

Skadi Loist; Evgenia (Zhenya) Samoiloova

How to capture the festival network: Reflections on the Film Circulation datasets

2023

<https://doi.org/10.25969/mediarep/19615>

Veröffentlichungsversion / published version

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Loist, Skadi; Samoiloova, Evgenia (Zhenya): How to capture the festival network: Reflections on the Film Circulation datasets. In: *NECSUS_European Journal of Media Studies*. #Ports, Jg. 12 (2023), Nr. 1, S. 363–390. DOI: <https://doi.org/10.25969/mediarep/19615>.

Nutzungsbedingungen:

Dieser Text wird unter einer Creative Commons - Namensnennung - Nicht kommerziell - Keine Bearbeitungen 4.0/ Lizenz zur Verfügung gestellt. Nähere Auskünfte zu dieser Lizenz finden Sie hier:

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Terms of use:

This document is made available under a creative commons - Attribution - Non Commercial - No Derivatives 4.0/ License. For more information see:

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

How to capture the festival network: Reflections on the Film Circulation datasets

Skadi Loist and Evgenia (Zhenya) Samoilova

NECSUS 12 (1), Spring 2023: 363–390

URL: <https://necsus-ejms.org/how-to-capture-the-festival-network-reflections-on-the-film-circulation-datasets/>

Abstract

The Film Circulation project is the first quantitative research in film festival studies that analyses the complex network relations of the sector using festival run data. This paper provides a detailed account of the project's dataset, including decisions on data model, collection, structuring, and enhancement. It also documents the dataset's sources, structures, limitations, and potential, with the goal of making the project more accessible and encouraging further collaborations in festival-related data analytics.

Film Circulation dataset with DOI at <https://doi.org/10.5281/zenodo.7887672> on Zenodo.

Keywords

data paper; datasets; film circulation; film festivals; festival runs

Background

The [Film Circulation project](#) emerged from discussions on the 'festivalisation' of the film industry, in which festivals have become an alternative window for distribution and exhibition.[1] The project aims to

study the festival sector's significant role in the life cycle of a film, including its development, financing, production, presentation, exhibition, and distribution.[2] By analysing the festival run of films, the project operationalises the festival network's complex structure to study the movement of films on the circuit and their influence on global film culture.

Theoretically and methodologically, the Film Circulation project is indebted to a number of research areas. Both media industries studies and new cinema history have shifted the focus away from films as aesthetic objects and contribute to a broader perspective in the analysis of film culture. Film festival studies operated in a similar vein and have long discussed the festival sector as an alternative distribution network and how it bestows symbolic capital on arthouse films.[3] Initial research on the festival sector was primarily based on single case studies and used qualitative methods.[4] Digital humanities (DH) have opened avenues to do critical cultural analysis with new digital as well as quantitative tools. New cinema history has made use of these digital methodologies, similar to an approach known as cultural analytics.[6] The network for the History of Moviegoing Exhibition and Reception (HOMER) and especially the Kinomatics project, which researches the industrial geometry of global cinema, film distribution, and exhibition, provided an inspirational context for the Film Circulation project.[6]

The Film Circulation project is the first project that shifted the qualitative, single case-study focus of festival studies to a quantitative approach and has been recognised in the field as a pioneering project for its novel methodological approach and scale in creating a global dataset of festival runs.[7] Apart from the mere task of building such a comprehensive dataset in the first place and the exploration of social computational methods and cultural analytics with this dataset, the project aims to answer some additional questions about the circulation of films and the structure of the festival sector itself. This includes descriptions of the project's core assumptions, a discussion of film circulation and the circulation power of festivals, as well as the structures of gender inequality in the festival sector.[8] This expansion of digital film festival studies has also been taken up by colleagues and offers promising avenues for further

HOW TO CAPTURE THE FESTIVAL NETWORK: REFLECTIONS ON THE FILM CIRCULATION DATASETS

research into film circulation, festival relations, and network structures of people and institutions.[9]

This article provides an account of the final dataset of the Film Circulation project including its sources, collection histories, structures, limitations, and potential. By doing so the paper aims to make the project more accessible, and in line with the FAIR(ER)[10] principles to provide a template for further data collection and to support interoperability in the hopes to invite further collaborations on festival-related cultural data analytics.

Data model

The Film Circulation dataset was created as a set of Excel spreadsheets and is organised as a relational database, meaning each table (csv file) consists of records with unique identifiers that allow linking records throughout various tables. The tables are organised into four segments, or sub-datasets, according to their origin and purpose for the project: 1. Film; 2. Survey; 3. IMDb; 4. Festival Library (see Figure 1). Four tables form the core of the database with the other 14 files offering variations and extension in 'join' tables.

In this Film Circulation relational database, films, festivals and related information can be linked based on the following keys:

- the unique film ID (unique.id), the ID assigned to each individual film,
- the festival ID (festival.id), the ID assigned to each individual film festival,
- the IMDb id (imdb.id), the ID that the Internet Movie Database uses, here available for films in our dataset that we identified on IMDb

Below we describe the most significant tables. For a more detailed description of tables, fields, and scripts, please consult the more extensive documentation that accompanies the deposited Film Circulation dataset on Zenodo.[11]

Film Circulation dataset model

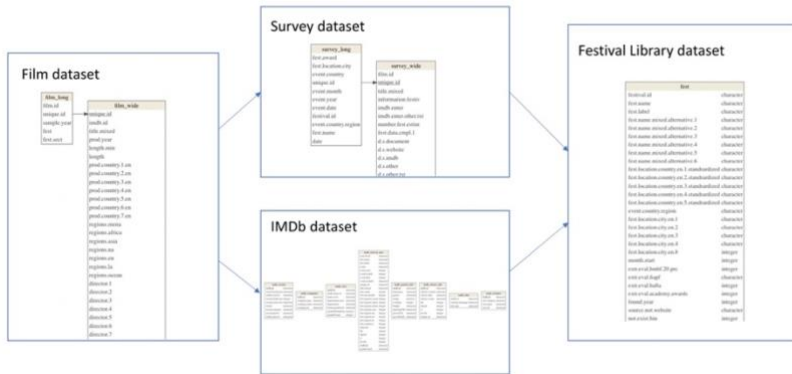


Fig. 1: Dataset scheme

The *Film Dataset* subset consists of a data scheme image file, a codebook, and two dataset tables in csv format. This dataset is the core of the project and comprises all films which set the sample frame for tracking the film circulation. It lists in file ‘1_film-dataset_festival-program_long’ all films and the sample festivals, festival sections, and the year of the festival edition that they were sampled from. In addition, it includes further core data for each film such as film title, production year and country, director name, length, genre attribution, and information on IMDb festival data.

The *Survey* subset consists of a data scheme image file, a codebook, and two dataset tables in csv format. In this subset survey responses on the festival runs of our sample films are made available (see file ‘2_survey-dataset_long-festivals_shared-consent’). In addition, the wide format file includes data on film IDs, film title, survey questions regarding completeness and availability of provided information, information on number of festival screenings, screening fees, budgets, marketing costs, market screenings, and distribution.

The *IMDb dataset* consists of a data scheme image file, a codebook, and eight dataset tables in csv format. It also includes the R scripts that we used for scraping and matching. Data from IMDb are organised in different dataset tables to account for the large amount of different data related to alternative titles, general info, companies, crew data, release information, festival runs, awards, and websites listed for the matched films. The table ‘3_imdb-

HOW TO CAPTURE THE FESTIVAL NETWORK: REFLECTIONS ON THE FILM CIRCULATION DATASETS

dataset_festival-runs_long' contains festival run data scraped from IMDb in a long format, i.e. each row corresponds to the festival appearance of a given film. The dataset does not include each film screening, but the first screening of a film at a festival within a given year. The data includes festival runs from 2011 up to 2019.

The *Festival Library* subset consists of a data scheme image file, a codebook, and one table in csv format. The table '4_festival-library_dataset_imdb-and-survey' contains data on all unique festivals that were collected as part of the festival runs for our sample films. This includes data on festival ID, sources, festival name and alternatives, location data for country, region and city, data on award recognition, founding year, whether a festival has a specialisation and if so the ascribed festival categories.

Data sources & data collection

The project collected data at two levels: films and festivals. At both levels, the data shows a notable heterogeneity. For example, films include all lengths (short, medium, and long films), films of various types (narrative, documentary, animation, experimental), various genres, and also historical films in archival and retrospective programs. Festivals include the complete range, from internationally-recognised industry festivals with FIAPF accreditation to very small local community festivals. As we will discuss in more detail below, the intended breadth and heterogeneity of films and festivals necessitated data collection from different data sources.

The datasets offer data that have been collected through festival catalogs, IMDb, and an online survey, as well as variables that have been generated or recoded within the analysis, for instance, categorisation by lengths (shorter or longer than 41 min), genre (animation, documentary, experimental, and fiction), region (e.g. Europe, MENA, Oceania) based on information of production country or festival location countries.

Film Dataset: Festival programs

The core Film Dataset comprises the data collected from the sampled festival programs of six festivals (Cannes Film Festival, Berlin International Film Festival [Berlinale], Toronto International Film Festival [TIFF], International Documentary Film Festival Amsterdam [IDFA], Clermont-Ferrand Short Film Festival, and Frameline San Francisco International LGBTQ+ Film Festival) over seven years (2011-2017). Data was collected from the programs of the respective festival editions. For festival season 2013 all the data was collected from the published festival programs, either in PDF form or from the festival websites. Since TIFF does not provide digital data from previous years, the data had to be collected from a printed catalog. As sources like web pages and PDFs were not uniformly structured, automated text mining was not feasible and data had to be collected and entered manually.

Midway into the project, while running first analytics, it became clear that the 2013 dataset (roughly 1,800 films) was too small for analysis of specific subsets.[12] Thus, the sample was expanded to include the programs of festival seasons 2011-2017. During the expansion of the sample size beyond the original festival season 2013, data for festival seasons 2011-2012 and 2014-2017 for Berlinale, Clermont-Ferrand, TIFF, and Frameline was kindly provided by the festivals.[13] The rest of the data was collected from digital archives available on the festival websites.

While the information provided in the catalogs varies from festival to festival, they contain common information that serves as master data. We collected data like title of the film, year of production, film length, country of production and co-production, name of director, film type and genre, festival, section, and festival year. Furthermore, the data was coded or recoded to include specific genre groups (animation, documentary, experimental, fiction, LGBTQ) or regions of production countries (MENA, Oceania, Sub-Saharan Africa, Asia and Central Asia, North America, Europe, Latin America and the Caribbean).

Survey

Between May 2019 and December 2022, we conducted a web-based survey among the production companies and filmmakers connected to the films in our sample.[14] The survey aimed to collect data on film circulation through festivals and other data not available through open sources. To reduce the drop-out rate, the survey was kept short and focused on five areas. These areas covered festival runs and markets, finances, festival consulting, distribution, and follow-up interviews. We asked for the full list of the festival runs, including festivals and markets attended. Finances were explored through questions on budget, marketing costs, submission and screening fees, and budget limitations. Questions on festival consulting and distribution were included, as well as an invitation for follow-up interviews, to which 60% of respondents agreed.

The online survey was distributed to 6,010 contacts associated with 6,755 unique films. The final sample resulted in 454 unique respondents, which equals a response rate of about seven percent. Contact details were primarily obtained from festival programs. Films produced before 1990 were excluded due to a lack of contact details and low response likelihood. Respondents were incentivised by the offer to enter or update their film data on IMDb. Of those surveyed, 69% took up the offer.

IMDb & scripts

From our defined sample of films and their characteristics based on the festival programs, we have written a web-scraping code for R that allowed us to identify the films from the core Film Dataset in the IMDb database. If the respective film was available on IMDb in the first place, we collected data on their festival runs. The data were collected between 2018-2020 using the web-scraping scripts for R.[15]

Overall, the web scraping process involved multiple steps to accurately match and scrape data from IMDb. Using various R packages and data from the core Film Dataset (specifically, film title and production year), we created a search URL for each film and scraped the data (title, production year, directors, genre, and running time) for the first page of suggested results. The

data of the suggested search results was then used to look for the best match to the film in the original dataset based on the festival program. To identify the quality of the matches, we applied a fuzzy matching approach that allowed us to create a probability score of the match based on the film title, names of directors, and production year (plus or minus one year, because the production year data in festivals are variable). The results were split into perfect matches, matches of high probability, matches of medium probability, matches of low probability, and pairs, where no matches were found. Matches of high and medium probability were checked manually, while matches of low probability/no matches were rejected and perfect matches accepted. For the matching and matching quality we used the R packages ‘stringdist’.[16]

For an analysis regarding gender inequality in the film festival sector,[17] we ascribed a binary gender of crew members based on their first names using the GenderizeR application.[18] This data is available in the crew dataset. In line with problematising a strictly binary conception of gender, for the 2013 festival season sample we tested manual coding methods to account for more nuanced gender presentation based on sources that resemble self-identification using available online sources such as Wikipedia and personal websites.[19] Due to high resource intensity, we were not able to manually code gender in this way for the whole dataset of all festival sample years.

Festival Library

The ‘Festival Library’ dataset has grown out of data collection on festival runs for our sample films collected on IMDb and via the survey; it consists of 3,860 unique festivals. The data were collected between 2018 and 2022. Once the festival screening information was collected, data had to be matched and homogenised to account for the different festival name versions.

Alternative names could result from name changes over the time period or different language versions of the festival names. For the analysis the international English name was prioritised.

Additional festival features, such as location, country, region, month of the event were manually searched (e.g. the host city) and coded based on available information on the web (e.g. the festival specialisation). Further

HOW TO CAPTURE THE FESTIVAL NETWORK: REFLECTIONS ON THE FILM CIRCULATION DATASETS

variables, such as festival categories (e.g. short, animation, documentary, genre, identity-based festivals, etc.), were coded based on a categorisation scheme developed in the project and discussed during the Workshop on Data Collection and Operationalisation of Film Festival Categories in May 2021.[20]

Further data on festival valuation was manually collected from festival listings available from industry organisations, such as film funding agencies (e.g. the German Film Funding Agency FFA, which gives out reference points for certain festival participation and festival awards), festival regulators (e.g. FIAPF accreditation) and awards organisations (BAFTA, Academy Awards). This coding is relevant for analysis of festival hierarchies.

Discussion

This section discusses some aspects of the data collection and the design of the data model that merit critical reflection on the benefit for scholars that plan to use the data in their research.

Project design & data collection

Current empirical research cannot disregard the emergence of new digital sources of data, referred to as found,[21] organic,[22] or trace data.[23] Such data sources have been changing the landscape of empirical approaches across natural and social sciences as well as humanities. In this project, we are incorporating techniques utilised in DH and computational social sciences. With respect to film research, DH presents opportunities for computational analysis with newly-accessible cultural data. Significantly, it enables the quantitative analysis of vast amounts of data without compromising the critical approaches of humanities and film studies.

The main challenge was to design a research project that relied heavily on data analysis in an area where data is scarce. Commercial data on film festivals are hardly available, presumably because they are not profitable within the industry, as festival runs are very time-specific and involve different industry players such as festivals, buyers, distributors, or agencies.

There is only a short time window when different market players are interested in the festival run. Film tracking in the festival sector is interesting for film buyers only as long as the films are not yet sold (in the respective territories). Film festivals only collect limited data upon submissions to find out which premieres are still available: world, international, continental, regional, or national. Information on premiere status is relevant for festivals on the one hand to be able to generate press attention and on the other hand it serves as a quality signal. In recent years, there has been an increase in consulting services of agencies that manage submissions to festivals. They track their films in the submission phase and have a list of festival invitations. However, since these data are part of the business model of the paid service providers, they are not accessible. Due to this complex situation in tracking the entire festival run, there is no easily accessible source for such data.

However, this limitation presented an opportunity for us to create a project by developing an operationalisation that aimed to generate a thorough and dependable dataset with elevated standards of data quality. By using this approach, we were able to overcome the obstacle of having limited data availability and achieve a more intricate comprehension of the subject matter.

Sample

The project operationalises the complex structure of the festival network through the festival run of films. For this, the project collected data on two different levels: we first defined a core dataset on the level of films and collected data on the screenings of those films at festivals. This created, in a second step, a dataset of festivals.

The project aimed to gather quantitative data on the circulation of a subset of international films in the festival sector. The goal was to examine the global operations of the festival sector, identify any potential differences based on film attributes, such as genre and production country and understand how festivals are linked through films, thereby forming the networked ecosystems. To accomplish this, a diverse and extensive sample of films and their festival runs was required. The sample was designed to be as varied as possible, while still providing a clearly delineated framework for

HOW TO CAPTURE THE FESTIVAL NETWORK: REFLECTIONS ON THE FILM CIRCULATION DATASETS

the source dataset. To create a manageable sample size for the laborious process of data collection and cleaning, data was initially collected from the festival programs of six major film festivals.

The six source festivals were chosen to allow for diversity of films that travel the circuit (Table 1). The sample includes three A-list festivals, i.e. festivals with an industry perspective and impact on the international festival circuit due to their premiere status and circulation power,[24] and three festivals with a specialisation. These festivals cover different regions and different moments of entry onto the festival calendar.

Sample festival	Category	Calendar	Location
Berlin International Film Festival (Berlinale)	A-list (FIAPF), industry festival	February	Germany, Europe
Cannes Film Festival	A-list (FIAPF), industry festival	May	France, Europe
Toronto International Film Festival (TIFF)	Best-of festival (FIAPF), industry festival	September	Canada, North America
International Documentary Film Festival Amsterdam (IDFA)	Specialized documentary film festival, with market	January	Netherlands, Europe
Clermont-Ferrand International Short Film Festival	Specialized short film festival, with market	January	France, Europe
Frameline: San Francisco International LGBTQ+ Film Festival	Specialized LGBTQ film festival	June	USA, North America

Table 1: Sample festivals

We included the complete selection of films from the festival programs, not just the competition sections, to also capture those with smaller budgets, less marketing clout, and without premiere status. This was done to account for their different trajectory on the festival network and circulation patterns. The originally chosen timeframe was the 2013 festival season to assure that the festival runs were already completed.

Based on this film sample, the next step was to collect data on screenings of the films at festivals to reconstruct the festival runs. Since no single authoritative source for festival runs is available, various sources were combined to achieve this: data from the IMDb Release section, film websites, and a survey that we sent to license holders – mainly producers and sales agents. This resulted in the two datasets of Survey and IMDb data. Finally, information about the film screenings provided input for the resulting fourth dataset of global film festivals, the Festival Library dataset.

Understanding the origin and logic of the sample is crucial for interpreting the results of the study. Our objective was to compare festival runs of films selected by A- and B-level festivals and examine smaller festivals in the festival network hierarchy. A random selection of films produced between 2011 and 2017 that is not defined by set source festivals would produce different findings. Hence, it is important to note that the sample is not representative of all the films and festivals within the chosen timeframe.

Film dataset

Collecting data on films from festival programs may seem straightforward, but it can be challenging due to the lack of easily accessible and uniformly structured data formats, such as web pages or PDFs. This problem is common in the festival sector, which often operates on a seasonal and underfunded basis.[25] Notably, the Toronto International Film Festival (TIFF), one of the largest film festivals in North America, is the only festival in the sample without an accessible digital archive. Although TIFF has an archive department and its own building, PDF catalogs for the selected years were unavailable, and print catalogs were only accessible at the Toronto Public Library.

There were additional challenges encountered while working with the data, such as issues with the structure and compatibility of the data. Some of the PDFs provided could not be easily processed, requiring manual work due to their non-uniform structure. Additionally, there was a lack of documentation and explanation of terminology used in some programs, such as inconsistent use of genre categories. For instance, the genre of some films had to be determined through manual research or additional sources since Cannes, Berlinale, and IDFA festival programs did not explicitly indicate genre information.

The dataset includes 9,972 films from six sample festivals between 2011 and 2017, spanning seven festival editions. Of these films, 593 (6%) were found in multiple festival programs, for instance, a film might have shown both at Berlinale and Frameline in the same or consecutive festival years. Thus, the core dataset is made up of 9,348 unique films.

HOW TO CAPTURE THE FESTIVAL NETWORK: REFLECTIONS ON THE FILM CIRCULATION DATASETS

Festival edition // Sample festival	2011	2012	2013	2014	2015	2016	2017	Total
<i>Berlinale</i>	229	232	355	240	232	276	274	1.838
<i>Cannes</i>	83	86	115	82	81	132	118	697
<i>TIFF</i>	347	379	322	401	403	405	356	2.613
<i>IDFA</i>	330	341	253	274	295	275	298	2.066
<i>Clermont-Ferrand</i>	137	168	437	161	167	166	165	1.401
<i>Frameline</i>	229	215	223	209	179	155	147	1.357

Table 2. Overview of the number of films collected from the sample festival programs in a seven-year period (2011- 2017).

Table 2 displays the number of films in the sample for each festival over the seven-year period. Data for TIFF, Frameline, and Clermont-Ferrand in years 2011-2012 and 2014-2017 were provided by the festivals themselves, while the rest were collected from digital festival archives. For 2013, data for all festivals were obtained from published festival programs, except for Clermont-Ferrand where data were obtained from the festival website for all festival sections. Films were sampled from both competitive and non-competitive sections for all festivals, except for Clermont-Ferrand in 2013. This explains slight discrepancies for the 2013 Clermont-Ferrand data. Key analyses were checked for robustness by comparing subsamples of Clermont-Ferrand with and without non-competitive sections to account for potential missing data issues.

The sample comprises 150 unique production countries, with only five films lacking production country data. Figure 2 displays the countries with the highest level of productions within the sample, which are the USA (23%, n=2,181), France (16%, n=1,526), Germany (10%, n=899), Canada (8%, n=738), the UK (7%, n=650), and the Netherlands (6%, n=545).



Fig. 2: Share of production countries in the entire sample (9,343 unique films) in percentages.

The primary factor influencing the composition of production countries in the sample is the high representation of films produced in the countries where the sample festivals are located. Among the sample festivals, Cannes, Berlinale, TIFF, and IDFA have a more balanced distribution of production countries, while Frameline and Clermont-Ferrand are more focused on local production. The number of unique production countries is the smallest for Frameline (n=63), followed by Cannes (n=84), Clermont-Ferrand (n=95), TIFF (n=109), IDFA (n=111), and Berlinale (n=113). In all sample festivals except TIFF, the location countries are also the most frequent production countries of the films presented. Frameline has the highest proportion of films produced in the USA (66%), followed by Clermont-Ferrand (41%) and Cannes (39%), which are dominated by French productions. For the Berlinale program, 25% of the films were produced in Germany, while IDFA and TIFF have 20% and 21% of home productions, respectively.

For a broader overview of the geographical representation of production countries, we grouped countries in regions as defined by the World Bank.[26] As we would expect, most of the productions come from Europe[27] (53%, n=4,950), followed by North America[28] (31%, n=2,884). Asian[29] productions (including East, South, and Central Asia) constituted eleven percent (n=1,001). Latin American and Caribbean[30] productions were represented in six percent of films (n=586). Similarly, the MENA[31] region was represented in five percent of films in the sample (n=490). The Pacific[32] region and Sub-Saharan Africa[33] were least represented with three percent (n=268) and two percent (n=150), respectively.

For one-fifth of the films (19%, n=1,752) the festival catalogs listed more than one production country. Among these co-productions, the most frequent combinations were France and Germany (n=144), Belgium and France (n=103), United Kingdom and United States (n=68), France and Italy (n=50), France and Switzerland (n=45), France and United States (n=44), Belgium and Netherlands (n=40), France and United Kingdom (n=37), Germany and Switzerland (n=37), and Germany and Netherlands (n=36).

One of the limitations of this study is that we are not able to delve deeply into the programming of the sample festivals. Due to the focus of our analysis on the films and their circulation patterns and the corresponding sampling frame, the dataset does not comprise information about the various

programming sections, screening dates, and specific venues within the festivals. This lack of information means that we cannot provide a comprehensive analysis of the hierarchies within the programming, such as which sections are considered the most prestigious or which venues attract the largest audiences. Provided with a larger sample of festivals with their full programs, future research could benefit from a more in-depth analysis of the programming structures of these festivals, such as overlap or uniqueness in the curation. This could shed more light on the various factors that shape the selection and presentation of films in festival contexts.

Survey dataset

Of the online survey sent out to 6,010 contacts associated with 6,755 unique films between 2018 and 2021, we achieved a response rate of about seven percent, resulting in 454 unique responses. Based on the experience of a pilot study, we found that many film producers and distributors collect data on the festival runs of their films. While surveys may be the most effective way to obtain a comprehensive understanding of these runs, there is a risk of low response rates, particularly with web-based surveys. Although this does not necessarily impact the representativeness of the sample, it can result in variable-specific nonresponse bias that requires further investigation. Nonetheless, a moderate response rate can still provide valuable insights into complete festival runs, especially for specific subgroups. Contact details were primarily obtained from festival programs. Films produced before 1990 were excluded due to a lack of contact details and low response likelihood. Respondents were incentivised by the offer to enter or update their film data on IMDb. Of those surveyed, 69% took up the offer.

The deposited dataset only contains survey responses from participants who agreed to share their data. To comply with GDPR and copyright regulations, some information such as production company details, contact information, and film synopses have been excluded from the published version. The survey dataset contains survey responses for 454 films, of which festival run data was provided for 206 films (45%). However, only data from participants who gave consent can be shared, resulting in 379 films with information on budgets, marketing, etc. in the wide-format table ('2_survey-dataset_wide-

no-festivals_shared-consent'), and 161 films with festival run data in the long-format table ('2_survey-dataset_long-festivals_shared-consent').

The vast majority of contacts (95%) belong to production companies, producers, and directors. Five percent of the contacts were email addresses of world sales companies where no production contacts were available. Our primary focus was on producers, as we anticipated that they would be more inclined to respond due to their direct involvement in filmmaking and authority to respond. In contrast, employees of a world sales company may be dependent on the decisions of their managers.

Of the respondents, 85 percent (n=384) identified themselves as film producers or directors, while the remaining respondents held other positions such as festival managers, distributors, sales and production managers, or interns.

Sample festival	Percentage of films responding to the survey with available festival run data (n)
Berlin International Film Festival (Berlinale)	34 % (n=71)
Festival de Cannes (Cannes)	2 % (n=5)
Toronto International Film Festival (TIFF)	9 % (n=18)
Clermont-Ferrand International Short Film Festival (Clermont-Ferrand)	23 % (n=48)
International Documentary Film Festival Amsterdam (IDFA)	12 % (n=24)
Frameline: San Francisco International LGBTQ+ Film Festival (Frameline)	16 % (n=34)

Table 3: Share of films sampled from each sample festival in the total survey dataset with available information on festival runs, n=206.

The dataset containing festival run data mainly consists of films selected from Berlinale, Clermont-Ferrand, Frameline, and IDFA (see Table 3). The survey faced difficulties in reaching films selected at Cannes and TIFF, which may be due to their higher prestige and representation by sales agents rather than producers. The sample of 206 films that provided festival data mainly represents minority genres, including 55% (n=113) of shorts, 32% (n=67) of documentaries, and 24% (n=50) with LGBT*Q themes. This indicates a response bias influenced by accessibility and industry hierarchies. The top two production countries represented in the responses are the United States (n=27) and Germany (n=23), followed by France (n=17) and Canada (n=11). The high response rate for films sampled at Berlinale (see Table 3) and the number of German films in the responses may reflect the project's German-based location.

IMDb Dataset

When taking a closer look at the IMDb data, several biases or omissions become apparent. Between 2018-2020, we collected data on IMDb for 7,851 films, accounting for 84% of the sample. However, only 7,150 films (76%) had corresponding festival data available, as some accounts were incomplete. Table 4 provides an overview of the IMDb variables and their corresponding share of missing data, with budget and box office data having the highest percentage of missing values.

IMDb variable	Percentage of the missing data
Release data (festivals, theatrical release, digital release, TV, DVD/Blue Ray) & festival awards	9 % (n=701)
Production countries	10 % (n=753)
Film language	14 % (n=1,075)
Runtime	11 % (n=845)
Genre	10 % (n=768)
Rating	22 % (n=1,719)
Budget	86 % (n=6,771)
Opening USA	88 % (n=6,880)
Gross USA	86 % (n=6,778)
Gross World	74 % (n=5,802)
Keywords of the film's topic	39 % (n=3,070)
Film title in other languages	27 % (n=2,111)
Plot summary	7 % (n=569)
Distribution & production company	15 % (n=6,652)
Credits of the crew members	0 % (n=0)
Websites and other online resources about the films	5 % (n=385)

Table 4: IMDb variables and share of missing data for the sample films, n=7,851.

Despite being the most widely used and comprehensive film database that is openly accessible, the IMDb structure is not entirely transparent, and is receiving relatively little systematic critical attention.[34] Since its inception as a crowd-sourced platform, the database has undergone several iterations and data integrations over the years, which keeps overall data quality unclear. While we use IMDb for its convenience, we recognise the likely selection bias for this source, as films available on IMDb may differ systematically from those not available, e.g. underrepresented film types. Although we cannot ensure comprehensive data quality of this source, we can evaluate its fitness for specific research questions. While conceding that no perfect method of data collection exists, we need to be transparent about its limitations and data generation processes. Researchers and practitioners working with data are increasingly adopting strategies of data integration to balance the strengths and weaknesses of various sources, given specific research questions.[35]

Regarding festival research, we were able to evaluate the quality and usefulness of IMDb data to some extent. Survey data from a smaller pilot study[36] analysed within the Circulation project revealed that festival runs recorded on IMDb are fragmented. The pilot study sample consisted of 39 films nominated for the Berlinale Teddy Award, for which festival run data were collected via a survey and identified on IMDb. When comparing the festival screenings reported by filmmakers in the survey with those listed on IMDb, it was found that there was no complete festival run recorded on IMDb (as shown in Figure 3a). However, a more detailed comparison of the two datasets revealed that IMDb festival data can be used to estimate the festival run in months (as shown in Figure 3b).

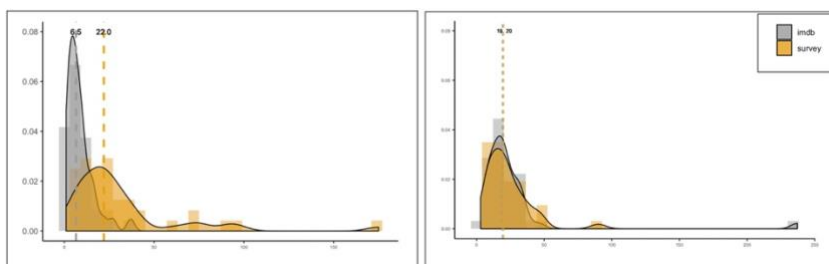


Fig. 3: Comparison of two distributions (IMDb versus survey data). Left: Comparison of the number of festivals in the festival run. Right: Comparison of the length of the festival run in months.

Table 5 shows the share of films with festival data available on IMDb by each sample festival. Festivals that are closer to traditional industry structures are better represented on IMDb than the three specialised festivals. Interestingly, of the three industry festivals, TIFF is least represented when compared to Berlinale and Cannes.

Sample festival	Percentage of films identified on IMDb with at least one festival screening available (n)
Berlin International Film Festival (Berlinale)	91% (n=1,595)
Cannes Film Festival	95% (n=657)
Toronto International Film Festival (TIFF)	81% (n=1,936)
Clermont-Ferrand International Short Film Festival	71% (n=933)
International Documentary Film Festival Amsterdam (IDFA)	65% (n=1,240)
Frameline: San Francisco International LGBTQ+ Film Festival	63% (n=789)

Table 5: Share of films with available festival data identified on IMDb by the sample festival.

HOW TO CAPTURE THE FESTIVAL NETWORK: REFLECTIONS ON THE FILM CIRCULATION DATASETS

Regarding the films listed on IMDb, there is an apparent under-representation of short films (i.e. films shorter than 40 minutes), with only 63% (n=2,495) of them being available on the platform. On the other hand, medium-length and long films (i.e. films longer than 40 minutes) are better represented, accounting for 88% (n=4,654) of the films identified on IMDb.

When examining the representation of films on IMDb by genre, we found that animation (81%, n=434), fiction (80%, n=4,270), and animated documentaries (82%, n=55) are the most well-represented, followed by documentaries (72%, n=2,267) and experimental documentaries (66%, n=38). On the other hand, experimental fiction films (52%, n=86) are the least represented. Furthermore, films with LGBT*Q topics are less likely to be available on IMDb with festival data, as only 70% (n=1,110) of them are represented compared to 78% (n=6,040) of other films.

Festival library

The Release Dates section on IMDb not only lists the start dates for theatrical releases by territory but also includes festival premieres for films. These festival screening dates provided the majority of festival library data for our sample films. While the survey sample size with a mere 7% response rate was relatively small, it still contributed 700 additional festivals that were not covered by the data gathered from IMDb. Of these additional festivals, the majority (72%) were classified as specialised events in our manual coding. Therefore, the combination of both data source and methods of data collection demonstrated a clear advantage.

Regarding the statistical description of the festival library, we found some interesting results. The foundation years of the festivals ranged from 1932 to 2019, as shown in Figure 13. However, 27% of the festivals (1,057) had missing data on their foundation years. We were not able to find 153 festivals on the web, and for 188 festivals we could not verify sufficient data to establish their uniqueness.

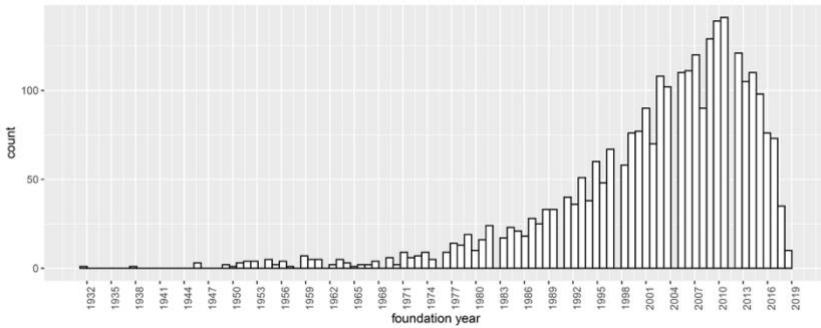


Fig. 4: Distribution of foundation years of the festivals with available data, n=2,803 festivals.

The most common festival months are October (n=580) and November (n=551), while December (n=168) and January (n=155) have the least amount of festivals in our sample.

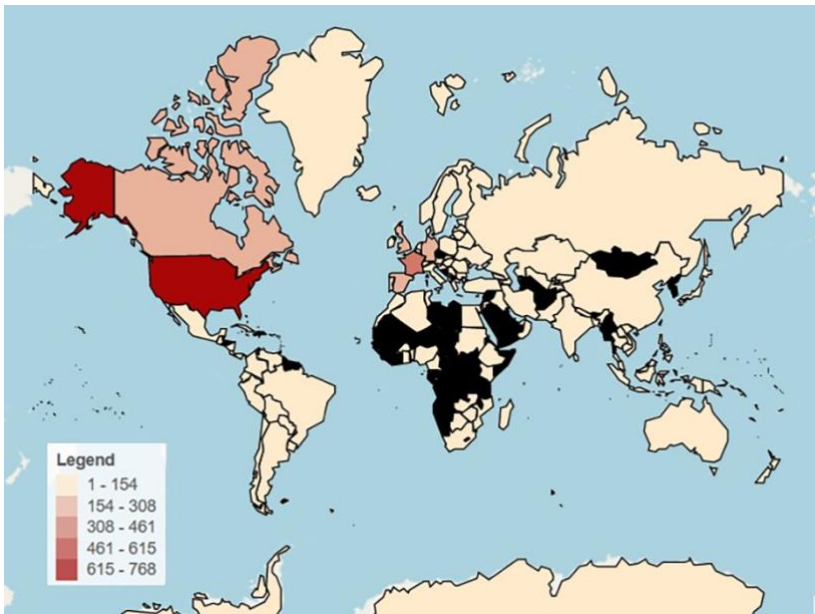


Fig. 5: Spatial distribution of the festivals from the festival library (n=3860).

Based on the basic information such as festival name, location, and date for the festival library dataset (n=3,860), we can provide insights into the festival industry in terms of our sample. The festivals are located predominantly in

HOW TO CAPTURE THE FESTIVAL NETWORK: REFLECTIONS ON THE FILM CIRCULATION DATASETS

Europe (n=1,897) and North America (n=1,011), followed by Asia (n=431), Latin America and the Caribbean (n=284), Oceania (n=90), the MENA region (n=87), and sub-Saharan Africa (n=36 festivals). The most common countries for festival locations include the United States (n=768), France (n=315), Canada (n=244), Spain (n=206), Germany (n=161), the UK (n=158), Italy (n=154), Japan (n=119), Poland (n=78), and the Netherlands (n=76).

It is important to keep in mind that these findings are limited to the festivals included in our library and are not necessarily representative of all festivals worldwide. We acknowledge limitations to the dataset, as we only traced the festival runs of the films in our sample and did not scrape and analyse current submission platforms like FilmFreeway, which could potentially yield a different distribution of festivals globally.

As discussed above, it is important to note that festivals tend to feature films that are produced in the same region as the festival. This practice is logical as festivals aim to promote local talent, support the local industry, and consider cultural proximity in their programming decisions. This trend is also reflected in the festival runs and thus the structure of our festival library. However, further research is necessary to validate these findings with a different sampling strategy. This is why combining existing data can be advantageous.

We manually searched, checked, and categorised each festival that we found as part of the festival runs. During this process, we classified festivals based on their position within the festival ecosystem and their program specialization.[37] However, we encountered a significant limitation when using a solely quantitative approach to categorise festivals in the large-scale, global dataset. We found it challenging to categorise festivals meaningfully from a distance. Studies of smaller datasets and locally or regionally-grounded projects, such as investigating the festival landscapes in Chile or the Basque country,[38] have already shown that festival categorisation is difficult even on a small scale and benefits significantly from triangulation with qualitative and ethnographic research. Therefore, further discussions about a systematic categorisation scheme and corresponding semantic systems would promote dataset interoperability and analytical comparability.

Conclusion

Our objective is to enhance the field of festival studies by delivering a thorough and refined comprehension of the festival ecosystem and its significance in the worldwide film industry.

However, we acknowledge that the array of films and festivals encompassed in the dataset presents both advantages and challenges. We understand the necessity of broadening and categorising films and festivals further to guarantee nuanced analysis. Additionally, we are in the process of devising the launch of a network dataset that will facilitate further analyses.

We recognise the drawbacks of the current dataset, which only encompasses festival seasons from 2011-2017, and its tendency to become outdated quickly, as is common with data projects. However, the resource-intensive work in data collection and enhancement for data updates and further research necessitates further funding to continue the work. Despite this, we strive to expand the dataset and foster collaborations with other film festival scholars and researchers in the broader realm of creative industries and cultural analytics. This is a significant reason why we are making the dataset available in open access.

To address regional bias, we aim to collaborate with international colleagues and add datasets from different regions. This will help to diversify our sample, which was originally based on six festivals with a Western (mainly European and North American) focus. We plan to collaborate with existing projects and datasets of festivals from Chile,[39] Central America,[40] the Basque country,[41] the Netherlands,[42] and beyond.

To discuss programming strategies at festivals more broadly, such as homogeneity or uniqueness in curating, we need to employ a different data gathering/sampling strategy. One possible strategy is to collect full festival programs from a larger set of festivals and trace shared films across festival programs. We plan to collaborate with existing projects that specialise in programs, such as the ILIAS project and the trans cinema project, to achieve this goal.

Acknowledgements

The project Film Circulation in the International Festival Network and the Influence on Global Film Culture on which this report is based was funded by the German Federal Ministry of Education and Research (BMBF) under the grant number 01UL1710X. The responsibility for the content of this publication lies with the author.

Authors

Skadi Loist is Assistant Professor for Production Cultures in Audiovisual Media Industries at the Film University Babelsberg KONRAD WOLF in Potsdam, Germany. Skadi's research focuses on film festivals and circulation, queer film culture, and diversity, inclusion and sustainability in screen industries. Skadi is a former NECS Steering Committee member and current Editorial Board member of NECSUS_European Journal of Media Studies. Publications include *Film Festivals: History, Theory, Method, Practice* (Routledge 2016). Their current research project is Gender Equity Policy Analysis (DFG/ESRC/SSHRC 2021-2024). ORCID 0000-0002-8299-4103.

Evgenia (Zhenya) Samoilova is a postdoctoral researcher in the project TrainDL at the University of Potsdam. Prior to that she was a researcher at the Film University Babelsberg KONRAD WOLF, at the department of Statistics and Methodology at the University of Mannheim, and at the GESIS Leibniz Institute of Social Sciences. She received her PhD in Sociology from the Bremen International Graduate School of Social Sciences (University of Bremen/Jacobs University). ORCID 0000-0003-3858-901X.

References

- Barnes, H. 'The Data-Driven Festival: Recordkeeping and Archival Practices' in *Documentary film festivals vol. 1: Methods, history, politics*, edited by A. Vallejo and E. Winton. New York-Cham: Palgrave Macmillan / Springer, 2020: 53-59.
- Bilteyst, D. and Meers, P. 'New Cinema History and the Comparative Mode: Reflections on Comparing Historical Cinema Cultures', *Alphaville*, no. 11, 2016: 13-32; <https://doi.org/10.33178/alpha.11.01>.

- Coate, B., Verhoeven, D., Palmer, S., and Arrowsmith, C. 'Using Big Cultural Data to Understand Diversity and Reciprocity in the Global Flow of Contemporary Cinema' in Proceedings of the International Symposium on the Measurement of Digital Cultural Products, edited by UNESCO Institute for Statistics, Montreal, 2016: 141-151; <http://dro.deakin.edu.au/view/DU:30091474>.
- de Valck, M. *Film festivals: From European geopolitics to global cinephilia*. Amsterdam: Amsterdam University Press, 2007.
- de Valck, M. and Loist, S. 'Film Festival Studies: An Overview of a Burgeoning Field' in *Film festival yearbook 1: The festival circuit*, edited by D. Iordanova and R. Rhyne. St. Andrews: St Andrews Film Studies, 2009: 179-215.
- Ehrich, M., Burgdorf, K., Samoilova, Z., and Loist, S. 'The Film Festival Sector and Its Networked Structures of Gender Inequality', *Applied Network Science*, 7, 1, 2022: 1-38; <https://doi.org/10.1007/s41109-022-00457-z>.
- Elsaesser, T. 'Film Festival Networks: The New Topographies of Cinema in Europe' in *European cinema: Face to face with Hollywood*. Amsterdam: Amsterdam University Press, 2005: 82-107.
- Falicov, T. 'The "Festival Film": Film Festival Funds as Cultural Intermediaries' in *Film festivals: History, theory, method, practice*, edited by M. de Valck, B. Kredell, and S. Loist. London-New York: Routledge, 2016: 209-229.
- Gass, L (ed.). 'Die Zukunft der Filmfestivals / The Future of Film Festivals', special issue, *Schnitt*, no. 54, 2009: 6-45.
- Golder, Scott A., and Michael W. Macy. 2014. "Digital Footprints: Opportunities and Challenges for Online Social Research." *Annual Review of Sociology* 40 (1): 129–52. <https://doi.org/10.1146/annurev-soc-071913-043145>.
- Groves, R. 'Three Eras of Survey Research', *Public Opinion Quarterly*, 75, 5, 2011: 861-871; <https://doi.org/10.1093/poq/nfr057>.
- Hill, C., Biemer, P., Buskirk, T., Japac, L., Kirchner, A., Kolenikov, S., and Lyberg, L. 'Introduction' in *Big data meets survey science: A collection of innovative methods*, edited by C. Hill, P. Biemer, T. Buskirk, L. Japac, A. Kirchner, S. Kolenikov, and L. Lyberg. Hoboken: Wiley, 2021: 1-8.
- Iordanova, D. 'The Film Festival as an Industry Node', *Media Industries*, 1, 3, 2015: 7-11; <https://doi.org/10.3998/mij.15031809.0001.302>.
- Iordanova, D. and Rhyne, R (eds). *Film festival yearbook 1: The festival circuit*. St Andrews: St Andrews Film Studies, 2009.
- Japac, L., Kreuter, F. Berg, M., Biemer, P., Decker, P., Lampe, C., Lane, L., O'Neil, C., and Usher, A. 'Big Data in Survey Research', *Public Opinion Quarterly*, 79, 4, 2015: 839-880; <https://doi.org/10.1093/poq/nfv039>.
- Loist, S. 'On the Relationships Between Film Festivals and Industry' in *Busan cinema forum: Seeking the path of Asian cinema: East Asia*, edited by Y. Lee. Busan: Busan Cinema Forum, 2011: 381-402.
- 'Zirkulation im Netzwerk: Eine Betrachtung zur Zirkulationskraft von Filmfestivals', *Zeitschrift für Medienwissenschaft*, no. 23, 2020: 55-63; <https://doi.org/10.25969/mediarep/14833>.
- 'Stopping the Flow: Film Circulation in the Festival Ecosystem at a Moment of Disruption' in *Rethinking film festivals in the pandemic era and after*, edited by M. de Valck and A. Damiens. Cham: Palgrave Macmillan, 2023a.
- 'Studying Film Circulation: Moving Film Festival Research to an Evidence-Based, Global Perspective' in *Film festivals research and methodologies: Dialogues between film festival scholars and practitioners*, edited by D. Ostrowska and T. Falicov. Amsterdam: Amsterdam University Press, 2023b.
- Loist, S. and Samoilova, Z. 'Open Media Studies und Digitale Methoden: Zur Erforschung von Filmfestivalruns', *Zeitschrift für Medienwissenschaft*, Open Media Studies Blog, 8 May 2019a: <https://zfmediawissenschaft.de/online/digitale-methoden-und-open-media-studies>.
- 'Getting Started on the Film Circulation Project: Studying Film Festivals with Various Data Sources', *Film Circulation*, 29 October 2019b: <http://www.filmcirculation.net/2019/10/29/getting-started-on-the-film-circulation-project/>; <https://doi.org/10.5281/zenodo.7934466>.
- 'First Results from Our Survey of Filmmakers on How Their Films Traveled Through Festivals', *Film Circulation*, 9 January 2020: <http://www.filmcirculation.net/2020/01/09/first->

HOW TO CAPTURE THE FESTIVAL NETWORK: REFLECTIONS ON THE FILM CIRCULATION DATASETS

- results-from-our-survey-of-filmmakers-on-how-their-films-traveled-through-festivals/
<https://doi.org/10.5281/zenodo.7934509>
- 'Data Collection and Operationalization of Film Festival Categories', working paper, Zenodo, 2021, <https://doi.org/10.5281/zenodo.7933399>
- *Film Circulation*, dataset [Data set], Zenodo, 2023a. <https://doi.org/10.5281/zenodo.7887672>.
- 'Evidenzbasiert, nicht datengetrieben: Herausforderungen im Einsatz quantitativer Forschungsmethoden im Feld der Festivalforschung' in *Produktionskulturen audiovisueller Medien: Neuere Perspektiven der Medienindustrie- und Produktionsforschung*, edited by S. Udelhofen, D. Göttel, and A. Riffi. Wiesbaden: Springer VS, 2023b (forthcoming).
- Loo, M. 'The Stringdist Package for Approximate String Matching', *The R Journal*, 6, 1, 2014: 111-122; <https://doi.org/10.32614/RJ-2014-011>.
- Manovich, L. 'Cultural Analytics, Social Computing and Digital Humanities' in *The datafied society: Studying culture through data*, edited by M. Schäfer and K. van Es. Amsterdam: Amsterdam University Press, 2017: 55-68.
- Naun, C. and Elhard, K.C. 'Cataloguing, Lies, and Videotape: Comparing the IMDb and the Library Catalogue', *Cataloging & Classification Quarterly*, 41, 1, 2005: 23-43; https://doi.org/10.1300/J104v41n01_03.
- Ostrowska, D. 'International Film Festivals as Producers of World Cinema', *Cinema & Cie*, 10, 14-15, 2010: 145-150.
- Peranson, M. 'First You Get the Power, Then You Get the Money: Two Models of Film Festivals', *Cineaste*, 33, 3, 2008: 37-43.
- Peirano, M. and Cruz, G. 'Festivales de cine en Chile', [dataset] Zenodo, 2022. <https://doi.org/10.5281/zenodo.6987256>.
- Petrychyn, J. 'Queering New Cinema History: Affective Methodologies for Comparative History', *TMG* 23, 1-2, 2020: 1-22; <https://doi.org/10.18146/tmg.588>.
- Samoilova, Z. and Loist, S. 'Film Circulation Project Questionnaire', Version 2019, Zenodo, 2019a. <https://doi.org/10.5281/zenodo.3581359>.
- 'Using a Feminist and Inclusive Approach for Gender Identification in Film Research' in *Complexities: Digital humanities*, edited by ADHO, Poster, 2019b: <https://dev.clariah.nl/files/dh2019/boa/0790.html> (accessed 9 September 2019).
- Stevens, K. 'Across and in-Between: Transcending Disciplinary Borders in Film Festival Studies', *Fusion*, 14, 2018: 46-59; <http://www.fusion-journal.com/across-and-in-between-transcending-disciplinary-borders-in-film-festival-studies/>.
- Stier, S., Breuer, J., Siegers, P., and Thorson, K. 'Integrating Survey Data and Digital Trace Data: Key Issues in Developing an Emerging Field', *Social Science Computer Review*, 38, 5, 2020: 503-516; <https://doi.org/10.1177/08944393198443669>.
- Vallejo, A. and Peirano, M. 'From the Field to the Database: Combining Methods in Film Festival Research' in *Film festivals research and methodologies: Dialogues between film festival scholars and practitioners*, edited by D. Ostrowska and T. Falicov. Amsterdam: Amsterdam University Press, 2023 (forthcoming).
- Vallejo, A., Nerekan, A., Vicario, B., and Fresneda, I. 'IKERFESTS dataset. Festivales de cine y audiovisuales en Euskal Herria / Euskal Herriko zine eta ikus-entzunezko jaialdiak', [dataset] Zenodo, 2022: <https://doi.org/10.5281/zenodo.6346570>.
- van Oort, T., Jermudd, A., Lotze, K., Pafort-Overduin, C., Biltereyst, D., Boter, J., Dibeltulo, S. et al. 'Mapping Film Programming Across Post-War Europe (1952)', *Research Data Journal for the Humanities and Social Sciences*, 5, 2, 2020: 109-125; <https://doi.org/10.1163/24523666-00502009>.
- van Oort, T. and Noordegraaf, J. 'The Cinema Context Database on Film Exhibition and Distribution in the Netherlands: A Critical Guide', *Research Data Journal for the Humanities and Social Sciences*, 5, 2, 2020: 91-108; <https://doi.org/10.1163/24523666-00502008>.
- van Vliet, H. *Festival atlas 2017: Film festivals*. Amsterdam: Amsterdam University of Applied Sciences, 2018.

- van Vliet, H., Dibbets, K. and Gras, H. 'Culture in Context: Contextualization of Cultural Events' in *Digital tools in media studies: Analysis and research*, edited by M. Ross, M. Grauer, and B. Freisleben. Bielefeld-New Brunswick: transcript, 2009: 27-42.
- Vanhaelemeesch, J. 'Common Ground: Film Cultures and Film Festivals in Central America', PhD Thesis, Faculty of Social Sciences, Department of Communication Studies, University of Antwerpen, 2021: <https://hdl.handle.net/10067/1770370151162165141>.
- Verhoeven, D. 'Show Me the History! Big Data Goes to the Movies' in *The Arclight guidebook to media history and the digital humanities*, edited by C. Acland and E. Hoyt. Sussex: REFRAME; Reframe Books in association with Project Arclight, 2016: 165-183.
- Verhoeven, D. 'The Library Is Open: Or Is It?' Keynote. VALA 2018 "Libraries, Technology and the Future", Melbourne, February 15, 2018.. Accessed June 12, 2023. <https://www.vala.org.au/vala2018-proceedings/vala2018-plenary-6-verhoeven/#>.
- Verhoeven, D., Loist, S., and Moore, P. '(Inter)Disziplinäre Routen und digitale Praxis: Kinomatics und die industrielle Geometrie der globalen Kinoforschung', *montage AV*, 29, 1, 2020: 173-184.
- Verhoeven, D., Moore, P., Coles, A., Coate, B., Zemaityte, V., Musial, K., Prommer, E. et al. 'Disciplinary Itineraries and Digital Methods: Examining the Kinomatics Collaboration Networks', *NECSUS_European Journal of Media Studies*, 9, 2, 2020: 273-298; <https://doi.org/10.25969/mediarep/15320>.
- Wais, K. 'Gender Prediction Methods Based on First Names with GenderizeR', *The R Journal*, 8, 1, 2016: 17-37; <https://doi.org/10.32614/RJ-2016-002>.
- Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Blomberg, N., et al. 'The FAIR Guiding Principles for Scientific Data Management and Stewardship', *Scientific Data*, 3, 1, 2016: 160018; <https://doi.org/10.1038/sdata.2016.18>.
- Zemaityte, V., Coate, B., and Verhoeven, D. 'Media Trade Beyond Country Borders: Five Types of Global Cinema Distribution', *SSRN Journal*, 2018: <https://doi.org/10.2139/ssrn.3228310>.

Notes

- [1] Peranson 2008; Gass 2009; Iordanova 2009.
- [2] Ostrowska 2010; Iordanova 2015; Falicov 2016.
- [3] Elsaesser 2005; de Valck 2007.
- [4] For an overview of film festival research see De Valck 2007; De Valck & Loist 2009; Iordanova & Rhyne 2009; Wong 2011; www.filmfestivalresearch.org/index/ffrn-bibliography.
- [5] Van Vliet & Dibbets & Gras 2009; Biltreyst & Meers 2016; van Oort et al. 2020; van Oort & Noordegraaf 2020; Manovich 2017.
- [6] <https://homernetwork.org/annual-conference/>; <https://kinomatics.com/>; Coate et al. 2017; Zemaityte & Coate & Verhoeven 2018; Verhoeven & Loist & Moore 2020; Verhoeven et al. 2020.
- [7] Stevens 2018; Petrychyn 2020; Vallejo & Peirano 2023.
- [8] www.filmcirculation.net; Loist 2020, 2023b; Loist & Samoilova 2019a, 2023b; Ehrlich & Burgdorf & Samoilova & Loist 2022.
- [9] For instance, in the local and regional projects on Chilean and Basque festival landscapes: Vallejo et al. 2022; Peirano & Ramirez Cruz 2022; Vallejo & Peirano 2023.
- [10] Wilkinson et al. 2016; For a development of FAIRER principles, which expand the concept from Findable, Accessible, Interoperable and Revisable to Ethical and Revisable, see Verhoeven 2018.

HOW TO CAPTURE THE FESTIVAL NETWORK: REFLECTIONS ON THE FILM CIRCULATION DATASETS

- [11] Loist & SamoiloVA 2023a.
- [12] Loist & SamoiloVA 2019b.
- [13] We would like to thank Florian Weghorn and Anne Marburger at the Berlin International Film Festival, Julien Westermann at the Clermont-Ferrand International Short Film Festival, Paul Struthers and Joe Bowman at Frameline: San Francisco International LGBTQ+ Film Festival, and Diana Sanchez at the Toronto International Film Festival for their help in arranging and providing festival data.
- [14] The survey questionnaire is accessible on Zenodo: SamoiloVA & Loist 2019a. First results for the survey responses to the 2013 sample are available in Loist & SamoiloVA 2020.
- [15] The scripts and detailed descriptions are available in the notes of the Zenodo dataset: Loist & SamoiloVA 2023a.
- [16] Loo 2014.
- [17] Ehrich et al. 2022.
- [18] Wais 2016
- [19] SamoiloVA & Loist 2019b.
- [20] Loist & SamoiloVA 2021.
- [21] Japac et al. 2015.
- [22] Groves 2011.
- [23] Golder & Macy 2014; Stier et al. 2020.
- [24] Loist 2020.
- [25] Barnes 2020.
- [26] World Bank. 'World Bank Country and Lending Groups', <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups> (accessed 15 May 2022).
- [27] European Countries present in the sample: Albania, Armenia, Austria, Belarus, Belgium, Bosnia & Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, France, Estonia, Faroe Islands, Finland, Germany, Greece, Iceland, Hungary, Ireland, Italy, Kosovo, Latvia, Liechtenstein, Luxembourg, Lithuania, Macedonia, Montenegro, Malta, Moldova, Monaco, Netherlands, Norway, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, Sweden, Switzerland, Spain, Ukraine, United Kingdom, Czechoslovakia, German Democratic Republic, North Macedonia.
- [28] North American countries present in the sample: Bermuda, Canada, United States.
- [29] Asian countries present in the sample: Afghanistan, Armenia, Azerbaijan, Bangladesh, Bhutan, Cambodia, China, Cyprus, Georgia, India, Indonesia, Japan, Kazakhstan, Kyrgyzstan, Laos, Malaysia, Myanmar (Burma), Nepal, North Korea, Pakistan, Philippines, Samoa, Singapore, South Korea, Sri Lanka, Taiwan, Thailand, Turkey, Uzbekistan, Vietnam.
- [30] Latin American and Caribbean countries present in the sample: Argentina, Aruba, Bahamas, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Jamaica, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Trinidad & Tobago, Uruguay, Venezuela.
- [31] MENA countries present in the sample: Algeria, Bahrain, Egypt, Iran, Iraq, Israel, Jordan, Kuwait, Lebanon, Morocco, Oman, Palestinian Territories, Qatar, Saudi Arabia, Syria, Tunisia, United Arab Emirates.
- [32] Pacific countries present in the sample: Australia, Fiji, Guam, New Zealand.

- [33] Sub-Saharan African countries present in the sample are: Benin, Burkina Faso, Cameroon, Chad, Congo - Brazzaville, Côte d'Ivoire, Ethiopia, Ghana, Kenya, Madagascar, Mali, Mauritius, Mozambique, Nigeria, Senegal, South Africa, Tanzania, Uganda.
- [34] Naun & Elhard 2005.
- [35] Hill et al. 2020.
- [36] A pilot study by Skadi Loist and Ann Vogel examined the film circulation for the films in competition for the Teddy Award at the Berlinale (2014). For this purpose, the festival runs of the films were collected via information from the films' websites, IMDb, and a survey.
- [37] Loist & Samoilova 2021.
- [38] See for instance the IKERFEST project on the Basque festival landscape (Vallejo et al. 2022) or the Festivales de Cine de Chile project, www.festivalesdecine.cl (Peirano & Ramirez Cruz 2022).
- [39] Vallejo et al. 2022.
- [40] Vanhaelemeesch 2021.
- [41] Peirano & Ramirez Cruz 2022.
- [42] <http://www.festivalatlas.nl/>; van Vliet & Dibbets & Gras 2009; van Vliet 2018.