

Caring for Audiovisual Data - Pre-Constituted Panel Proposal

Chair: Christian Gosvig Olesen (University of Amsterdam)

Panel Description

The panel “Caring for Audiovisual Data” zooms in on the challenges of curating audiovisual research data as well as film archival metadata to address issues of data privacy, datafication of historical audiovisual media, and data standards for archives. Drawing on perspectives from scholarly and archival projects, the panel’s papers collectively depart from the observation that datafication of audiovisual media and archives and the emergence of digital scholarship, raise significant ethical challenges for analyzing, managing and caring for audiovisual data. While principles for data sharing and management such as FAIR (Findable, Accessible, Interoperable, Reusable) and CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) are becoming increasingly vital in developing care-based perspectives, a fundamental reflection still needs to take place before such initiatives can be integrated into scholarly and archival practice. Contributing to this endeavor from different theoretical angles, the panel critically elucidates how power structures and knowledge regimes are embedded in different types of data – from the legacy data of digitized collections to scholarly research data, and data pertaining to personal video streaming behaviors – to articulate new workflows and approaches. On the one hand, this entails focusing on the processes of collecting datasets from one's own Netflix viewing history to take ownership of personal data, to analyzing patterns embedded in an entire film department's video library catalogue. On the other hand, the panel highlights approaches that criticize digital tools and methods for altering audiovisual (meta)data as well as their embedded histories through common practices such as classifying, cleaning or upscaling of data.

Abstracts

Abstract 1

Emil Stjernholm (Lund University): *The Curation of Audiovisual Data, Artificial Intelligence and the Contested Practice of Upscaling*

Attending to the digital humanities’ visual turn, which foregrounds audiovisual media as data utilizing the increasing abilities of machines and computer vision methods to analyze images, video, and audio in new ways (Fickers, Snickars & Williams 2018; Burghart et al. 2020), this paper zooms in on the contested

practice of *upscaling*, and its epistemological, methodological and aesthetic challenges for digital film and media studies. In 2020, a so-called “AI-enhanced” version of *Arrival of a Train at La Ciotat* (The Lumière Brothers, 1896) went viral, showcasing a modified version of the film which was upscaled to 4K digital resolution and transformed from 16 frames per second to 60 frames per second. Denis Shiryayev’s “restorations”, made using machine learning algorithms, has sparked criticism. Whereas film theorist Antonio Somaini (2021) argues that upscaling “may alter significantly our experience of visual documents of the past”, archivist and scholar Rick Prelinger (2021) cautions that “prettified AI-enhanced digital objects” might “push the original, less attractive evidence out of view”. AI enthusiasts *care* for neural networks and their capabilities could thus also imply *carelessness* with regard to the treatment of media historical artefacts. Today, film and media scholars working with digital methods engage in everyday self-curatorial practices relating to audiovisual datasets. Such practices include but are not limited to the organization and curation datasets (e.g. the use of computational tools to prepare and study the data), the process of data cleaning (e.g. the removal of noise in audio or audiovisual data), or transformation of the data (e.g. altering the format, frame rate or definition of videos). Accordingly, drawing on my experiences working with audiovisual media as data within two ongoing research projects, this paper focuses on the following question: what are the pitfalls and affordances related to *upscaling* as a low-level curatorial practice in the digital humanities?

References

- Burghardt, M. et al. (2020), “Digital humanities & film: Analyzing the modalities of moving images”, *Digital Humanities Quarterly*, vol. 14, no. 4.
- Fickers, A., Snickars P. & Williams M. (2018), “Editorial special issue: Audiovisual data in digital humanities”, *VIEW Journal of European Television History and Culture*, vol. 7, no. 14.
- Prelinger, R. (2021), “NFTs and AI Are Unsettling the Very Concept of History”, *Wired*, 20 April.
- Somaini, A. (2021), “Film, media and visual culture studies, and the challenge of machine learning”, *NECSUS European Journal of Media Studies*, vol. 10, no. 2.

Abstract 2

Isadora Campregher Paiva (Goethe University Frankfurt): *Taking care of our own data: reclaiming data ownership through analysis in the streaming age*

In the age of “surveillance Capitalism” (Zuboff, 2019), the business model of media companies has increasingly shifted towards harvesting vast amounts of user-data in order to target consumers with productions perfectly matched to their preferences. As users, we have become accustomed to handing over our data to such companies, often impatiently accepting terms of service without reading them. While the gap between technological capabilities and governmental regulations is largely in favor of companies, the European Union’s General Data Protection Regulation (GDPR) did bring about a change in policies, opening some room for resistance. As a result of these changes, for instance, it is now possible for Netflix users to request access to the data the company has gathered about them. Still, with limited data literacy and resources, this has no impact on most users, who are unaware of this possibility and would not know what to do with the data even if it was easily accessible. What can we do as film and media scholars to empower ourselves – and potentially others – to take ownership of our own data in meaningful ways? In this presentation I will share my process of analysing my own Netflix data and creating my own recommendation algorithm using R code. I will argue that programming is more suitable to the application of FAIR (Findable, Accessible, Interoperable, Reusable) and CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) principles than point-and-click tools due to its transparency, reproducibility and open accessibility. In keeping with these principles, I will go over how my script can be used by others to gain insight into their own data.

References

- GO FAIR. “FAIR Principles.” Accessed January 25, 2023. <https://www.go-fair.org/fair-principles/>.
- Global Indigenous Data Alliance. “CARE Principles.” Accessed January 25, 2023. <https://www.gida-global.org/care>.
- Zuboff, Shoshana. (2019), "The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power". New York: Public Affairs.

Abstract 3

Josephine Diecke (Philipps-Universität Marburg) and Simon Spiegel (University of Zurich): *More Varda than Godard: Exploring Audiovisual Collections through Metadata*

Curating audiovisual collections not only concerns the maintenance and care of actual artifacts, but also the associated metadata. This challenge is currently being taken up by the Department of Film Studies at the University of Zurich (FIWI). Dating back to its founding in 1989, FIWI has built up a film library

holding more than 55'000 audio-visual works on DVD and Blu-ray. The associated FileMaker database comprises rich filmographic data as well as keywords from a defined glossary for all films in the collection. While the current solution has worked reliably for more than three decades, it is now undeniably outdated. All information is stored in a "flat" database, and the individual records are barely linked to each other. In addition, no standardized identifiers are used, and there is no support for standardized web interfaces. Tackling this situation, preparations are currently under way to evaluate the textual metadata and to make it available following FAIR principles.

This paper focuses on specific steps that will be taken to curate and evaluate the collection, including aspects of data quality, data enrichment and linking with standardized identifiers, as well as initial evaluations of the metadata in regards to distributions of the collected audiovisual works. An important aspect will be the integration of common standards and ontologies such as the *Functional Requirements for Bibliographic Records* and *EN15907*. This raises questions about the future design of the database and the possibilities for research and teaching such as "which works, persons, or genres are particularly often or rarely represented?" or "what does the number of manifestations tell us about the work but also about the previous research foci at FIWI?". These and other questions will be addressed in the paper to lay ground for further exchange on the topic within the community.

References

- Fairbairn, Natasha; Pimpinelli, Maria Assunta; Ross, Thelma (2016), *The FIAF Moving Image Cataloguing Manual*. In: Linda Tadic (ed.). Bloomington: Indiana University Press.
- IFLA Study Group on the Functional Requirements for Bibliographic Records (2009): "Functional Requirements for Bibliographic Records: Final Report". IFLA Repository. Accessed January 27, 2023. <https://repository.ifla.org/handle/123456789/811>.
- Wilkinson, Mark D. et al. (2016): "The FAIR Guiding Principles for Scientific Data Management and Stewardship". In: *Scientific Data* 3.1, 1–8.

Abstract 4

Sarah-Mai Dang (Philipps-Universität Marburg): *The Power of Data: Why Digital Film History Needs Data Feminism*

There is no such thing as „raw data“ – data are produced by specific people for specific reasons (Gitelman and Jackson 2013). Data are therefore always already an interpretation of the world (Filmore-

Handlon 2016). They reflect political, cultural and social conditions, and thus the distribution of power. Who gets counted counts a main argument of the book *Data Feminism* (2020) by Catherine D'Ignazio's and Lauren Klein. Data feminism is about critically analyzing data practices while actively applying (counter)data strategies to address inequalities and structural oppression.

In this presentation, I will argue why data feminism offers a fruitful framework for scrutinizing data-driven projects in film history. D'Ignazio and Klein outline seven principles of data feminism, some of which are intertwined: 1) examine power, 2) challenge power, 3) value emotion and embodiment, 4) rethink binaries and hierarchies, 5) embrace pluralism, 6) consider contexts, 7) make labor visible. Using various examples dealing with film historical data, I will focus on classifications, missing data, and labor. For example, how do we deal with the category of gender? How do we deal with absences and blind spots, dominance and privilege? How do we acknowledge the intensive labor of data-driven projects? Who is involved in film historical research and who benefits from it? With these and other questions, the paper aims to initiate a self-reflexive discussion about our goals and practices in the field of digital film history.

References

D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. Strong Ideas Series. Cambridge, Massachusetts: The MIT Press, 2020.

Fillmore-Handlon, Charlotte. "5 Things Data Cannot Do." July 8, 2016.

<http://projectarclight.org/arguments/5-things-data-cannot-do/>.

Gitelman, Lisa, and Virginia Jackson. "Introduction." In *"Raw Data" is an Oxymoron*, edited by Lisa Gitelman. Infrastructures Series. Cambridge, Massachusetts; London, England: The MIT Press, 2013, pp. 1–14.

Bios

Emil Stjernholm is an Assistant Professor in Media and Communication Studies at the Department of Communication and Media, Lund University. Currently, he is working on two digital humanities research projects: *Televising Information: Audiovisual Communication of Swedish Government Agencies* (Swedish Research Council, 2020–2023) and *Modern Times 1936* (Riksbankens Jubileumsfond, 2022–2025, <http://modernatider1936.se/en>). His areas of research include digital history, propaganda studies and television history. Together with Fredrik Norén and Claire C. Thomson, he has recently edited the volume *Nordic Media Histories of Propaganda and Persuasion* (Palgrave Macmillan, 2022). He has

previously published articles in journals such as the *Media History*, *NECSUS European Journal of Media Studies*, *Historical Journal of Film, Radio and Television*, and *VIEW Journal of European Television History and Culture*.

Isadora Campregher Paiva is a PhD candidate and lecturer at the Goethe University Frankfurt. She holds master's degrees in sociology (UFRGS, Brazil) and film studies (Goethe University). As a research associate in the project "Digital Cinema-Hub", she is dedicated to the development and application of digital methods to the study and teaching of film. She is particularly interested in using mixed methods to study film history, and her current research is focused on film credits.

Josephine Diecke is a postdoctoral researcher and lecturer in Film and Media Studies at Philipps-Universität Marburg, and the Academic Coordinator of the project "Digital Cinema-Hub" funded by the Volkswagen Foundation (2021-2026). Her research interests include film and media historiography and archaeology, moving image archiving and conservation practices, and digital methods. After graduating in Film Studies from the universities of Mainz and Frankfurt in Germany, she supervised and assisted in a number of film digitization projects, as PhD student at the University of Zurich, and as film lab technician for German and Swiss service providers. In 2022 she received her PhD with a thesis on the discursive construction of the German color film process Agfacolor. She is co-editor of the [Open Media Studies Blog](#) and serves as co-leader of the NECS workgroup Digital Methods.

Simon Spiegel is a senior researcher at the Department of Film Studies at the University of Zurich and Privatdozent at the University of Bayreuth. From 2019 to 2021 he was scientific research manager in the ERC Advanced Grant research project *FilmColors* and from 2014 to 2018 he was a collaborator in the research project *Alternative Worlds. The Political-activist Documentary Film* funded by the Swiss National Science Foundation which resulted in his professorial thesis on utopias in nonfiction films (published 2019 with *Schüren as Bilder einer besseren Welt*, English translation with Palgrave Macmillan in 2021 as *Utopias in Nonfiction Film*). From 2012 to 2014 he was part of the interdisciplinary SNF research project *Analog vs. Digital – The Emotional Impact of Film Recording Processes on the Audience* which dealt with the emotional impacts of different recording technologies. He has published widely on science fiction film, utopias and related subjects and is chief editor of the interdisciplinary *Zeitschrift für Fantastikforschung*.

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. Based on her current research and teaching focus on digital film historiography, data visualization, feminist theory, open science, and media aesthetics, she also founded the international DFG research network “New Directions in Film Historiography. Digital Tools and Methods in Film and Media Studies” (2019-2022). In addition, she initiated the collaborative Open Media Studies blog and the hybrid publishing project oa books, where she published her dissertation and blogs about academic publishing. She received a doctoral degree in Film Studies from Freie Universität Berlin and a Master of Arts from University of Michigan and has published widely on the effects of digitization in the humanities. (<https://orcid.org/0000-0002-1960-247X>, @SarahMaiDang, @SarahMaiDang@fedihum.org).

Christian Gosvig Olesen is Assistant Professor Digital Media and Cultural Heritage at the University of Amsterdam, where he teaches courses in film and media studies and in moving image archiving in the MA Preservation and Presentation of the Moving Image. His research interests include moving image archiving theory and history, film and media historiography, digital methods, found footage and sound studies. For the Dutch CLARIAH Media Suite infrastructure, he currently coordinates and develops the Learn teaching and training initiative. Previously he has carried out research on digital methods and artistic research for audiovisual archives in the projects *The Sensory Moving Image Archive: Boosting Creative Reuse for Artistic Practice and Research* (2017-2020) and *MIMEHIST: Annotating EYE's Jean Desmet Collection* (2017-2018).