be numbers, or that names cannot be numbers). This can be a barrier to further data processing accuracy, as errors that could easily be prevented by technical specifications can creep into the database. This observation illustrates how differently people look at data, and how different backgrounds – in this case computer science and film historical and archival concepts and premises – determine priorities and goals in data processing.

As for the occupations, the DFF works with a terminology that is intended to be as reliable and pragmatic as possible. It is shaped by internal considerations, international film historiographical debates, and authority data of their project partners. The origin of the data is recorded as well as possible, as Bianca Sedmak explained in the interview.³⁷ All sources used are entered in a database, which can only be accessed via the internal ZDB user interface. If the referenced material is available in digital form, it is archived on an internal server. This includes press materials, scans of FSK cards, censorship cards, etc. The same procedure

³³ Women Film Pioneers Project, “Guidelines: Profiles – Women Film Pioneers Project,” accessed March 3, 2023, <https://wfpp.columbia.edu/guidelines-profiles/>.

³⁴ Dicke et al., “Women Film Pioneers Explorer.”

³⁵ Jane M. Gaines, interview by Pauline Junginger, February 5, 2023.

³⁶ Deutsches Filminstitut (DIF) e.V, “EN 15907,” filmstandards.org, accessed March 1, 2023. http://filmstandards.org/fsc/index.php?title=EN_15907.

³⁷ Bianca Sedmak, interview by Sarah-Mai Dang, M. Leonie Biebricher, and Pauline Junginger, May 19, 2022.

applies to correspondence such as e-mails or, in rare cases, letters.³⁸ Early film is documented in detail because, Sedmak explains, the DFF has incorporated all the encyclopedias, including censorship maps and secondary sources. Primary sources hardly still exist. It is also quite possible that some sources have just not been discovered yet, or even looked for. Sedmak points out that the DFF very much depends on encyclopedias and other research, and experts in the field of film history.³⁹

The DFF has several controlled topic-based vocabularies for their database to establish consistent spellings, as Bianca Sedmak explained.⁴⁰ Their occupation vocabulary is divided into two areas. One area includes the credits of film-related work – that is, the relation between the film and the person. A large proportion of these terms is derived from the most commonly used terms in movie credits. The other area includes person-related work, which is related to the GND (Ge-meinsame Normdatei) of the German National Library (DNB – Deutsche Nationalbibliothek). This vocabulary was adapted in the course of the collaboration with the GND. Since this vocabulary is synchronized and imported into the GND, it must be standardized and comply with the job descriptions of the GND. Both lists contain controlled vocabulary, so they cannot be manipulated manually. Each list includes around 140 to 150 job titles. The titles in the list of film-related works are in turn divided into 16 top categories. These categories are decisive for the occupation that is displayed on the person page at filmportal.de.

The DFF is far from being able to map every occupation in personal records, Sedmak notes.⁴¹ With regard to interoperability with the German National Library, it is also not necessary to represent every profession in the database. For reasons of clarity, the terms are therefore limited to the most common ones.⁴² For “model,” for example, there is no term. It is resolved by using “actress” (“Darstellerin”) for the occupation if the person did some acting, or “participation” (“Mitwirkung”) if it is a documentary film about a model. The DFF seeks to describe the actual work of a person as precisely as possible, which is why non-film-related job titles such as “politician” or fields of work such as “medicine” can also be found among the person-related occupations.⁴³

³⁸ Sedmak, interview, May 19, 2022.

³⁹ Sedmak, interview, May 19, 2022.

⁴⁰ Sedmak, interview, May 19, 2022.

⁴¹ Sedmak, interview, May 19, 2022.

⁴² I thank Bianca Sedmak for the additional insights provided in her feedback on an earlier version of this chapter.

⁴³ Sedmak, interview, May 19, 2022.

The modeling of job titles is quite understandable, although not uncomplicated. When it comes to gender classifications (e.g., “female,” “male,” “non-binary”), however, it not only gets complicated, but also messy. The DFF deals with these attributions in different ways. Depending on the specific table of the database, whether it is person-related or film-related, internal, or published data, the DFF uses gendered job titles or descriptions that refer to film divisions rather than job titles. For instance, they use “Szenenbild” (scenography) rather than “Szenendesigner:in” (which refers to a scene designer who is female, male, or non-binary). For some professions, however, both variants exist, for example, “Darsteller” (male actor) and “Darstellung” (acting). While the job titles in personal records are predominantly gendered, film-related categories that describe how a person is related to a film are sometimes gendered, and then based on the masculine form. To account for historical and cultural developments, multiple variants, in turn, have been retained, like “Szenenbild” (scenography) and “Bauten” (buildings). These terms appear synonymous but refer to different professions. Sedmak explained in the interview that some distinctions or nuances are not published on filmportal.de, but documented in the internal ZDB.⁴⁴ The different handling of gender attributions is remarkable but in a way also understandable. Gendered categorizations are sensitive and pose a political conceptual, and technical challenge.

It becomes apparent that standardizations cannot provide all-encompassing solutions for linking data and ensuring sustainable reuse, let alone comprehensive documentation of film historical aspects. Moreover, the question arises as to what is actually involved in the demand for standards. Which data or procedures should be standardized, and for what purpose? What is lost with standardization; what is gained?⁴⁵ While the implementation of standard metadata can foster interoperability and collaboration, it might also reinforce blind spots. Data can help make particular histories visible, but might also obscure certain aspects of the past.⁴⁶ The issue of standards is relevant not only in terms of job titles, but also in terms of gender. For example, if a record of film historical occupations does not have a gender assigned to it, either for the job title or the linked person,

⁴⁴ The current data model of the ZDB is described in a semi-public Wiki, where the DFF also provides information on the controlled vocabulary of occupations. Deutsches Filminstitut (DIF) e.V, “Tätigkeiten – DIF Filmographie Wiki,” filmstandards.org/difzf/index.php?title=Tätigkeiten.

⁴⁵ Dang, “O.J.”

⁴⁶ Sarah-Mai Dang, “Representing the Unknown: A Critical Approach to Digital Data Visualizations in the Context of Feminist Film Historiography,” in *How Film Histories Were Made: Material, Methods, Discourses*, ed. Malte Hagener and Yvonne Zimmermann (Amsterdam: Amsterdam University Press, 2023), 467–493.

we cannot analyze, for instance, how many women or men or non-binary people have directed a film, or how this might have changed over time. Standards are crucial not only in terms of data modeling, but also in terms of what data we include or exclude. This raises the question of what information we consider key in the first place – and what we wish to be able to be retrieved in the course of further research.

Since the WFPP features women workers, it does not include male pioneers. The website understandably does not provide an option to search for gender, but nor does filmportal.de. However, the DFF's internal database can be searched by gender. The ZDB contains three gender categories: "female," "male," and "undefined," where "undefined" can refer to both "non-binary" and "unknown": a fourth category would actually be necessary to distinguish between the two. "Unknown" is usually employed by the GND in the sense of "undetermined" and not in the sense of "non-binary."⁴⁷ As Sedmak stated in the interview, gender categorization is important to the DFF and internal annotations are made in the ZDB when uncertainties or new findings arise. However, the representation of gender is rather secondary for filmportal.de, where the personal data pages do not say male or female, but only "Darsteller" (actor). The DFF team has been working on this for a long time and are strongly advocating that wording should be more accurate. But that would be an extensive undertaking that requires many resources.

As for databases that explicitly focus on gender representation, in my view the BFI Filmography (2012–2017)⁴⁸ can serve as good practice example. In a detailed documentation of their data modeling, which was accessible together with the data online, the curators describe which gender attributions they have made, how, and for what reason. In doing so, they reflect on the advantages and disadvantages of their approach. They rightly point out that, while their method is not perfect as binary categories and external gender attributions leave out nuances, the focus on gender can still foster further discussions about equality in film industries.⁴⁹

Of course, not all the implications of categorization can always be considered, and it cannot be the goal to create a perfect database because, as already stated, there is no such thing, but we need to situate and, in doing so, critically reflect on the underlying processes of databases. Categorizations affect how we evaluate a person's role in history. Assigning specific metadata to people or objects is a powerful act of meaning-making.⁵⁰ By ignoring gender or other political categories, or

⁴⁷ Sedmak, interview, May 19, 2022.

⁴⁸ See <https://filmography.bfi.org.uk/>, accessed March 8, 2023.

⁴⁹ See <https://filmography.bfi.org.uk/>, accessed March 8, 2023.

⁵⁰ Miriam Posner, "What's Next. The Radical, Unrealized Potential of Digital Humanities," in *Debates in the Digital Humanities*, ed. Matthew K. Gold and Lauren F. Klein (Minneapolis, London:

by focusing exclusively on one individual, the director of a film, for instance, numerous people and facets of film history are neglected.⁵¹ We also must pay attention, as Flanders argues, to what function attributions have, how they are used, and how they are conceived, whether they are applied as fact-stating descriptions that perpetuate a form of othering, or as a “category of discovery.”⁵²

3 Access and Reuse

Where is the data stored? How can a database be accessed? How is the interface designed? How can the data be retrieved? How is the data licensed? Who uses the database? How can users contribute?

As research practices change in the wake of digitalization, research results beyond the classical publication become more and more relevant.⁵³ Whereas five years ago Adelheid Heftberger and Marion Goller were critical of the very low interest in open access and reuse of research data in film studies, this has fortunately changed considerably in recent times.⁵⁴ As open science activists rightly argue, sharing research data of any kind and making it reusable without technical, financial, or legal barriers is the prerequisite for fruitful, effective, and inclusive knowledge production.⁵⁵ Reuse logically requires findability, accessibility, and interoperability, which is why the FAIR data principles – findable, accessible, interoperable, reusable – were created.⁵⁶ Since data reflect political, cultural, and social conditions, and thus the distribution of power, the Global Indigenous Data Alliance (GIDA) has developed the CARE principles (collective benefit, authority

University of Minnesota Press, 2016), accessed March 21, 2023, <https://dhdebates.gc.cuny.edu/read/untitled/section/a22aca14-0eb0-4cc6-a622-6fee9428a357>.

⁵¹ Dang, “Representing the Unknown,” 469.

⁵² Flanders, “Building Otherwise,” 296–298.

⁵³ Dang, “Forschungsdatenmanagement in der Filmwissenschaft,” 128–130.

⁵⁴ Marion Goller and Adelheid Heftberger, “Die Öffnung von Forschungsdaten in den Film- und Medienwissenschaften: praktische und urheberrechtliche Herausforderungen,” *<intr>2Dok [S] – Fachinformationsdienst für internationale und interdisziplinäre Rechtsforschung* (2018), accessed March 21, 2023, <https://doi.org/10.17176/20180515-233758>.

⁵⁵ Heinz Pampel and Sünje Dallmeier-Tiessen, “Open Research Data: From Vision to Practice,” in *Opening Science: The Evolving Guide on How the Internet Is Changing Research, Collaboration and Scholarly Publishing*, ed. Sönke Bartling and Sascha Friesike (Cham: Springer International Publishing, 2014), 213–224, accessed March 21, 2023, https://doi.org/10.1007/978-3-319-00026-8_14.

⁵⁶ FORCE11, “The FAIR Data Principles,” accessed March 21, 2023, <https://www.force11.org/group/fairgroup/fairprinciples>.

to control, responsibility, ethics) to complement the FAIR data principles by drawing attention to the power imbalances and historical contexts that they believe are being ignored by the current open data movement.⁵⁷ Who benefits from data-driven projects and who does not is an important question that applies not only to multinational technology companies but also to academia. As Catherine D'Ignazio and Lauren Klein argue, “asking these *who questions* allows us, as data scientists ourselves, to start to see how privilege is baked into our data practices and our data products.”⁵⁸

WFPP is regularly archived in the Wayback Machine, and pioneer profiles are deposited as PDF files, tagged with a DOI (digital object identifier), in Columbia’s digital repository, Academic Commons. Although it is not clear on the website, the content created by the WFPP – not the images or films of third parties – has a Creative Commons Attribution license, according to Saccone. The biographical dataset I requested for the Women Film Pioneers Explorer case study is now archived on Academic Commons as well.⁵⁹ All the information on WFPP is freely accessible and reusable. This is somewhat different from the DFF. As already mentioned, filmportal.de is an output of the ZDB. It is detached from the DFF’s central database and operates via a content management system programmed for its specific requirements, based on Drupal.⁶⁰ This means that the ZDB and filmportal.de contain slightly different information. Since the DFF is our project partner, the DAVIF research group has the privilege of accessing both datasets. Many of the contents and objects presented online on filmportal.de are protected by copyright and exploitation rights and exclusively for non-commercial use. While the portal is a public platform, the ZDB is for internal use or research projects only.⁶¹ In addition, there is a web service for personal and corporate data such as

⁵⁷ Global Indigenous Data Alliance, “CARE Principles,” accessed January 23, 2023, <https://www.gida-global.org/care>.

⁵⁸ Catherine D'Ignazio and Lauren F. Klein, *Data Feminism* (Cambridge, MA: MIT Press, 2020), 26.

⁵⁹ Jane M. Gaines and Columbia University Libraries, “Women Film Pioneers Project Biographical Data,” dataset compiled December 7, 2020, <https://doi.org/10.7916/m4dc-n768>.

⁶⁰ Kleingers, “Filmportal.de,” 212.

⁶¹ The filmographic data of the ZDB can be exported as XML data via an OAI-PMH interface. As Kleingers explained, this is used, for example, by the partners of the DFF, like the federal archive (Bundesarchiv) or the Goethe-Institut. The web service (another interface) is used by the German Film Academy (Deutsche Filmakademie), among others. Access to these interfaces is not generally available to the public, but has to be granted by the DFF. The reuse of the data obtained via this interface is regulated in respective contracts. The website filmportal.de has another interface for the export of object-related metadata (i.e. information on digital photos, videos, texts, etc., that are held in the portal). The data provided via this interface are used by the German Digital

film distribution or production companies.⁶² New entries of personal information go directly into the GND and the GND links to the corresponding information on filmportal.de (and vice versa).⁶³ Backup copies are made regularly by the DFF itself.

Legal issues concerning immaterial goods, their ownership, and conditions of use, are usually highly complex and can only be decided on a case-by-case basis. Thus, the question of which research data can be shared and used is not easy to answer in principle.⁶⁴ In this regard, however, I was surprised to learn from a legal consultant that as long as the data obtained is not used for commercial but only for scholarly purposes, web scraping is legal. That is, as long as one does not want to build a competing product to IMDb, for instance, it is legal to download data from the website. In dealing with data usage agreements (Nutzungsvereinbarungen), in which my research assistant Pauline Junginger and I have invested considerable time, I have learned that the threshold of originality (Schöpfungshöhe) of datasets can be reached relatively quickly. On the other hand, however, if data are only facts, they are not protected by copyright. But if someone or an institution has invested a lot of time and money in a database, ancillary copyright (Leistungsschutzrecht) takes effect. Again, as always with legal issues, it depends on the particular case. It is also important to distinguish between data types, metadata, films, texts, or other – digitized or born digital – objects we are dealing with. The legal uncertainties that still prevail in this context may in any case prevent the free use of research data and the advancement of digital scholarship. Data is resource-intensive and thus expensive.⁶⁵ This is another reason why, in addition to the promotion of open data, we need to further familiarize ourselves with the legal framework.

Filmportal.de is used for general and scholarly inquiries. In addition to film historians, filmmakers are increasingly asking for information about their work to be updated. Thus the website serves, as Kleingers explains, both as a source of information and as a presentation platform.⁶⁶ With regard to the accessibility of the data, it should be noted that the implementation of English will be further

Library, the European Film Gateway, and the Europeana, among others. I thank David Kleingers for the additional insights provided in his feedback on an earlier version of this chapter.

62 Kleingers, “Filmportal.de,” 212.

63 Kleingers, “Filmportal.de,” 213.

64 Linda Kuschel, “Wem ‘gehören’ Forschungsdaten?” *Forschung & Lehre*, September 12, 2018, accessed March 21, 2023, <https://www.forschung-und-lehre.de/forschung/wem-gehoeren-forschungsdaten-1013/>.

65 D’Ignazio and Klein, *Data Feminism*, 41–47.

66 Kleingers, “Filmportal.de,” 207.

developed. To date, the interface allows at least a rudimentary search with English terminology. For reasons of resource, the WFPP's content is currently only available in English. Like the DFF, the WFPP serves different interests and users. Remarkably, apart from film historical research, it is also used as a genealogical source. In the interview, Saccone reported that many family members reach out to WFPP. Once a family member asked the editorial team to rewrite a profile of a woman because in their opinion she was only secondarily a filmmaker. In conversation with the author, however, who had done extensive archival research, they decided to continue to represent this woman who made one film and perhaps did not see herself as a filmmaker, and the profile has remained. This was an ethical decision to increase the visibility around such women.⁶⁷

In principle, everyone can contribute to the WFPP. In doing so, authors must adhere to the standards of the project, which are ensured by peer review. Although these specifications are not of a technical nature in the sense that one might associate with RDM, they are part of a standardized approach that has been developed for the project and is continuously adjusted.

As other film and media scholars have shown and I have argued elsewhere, while film historical databases such as the ZDB/filmportal.de or the WFPP provide valuable resources for film historical research, it is important to further explore digital technologies for creative reuse of existing data.⁶⁸ For instance, as already indicated, data visualizations can open up new perspectives on data and facilitate critical reflection on historical sources. The way data is presented has a major impact on how we can access, explore, and reuse it.⁶⁹ By means of scaling, for instance, data visualizations can provide orientation, as Deb Verhoeven claims. Due to the size of big datasets, anomalies or absences do not become apparent unless experiments are conducted to examine them.⁷⁰

⁶⁷ Kate Saccone, interview by Sarah-Mai Dang, M. Leonie Biebricher, and Pauline Junginger, March 25, 2022.

⁶⁸ Dang, "Representing the Unknown."

⁶⁹ 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), accessed March 21, 2023, <http://www.digitalhumanities.org/dhq/vol/14/4/000528/000528.html>.

⁷⁰ Deb Verhoeven, "Show Me the History! Big Data Goes to the Movies," in *The Arclight Guidebook to Media History and the Digital Humanities*, ed. Charles R. Acland and Eric Hoyt (Folmer: REFRAME Books, 2016), 172, accessed March 21, 2023, <https://projectarclight.org/wp-content/uploads/ArclightGuidebook.pdf>.

Conclusion

Since RDM is of great importance to our studies, as I hope to have shown, and this importance will grow rapidly, we should not regard it as an annoying necessity that has to be addressed in the next research proposal, but as an opportunity to help shape critical digital scholarship. RDM should be considered an integral part of our research as data becomes more and more important – not only to film and media studies, but to other disciplines as well. It is therefore crucial for researchers to reveal how the data they apply has been shaped. This is necessary to acknowledge the limitations and messiness of one's own methods.

While new approaches and practices can offer exciting perspectives on film culture and its past, additional expertise and strategies are needed. Data-driven approaches require a deep understanding of the underlying data, how it was created, by whom, and for what purpose. Up to now, a distinction has generally been made between technical and content-related work in data-driven projects. However, in view of the multi-layered entanglements, a closer collaboration between the two areas is required in order to do justice to the numerous aspects of RDM.

Needless to say, the comparison of the WFPP and the DFF is not conclusive. In addition to providing insights into particular RDM practices, their premises and implications, it should, first and foremost, serve as a starting point for further inquiries of other initiatives, not least our own RDM approaches and data-driven research projects. A closer look at how film historical metadata has been collected, selected, modeled, organized, preserved, and shared, has shown that data is shaped by a complex interplay between people, institutions, and infrastructures, as well as practical, technological, and theoretical premises.

As demonstrated, data may vary not only in quality but also in type and model due to different assumptions and goals. This makes it difficult to map data across projects. Since each of the WFPP and DFF databases was developed independently for specific purposes and in different languages, it seems impossible to link these two. However, comparing heterogeneous data, as in this case, can provide many insights into the conditions of historiographic knowledge production. It helps in bringing initiatives together – not necessarily on a technical level, but certainly in terms of film historiographical questions – and to further stimulate critical data discourses.

To understand the RDM of the DFF and the WFPP, the semi-structured interviews were extremely helpful. In addition to discoveries made through the data visualizations, many insights were gained through the conversations with the individuals involved in the work. The exchange has intensified the cooperation with our project partners, but, ideally, all projects should document their data practices. We can learn from other disciplines such as ethnography and social sci-

ences about documentation protocols. For instance, Van der Heijden and Kolkowski took inspiration from history of science in how they document experiments in laboratory settings.⁷¹ Library science could also be helpful: it might be interesting to look more closely at the history and infrastructure of libraries in relation to RDM within academic research practices, not least because of the increasing collaboration between researchers and librarians in the storage, access, and re-use of data in data-driven projects. Systematic documentation through data papers, for example, which is also becoming more widely discussed in film and media studies,⁷² would not only help one to better understand one's own premises and practices, but also open up scholarship. By situating relevant datasets like we do with scholarly publications, we can strengthen data-driven projects and further advance digital film history.

References

Dang, Sarah-Mai. "Representing the Unknown: A Critical Approach to Digital Data Visualizations in the Context of Feminist Film Historiography." In *How Film Histories Were Made: Material, Methods, Discourses*, edited by Malte Hagener and Yvonne Zimmermann, 467–493. Amsterdam: Amsterdam University Press, 2023.

Dang, Sarah-Mai. "The Women Film Pioneers Explorer. What Data Visualizations Can Tell Us about Women in Film History." *Feminist Media Histories* 9, no. 1 (2023).

Dang, Sarah-Mai. "O.J. – Recherchepraktiken, Datenquellen und Modellierungen." In *Doing Research. Wissenschaftspraktiken Zwischen Positionierung und Suchanfrage*, edited by Sandra Hofhues and Konstanze Schütze, 330–337. Bielefeld: Transcript, 2022.

Dang, Sarah-Mai. "Forschungsdatenmanagement in der Filmwissenschaft: Daten, Praktiken und Erkenntnisprozesse." *montage AV Zeitschrift für Theorie und Geschichte audiovisueller Kommunikation* 29, no. 1 (2020): 119–140.

Dang, Sarah-Mai. "Unknowable Facts and Digital Databases: Reflections on the Women Film Pioneers Project and Women in Film History." *Digital Humanities Quarterly* 14, no. 4 (2020). Accessed March 21, 2023. <http://www.digitalhumanities.org/dhq/vol/14/4/000528/000528.html>.

Deutsches Filminstitut (DIF) e.V. "Tätigkeiten – DIF Filmographie Wiki." [filmstandards.org](http://filmstandards.org/difzf/index.php?title=Tätigkeiten). Accessed March 8, 2023. <https://filmstandards.org/difzf/index.php?title=Tätigkeiten>.

⁷¹ Tim van der Heijden and Aleksander Kolkowski, *Doing Experimental Media Archaeology: Practice* (Berlin, Boston: De Gruyter Oldenbourg, 2023), <https://doi.org/10.1515/9783110799767>.

⁷² Data papers in journals like *Research Data Journal for the Humanities* or *Social Science* or *Journal of Open Humanities Data* are very intriguing and offer important insights into RDM; see, for example, Thunnis van Oort and Julia Noordegraaf, "Structured Data for Performing Arts History: An Introduction to a Special Issue of Data Papers: Arts and Media," *Research Data Journal for the Humanities and Social Sciences* 5, no. 2 (2020): 1–12. <https://doi.org/10.1163/24523666-bja10008>.

Deutsches Filminstitut (DIF) e.V. “EN 15907.” [filmstandards.org](https://filmstandards.org/fsc/index.php?title=EN_15907). Accessed March 1, 2023.

https://filmstandards.org/fsc/index.php?title=EN_15907.

Deutsches Filminstitut & Filmmuseum. “Filmportal.de – Alles Zum Deutschen Film.” Filmportal. Accessed July 15, 2023. <https://www.filmportal.de/>.

Dickel, Henri, Matija Miskovic, Kharazm Noori, Christian Schmidt, Atefeh Soltanifard, Sarah-Mai Dang, and Thorsten Thormählen. “Women Film Pioneers Explorer.” Accessed March 21, 2023.

<https://www.online.uni-marburg.de/women-film-pioneers-explorer/>.

D’Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. Cambridge, MA: MIT Press, 2020.

Einwächter, Sophie G. “Forschungsdaten (in) der Film- und Medienwissenschaft – Sophie G. Einwächter über vorurteilsbehaftete Begrifflichkeiten und fruchtbare Momente in der Lehre.” *Open Media Studies Blog* (2019). Accessed March 21, 2023. <https://mediastudies.hypotheses.org/1314>.

Fickers, Andreas, Juliane Tatarinov, and Tim van der Heijden. “Digital History and Hermeneutics – Between Theory and Practice: An Introduction.” In *Digital History and Hermeneutics: Between Theory and Practice*, edited by Andreas Fickers and Juliane Tatarinov, 1–20. Berlin, Boston: De Gruyter Oldenbourg, 2022.

Flanders, Julia. “Building Otherwise.” In *Bodies of Information. Intersectional Feminism and Digital Humanities*, edited by Elizabeth Losh and Jacqueline Wernimont, 289–304. Minneapolis: University of Minnesota Press, 2018.

FORCE11. “The FAIR Data Principles.” Accessed March 21, 2023. <https://www.force11.org/group/fair/group/fairprinciples>.

Forschungsdaten.info. “English Pages | Forschungsdaten Und Forschungsdatenmanagement.” Accessed February 3, 2023. <https://forschungsdaten.info/english-pages/>.

Gaines, Jane M., and Columbia University Libraries. “Women Film Pioneers Project Biographical Data.” Dataset compiled December 7, 2020. <https://doi.org/10.7916/m4dc-n768>.

Gaines, Jane M., Radha Vatsal, and Monica Dall’Asta. *Women Film Pioneers Project*. New York: Columbia University Libraries, 2013. Accessed May 2, 2022. <https://wfpp.columbia.edu/>.

Gitelman Lisa, and Virginia Jackson. “Introduction.” In “Raw Data” Is an Oxymoron, edited by Lisa Gitelman, 1–14. Cambridge, MA: MIT Press, 2013.

Global Indigenous Data Alliance. “CARE Principles.” Accessed January 23, 2023. <https://www.gida-global.org/care>.

Goller, Marion, and Adelheid Heftberger. “Die Öffnung von Forschungsdaten in den Film- und Medienwissenschaften: praktische und urheberrechtliche Herausforderungen.” *<int>2Dok [5] – Fachinformationsdienst für internationale und interdisziplinäre Rechtsforschung* (2018). Accessed March 21, 2023. <https://doi.org/10.17176/20180515-233758>.

Hoekstra, Rik, and Marijn Koolen. “Data Scopes for Digital History Research.” *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 52, no. 2 (2019): 79–94. <https://doi.org/10.1080/01615440.2018.1484676>.

Jannidis, Fotis, and Julia Flanders. “A Gentle Introduction to Data Modeling.” In *The Shape of Data in the Digital Humanities: Modeling Texts and Text-Based Resources*, edited by Julia Flanders and Fotis Jannidis, 26–94. New York: Routledge/Taylor & Francis, 2019.

Kleingers, David. “Filmportal.de – Die zentrale Internetplattform zum deutschen Film.” In *Handbuch Kulturportale: Online-Angebote aus Kultur und Wissenschaft*, edited by Ellen Euler, Monika Hagedorn-Saupe, Gerald Maier, Werner Schweibenz, and Jörn Sieglerschmidt, 204–214. Berlin, Munich, Boston: De Gruyter Saur, 2015. <https://doi.org/10.1515/9783110405774-01>.

Kuschel, Linda. "Wem 'gehören' Forschungsdaten?" *Forschung & Lehre*, September 12, 2018. Accessed March 21, 2023. <https://www.forschung-und-lehre.de/forschung/wem-gehoeren-forschungsdaten-1013/>.

Noordegraaf, Julia, Kathleen Lotze, and Jaap Boter. "Writing Cinema Histories with Digital Databases: The Case of Cinema Context." *TMG Journal for Media History* 21, no. 2 (2018): 106–126. <https://doi.org/10.18146/2213-7653.2018.369>.

Oort, Thunnis van, and Julia Noordegraaf. "Structured Data for Performing Arts History: An Introduction to a Special Issue of Data Papers: Arts and Media." *Research Data Journal for the Humanities and Social Sciences* 5, no. 2 (2020): 1–12. <https://doi.org/10.1163/24523666-bja10008>.

Pampel, Heinz, and Sünje Dallmeier-Tiessen. "Open Research Data: From Vision to Practice." In *Opening Science: The Evolving Guide on How the Internet Is Changing Research, Collaboration and Scholarly Publishing*, edited by Sönke Bartling and Sascha Friesike, 213–224. Cham: Springer International Publishing, 2014. https://doi.org/10.1007/978-3-319-00026-8_14.

Posner, Miriam. "What's Next: The Radical, Unrealized Potential of Digital Humanities." In *Debates in the Digital Humanities*, edited by Matthew K. Gold and Lauren F. Klein. Minneapolis: University of Minnesota Press, 2016. Accessed March 21, 2023. <https://dhdebates.gc.cuny.edu/read/untitled/section/a22aca14-0eb0-4cc6-a622-6fee9428a357>.

Van der Heijden, Tim, and Aleksander Kolkowski. *Doing Experimental Media Archaeology: Practice*. Berlin, Boston: De Gruyter Oldenbourg, 2023. <https://doi.org/10.1515/9783110799767>.

Verhoeven, Deb. "Show Me the History! Big Data Goes to the Movies." In *The Arclight Guidebook to Media History and the Digital Humanities*, edited by Charles R. Acland and Eric Hoyt, 165–183. Falmer: REFRAME Books, 2016. Accessed March 21, 2023. <https://projectarclight.org/wp-content/uploads/ArclightGuidebook.pdf>.

Women Film Pioneers Project. "About the Project – Women Film Pioneers Project." Accessed May 3, 2022. <https://wfpp.columbia.edu/about/>.

Women Film Pioneers Project. "Editorial Team and Acknowledgments – Women Film Pioneers Project." Accessed February 17, 2023. <https://wfpp.columbia.edu/editorial-team-and-acknowledgments/>.

Women Film Pioneers Project. "Guidelines: Profiles – Women Film Pioneers Project." Accessed March 3, 2023. <https://wfpp.columbia.edu/guidelines-profiles/>.

Filmography

BATTLESHIP POTEMKIN, dir. Sergei Eisenstein. Soviet Union, 1925.

