

5. Constructing an Open, Participatory Database on Gender (In)Equality in the Italian Film Industry

Methodological Challenges

Rosa Barotsi, Mariagrazia Fanchi and Matteo Tarantino¹

ABSTRACT

Through CENTRIC (CrEative NeTwoRks Information Cruncher), an on-line search tool developed by the Digital Humanities Lab at the Catholic University of Milan, data was gathered by the authors on ten professional roles in Italian feature film productions from the institution of Legge Corona in 1965 and until 2020, with plans to update its findings year by year. The aim was to understand the development of gender equality dynamics in the film industry across this nearly 60-year timespan. Developed in the context of the Horizon2020 project CineAF, the CentricWEBI/CineAF database wishes to serve as a valuable resource for those who are intent on promoting gender equality in policy, film programming and educational curricula. As part of this goal, the ambition is for the database to be open and participatory. A Pilot project was conducted in Spring 2022 in order to test the usability of the platform, and its capacity to generate knowledge about women professionals and reflections on gender inequality. The article focuses on some of the challenges and ethical issues that were made evident thanks to this preliminary exercise, and suggests some possible solutions. These challenges include the ethical implications related to the collection of gender data; the meaningful application of FAIR principles in the humanities and social sciences; and the epistemological pitfalls of film industry data.

KEYWORDS

Gender equity; big data; knowledge ecology; audiovisual industries; film policy.

¹ Rosa Barotsi (University of Modena and Reggio Emilia) – rosa.barotsi@unimore.it; Mariagrazia Fanchi (Catholic University of Milan) – mariagrazia.fanchi@unicatt.it; Matteo Tarantino (Catholic University of Milan) – matteo.tarantino@unicatt.it. All authors contributed to the writing of this article. Rosa Barotsi wrote pages 99-102, Mariagrazia Fanchi wrote pages 88-93, and Matteo Tarantino wrote pages 94-99.

Introduction

What gender equality initiatives are possible and probable depends on what gender knowledge is produced and circulated, and which aspects of gender inequality that knowledge makes visible.

– Eikhof et al. 2019: 841

As Eikhof, Newsinger, Luchinskaya, and Aidley suggest in their analysis of gender inequality in the British screen industries, the availability of data on the gender gap is a necessary condition for planning actions that are able to generate real impact, but it is not a sufficient one. Indeed, as the authors highlight, how data are collected and communicated influences the possibility of launching effective policies supporting gender equality. In the UK, for example, at the time this essay was being written, the social discourses on inequalities tended to focus on the scarcity (or absence) of women in top roles,² generating an “ecology of knowledge” strictly focused on the numbers and percentages estimating the presence of women in different professional roles – and especially in the decision-making ones, but much more imprecise when it came to describing their working conditions, with the consequence of pushing political intervention more towards increasing the number of women employed in the screen industries than towards protecting them: increasing their salary, making their jobs more stable, improving the balance between professional and personal life, and creating the best conditions for overcoming the gender gap, starting from education and training paths.³

² There are, however, some important exceptions to this trend, such as the ‘Raising Films Project’. <https://www.raisingfilms.com/about/> (last accessed 13-04-23).

³ Iñaki Permanyer had carried out a similar analysis, some years earlier, on the criteria used by the European Institute for Gender Equality to measure the gender gap, show-

This paper seeks to problematise some aspects of the relationship between the “ecology of knowledge”, data, and gender equality policies, starting from the CENTRIC project.

CENTRIC (Creative Networks Information Cruncher) is a software developed by the Digital Humanities Lab of Università Cattolica in 2018.⁴ CENTRIC automatically collects and analyses data relating to Italian films – such as year of production, format, genre, production companies, and year of distribution. It also searches and archives data on film crews – at this time concerning ten primary professions: director, screenwriter, editor, producer, music director, director of photography, set design, costume and make-up supervisor, and special effects director –, used to calculate a set of concise indexes devoted to evaluating the gender balance (for instance the Gender Balance Score – Fanchi 2021; Fanchi, Tarantino and Barotsi forthcoming) and describing the level of integration of male and female professionals in the Italian film industry (for instance the gd-score – Fanchi and Tarantino 2019, 2021).

Following Eikhof et al. (2019), in order to understand the meaning and the value of the data generated by CENTRIC, and the representations of the gender gap in the Italian screen industry that it produces, it is essential to understand the specific “knowledge ecology framework” (Eikhof et al. 2019: 841) within which CENTRIC was developed. In other words, as for any other method and scientific approach, CENTRIC, and the kind of knowledge it produces, reflects a set of conditions and coordinates, historically contextualised, that have contributed to defining its priorities, focusing on specific aspects of the complex issue of gender balance rather than others, and that have influenced its solutions, including methodological and technical ones. In the next section, we will define the main theoretical and cultural coordinates that led to CENTRIC.

ing that the indexes used by the Institute tended to overestimate the gender gap in the countries whose economies were less developed and, conversely, underestimate it in the countries with stronger economies (Permanyer 2015).

⁴ <https://almed.unicatt.it/almed-ldh-lab-digital-humanities-fou-main-4970> (last accessed 26-03-23). CENTRIC was developed in 2018 by Matteo Tarantino and it was first used, by two of the authors, to collect and analyze the crews of 3542 Italian films (both shorts and feature films) that received, from 2004 to 2016, public funding as ‘cultural interest films’ (‘film di interesse culturale’) under the so-called Urbani Law (DL 28/2004) (Fanchi 2020, Fanchi and Tarantino 2021).

The CENTRIC Framework

The development of CENTRIC has been influenced by a complex set of factors: cultural, social, legislative and scientific.

In the second decade of the new millennium, the awareness of the profound gender inequalities in the cultural and creative industries, even in the most economically advanced countries (Conor et al. 2015), grew for both policymakers and scholars. Along with the commitment of the European Union and the Council of Europe on gender equality issues, this led to more systemic approaches to the analysis of the presence and roles of women in the screen industries. The ambition was not only to provide a snapshot of gender inequality, but also to promote pluralism and inclusiveness where culture is produced and to reflect on methods and tools that can overcome the absence of data on women's work. The tendency of economic systems to relegate women to the 'grey area' of the labour market, with temporary, informal or 'labile' contracts (Morgan and Nelligan 2015), in fact, makes the presence and contribution of female workers challenging to recognise. This is also because there is a lack of categories suitable to describe and detect the specific forms that women's work assumes. Barbara Montesi wrote:

Statistical survey, which has been and is, a source of primary importance for verifying the dimensions of women's access to different professions and their evolution over time has contributed to the codification of the bias for the 'awkwardness' often shown by statistical surveys in framing the multiple activities of women (Montesi 2002: 148).

In that climate, the new data-driven approaches seemed to represent a veritable turning point; they offered, in fact, the possibility to disclose otherwise invisible correlations, shedding light on the patriarchal mechanisms that still operated in advanced economies and to identify new and more perspicuous categories capable of representing women's work. Mark Lutter's well-known study of the casts of 350,000 U.S. film productions between 1929 and 2010, for example, was welcomed by the scientific community as a truly innovative approach. Using data-driven methods and, especially, Social Network Analysis, Lutter demonstrated that the different career possibilities for the actresses and actors employed in Hollywood cinema depended on the kind and 'quality' of the professional relationships they were able to activate. For instance, Hollywood actresses had less linear job trajectories than their male peers and significantly higher dropout rates when they were

working in ‘closed’ creative teams composed of professionals who tended to collaborate exclusively with each other. Conversely, the risk of failure and career interruption was lowered for actresses involved in projects with ‘open’ creative teams.

Lutter’s essay showed the relevance of ‘relationship capital’ as a resource for female professionals against segregation and gender inequality, highlighting the opportunities of an alternative methodology to the simple counting of men and women employed or represented in film or television (Verhoeven et al. 2020). It also demonstrated the fruitfulness of an approach that did not focus exclusively on what was happening onscreen but on the complex offscreen dynamics.

The focus on the offscreen dimension, at that time, was also supported by a general shift in the scientific paradigm and by the turn from the idea of cultural production as an individual initiative to a collaborative model, typical of digital cultures (Hartley et al. 2015); in this perspective, the absence of women or talents belonging to socially or historically minoritised groups in production initiatives was presented not only as a problem of social justice but as a weakness in the functioning of the media system, with detrimental effects on the cultural and creative industries (Graham and Gandini 2017). The persistence of discriminatory mechanisms, in fact, undermined the principle of ‘everyone’ – which, according to Hartley, Wen and Li (2015), is one of the new millennium’s creative industries’ three pillars, alongside the principles of ‘everywhere’ and ‘everything’, on which their ability to be innovative and generate value depends.

The CENTRIC project took shape at this particular conjuncture, also prompted by the approval at the end of 2016 of the long-awaited new Law on Cinema and Audiovisuals (Law 220, also known as the Franceschini Law), that replaced the previous law, approved in 1965 (Law 1234, known as the Corona Law), and by the absence, in the text of the Law, of any reference to gender equity policies (Fanchi 2020). Even if, in the following months, a first set of incentives for gender equality were introduced, through some implementing decrees (07/31/2017 and 11/09/2017), at the time of its approval, the Law appeared inexplicably indifferent to the issue of the gender gap.

The CENTRIC project took shape within this framework, setting itself three main objectives:

1. establish a sustainable continuous survey system capable of recording women’s presence in the Italian cinema and audiovisual sector and its changes;

2. extend the focus to a range of professions beyond director or screenwriter, to trace the contribution of female professionals to technical roles as well;
3. focus on relationships, measuring the quality and duration of collaborations and more generally the working conditions of women and men in the Italian cinema and audiovisual sector.

However, during the development of the software devoted to data collecting and management, CENTRIC was faced with some new challenges. These challenges were connected, on the one hand, to the quality and completeness of the available information on the workers employed in Italian cinema; and, on the other hand, to the crucial issue of the automatic gender attribution to professionals, and the ethical and political matters it raises. Moreover, the scientific and disciplinary debate, after an initial moment of uncritical enthusiasm for the use of big data, began to detect some critical, even profound, issues linked to the use of data-driven approaches in studies aimed at assessing and overcoming gender inequalities (D'Ignazio and Klein 2020). In particular, with specific reference to the application of such approaches to the field of cinema and audiovisuals, two main issues emerged: the risk of a regression to a biologically deterministic and binary approach, even within frameworks that do not embrace the gender binary perspective (Eikhof et al. 2019, Guyan 2021); and the inability of data-driven approaches to reflect intersectional dynamics (Cullen and Murphy 2018, Jansson and Wallenberg 2020).

Two additional facts helped direct the project and define its particular 'ecology'.

The first one was the beginning of a collaboration with the Italian Ministry of Culture for the assessment of public aid to cinema and audiovisual media. In 2019, Università Cattolica, together with the consulting firm PTSCLAS, was awarded the tender for the analysis and evaluation of the impacts of aid provided by the Italian State to screen industries within the framework of Law 220/2016.⁵ The evaluation model included, in addition to the analysis of industrial and economic effects of public funds

⁵ "Valutazione di impatto", *Direzione Generale Cinema e Audiovisivo* <https://cinema.cultura.gov.it/comunicazione/eventi-e-pubblicazioni/valutazione-di-impatto/> (last accessed 26-03-23).

to the sector, the analysis of social and cultural impacts too, including the ability of the State to encourage, through its various measures, the overcoming of the gender gap in screen industries. This project led to the development of specific sections dedicated to the topic within the Annual Report of ‘Valutazione di Impatto Legge Cinema’, and to the establishment of the ‘focus’: ‘Gender Balance in Italian Film Crews. Data and Research Policies’, an annual Report, first presented in 2021 at the Venice Film Festival, aimed at measuring gender balance in Italian production initiatives, film and audiovisual works, that received clearance for distribution in Italy from 2017, when the new Law came into force.⁶

The second variable, which intervened significantly in CENTRIC’s ‘knowledge ecology’, was the launching, in 2020, of the project ‘A Cinematic Archive for the Future: Women’s Films in Italy, 1965-2015’ (CineAF). CineAF was funded by Horizon 2020 under the Marie Skłodowska-Curie Actions grant.⁷ It was devoted to making visible the contribution of women professionals to the history of Italian cinema, creating a broader awareness that could inspire interventions, in terms of policy and beyond, for the promotion of gender equality in the audiovisual field.

The meeting between CENTRIC and CineAF allowed for an improved approach to gender imbalance in Italian cinema, developing a methodology nestled within the emergent data-based research framework and combining quantitative and qualitative methodologies with feminist research principles and a humanities perspective.

⁶ “Focus parità di genere”, *Direzione Generale Cinema e Audiovisivo* <https://cinema.cultura.gov.it/per-gli-utenti/approfondimenti-tematici/focus-parita-di-genere/> (last accessed 26-03-23).

⁷ The project took place between 2020-2022 and was conducted by Rosa Barotsi under the supervision of Mariagrazia Fanchi, with the collaboration of Matteo Tarantino. One of CineAF’s objectives was to investigate the presence and work of women directors in the Italian film industry between 1965 and 2015, although this was eventually extended to the present. Specifically, at least at this stage, the research focused on women-led feature films that obtained clearance (nulla osta) for national theatrical distribution. <http://cineaf.eu/> (last accessed 04-04-23). This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant agreement No. 891966.

CENTRICwebi

Starting from these premises and taking advantage of the favourable conditions described above, the CENTRICwebi website was launched in 2021.⁸

CENTRICwebi makes available data extracted from CENTRIC and archived in CineAF, allowing it to be queried through a system of filters and visualizations. CENTRICwebi includes two main domains of data:

- Data regarding the audiovisual works, including title, unique identification, production companies, year of production, length and format, nationality, genre, as well as a variety of scores (mostly on gender balance) calculated by the system; for a part of the corpus, the dataset also includes box office and production costs.
- Data on the production crews, focusing on ten professional roles: director, screenwriter, producer, editor, director of photography, heads of department in music, costumes, makeup, set design, and Special Effects, and including such information as the number of professionals in each role, name/surname, and gender; the latter is automatically assigned by the software based on the professional's first name. We will return to this point, which is decisive for the meaning and value of the CENTRIC/CineAF 'ecology of knowledge'.

As of October 13, 2022, the database contains 132,882 entries for professionals, corresponding to 47,554 individuals.

ATTRIBUTED GENDER	% OF TOTAL
Company	21%
Female	21%
Male	56%
Other/Unknown	2%
TOTAL	100%

TABLE I

Distribution of professionals in the database by attributed gender.

Source: CENTRIC/CineAF. Università Cattolica del Sacro Cuore, Milan.

⁸ Like the CENTRIC software, the CENTRICwebi platform was created by Matteo Tarantino. The intellectual property of the platform was acquired by Università Cattolica del Sacro Cuore. The platform is currently in beta version, and it can be accessed at the following address: <http://80.211.232.48:8000/home/> (last accessed 04-04-23).

As of 2023, therefore, the CENTRIC/CineAF project evolved into a tool with the following characteristics. First, it continuously monitors the crews of Italian films and audiovisual works, which received clearance for distribution (Law 161/1962) and, since November 2, 2021, the ‘classification of works for the protection of minors’ (Law 220/2016), from 1964 to the present.⁹ It extracts and cross-references the information from multiple movie databases. The data are automatically updated twice a month. At this point it automatically attributes gender according to binary classification. As gender information is not included as a field in online databases, genders are inferred through probabilistic estimates. Such estimates consider the given names and, when available, short biographies. We chose not to process images, even when present, for ethical reasons. Thus, the resulting “attributed gender” should be considered not as the gender of the *individual* but as that of the *name* (and pronouns used in the available biographies) that the individual employed in the credits of the work. We must keep in mind the scientific, ethical, and political limitations of this approach. At the same time, this raises the methodological question of what social indicators are available to perform any sizeable data-driven analysis of gender dynamics. We will return to this point in the last section. Finally, the system aggregates data into synthetic indexes, aiming to offer a more powerful lens for observing and understanding the features that discrimination based on gender takes in the Italian screen industries, reasoning not only in terms of numerical or percentage presences. Gender Balance Score (GBS) is the most important of these indexes. The GBS is calculated for each profession in each work; it measures the ratio of male professionals to females, ranging from -1 when 100% male to +1 when 100% female.¹⁰ In the current design of the system, only such aggregated indexes are visible to the general public.

Data collected in CENTRIC are managed in compliance with the FAIR guiding principles, particularly concerning issues regarding findability,

⁹ Our core official sources proceed from the Italian Ministry of Culture’s distribution/censorship clearance records (<https://cinema.cultura.gov.it/database-nulla-osta-film/>, last accessed 04-04-23) and datasets regarding films applying for public funding (<https://cinema.cultura.gov.it/comunicazione/eventi-e-pubblicazioni/valutazione-di-impatto/>; last accessed 04-04-23).

¹⁰ GBS operates according to the following formula: $(\text{women} - \text{men}) / (\text{women} + \text{men} + \text{unidentified})$. There is no qualitative dimension related to positive or negative values.

re-usability, and accessibility; all data can be downloaded and reused.¹¹ Furthermore, CENTRIC was conceived as a participatory archive. This choice has a twofold purpose. The first is to make the project independent of the researchers who launched it and thus enable it to continue over time, interweaving other ‘knowledge ecosystems’, and, in this way, improving the quality of data and their ability to inspire policies and actions capable of effectively supporting gender equality. The second reason concerns the nature of the data and the imprecision and lacunae of some information domains. As mentioned at the beginning of this paper, information about film workers is public and accessible but less complete and correct than one might imagine. The participatory approach allows for correcting these errors, integrating missing information, and acting on the crucial issue of automatic gender attribution.

Among the many issues we have addressed during the development of the project and dataset, we would like to focus on two specific points: the exhaustiveness of the data and the crucial issue of automatic gender attribution to professionals.

A Trace of the Instability of Online Film Information: The ‘Pilot Experiment Initiative’

Despite the abundant information about films (at least from the institution-alisation phase of cinema onward), cinema configures a field of “dirty data”: different countries produce fragmented institutional datasets of various types of films, rarely complete and mostly related to public funding; access to databases is slow and heavily regulated; there exists a plethora of private initiatives (including Amazon’s IMDB) that try to fill the void, offering partial, non-comparable and sometimes contradictory information. Especially, the further one moves away from Hollywood, the more inaccurate cinema becomes as a domain of data, even with reference to basic master data, such as the title: titles can be multiple, of course, in different languages, have different spelling, and so on.

¹¹ At the time this text was being written, a set of ethics guidelines and rules were being drafted to protect sensitive data, as required by law, and to ensure that the epistemological limits of the dataset (especially regarding automatic gender attribution) are clear to any users outside the research team.

To assess the completeness of the data collected automatically by CENTRIC and thus provide a measure of the (possible) inaccuracy of the available public cinema data, between February 22, 2022, and May 31, 2022, the *Pilot Experiment initiative* was launched.

During this Pilot Experiment, 267 editors, operating under the supervision of two of the authors, worked on 15 selected years (1964-1967 and 2008-2020, for a total of 3944 works or 34.5% of the total works and corresponding to 39% of the crew entries) manually checking the information collected.

Editors were assigned a package of works and manually double-checked all information entered with respect to those works and the corresponding crew by examining available information (such as biographies, interviews etc.). Each edit had to be manually approved by a supervisor.

This resulted in a total of 21,557 edits performed on the database. Crew edits (including modification, insertion, and deletion) amounted to 82.18%; these modifications accounted for 28.01% of the total. Thus around 36% of the crew information in the sample required corrections.

Gender was the core concern of the project underpinning CENTRIC. We expected a significant instability of gender information. This was not confirmed by the data. The overall rate of professionals whose attributed gender information was modified is the same for both attributed gender categories, at around 3.6%. Thus about 96.4% of the subsample had the appropriate gender attribution based on the available information and according to our method. However, we must always keep in mind that gender values could only be expressed through the Male/Female dichotomy and through an imperfect proxy, as the available information currently does not allow for efficient estimation through other options.

ATTRIBUTED GENDER	GENDER EDITS	TOTAL	% GENDER EDITS
Female	195	5,219	3.7%
Male	532	14,805	3.6%

TABLE 2

Gender distribution of professionals in the database whose attributed gender information was edited.

Source: CENTRIC/CineAF. Università Cattolica del Sacro Cuore, Milan.

Edits in themselves must be coupled with analyses of rejection rates, that is, the number of times a typology of edit was refused by the supervisor. In our case, gender showed a low degree of ambiguity with a rejection rate around

1%, whereas edits on profession and film title were rejected at much higher rates, peaking at 5.34% for film titles. However, in our case, the value of this data is severely skewed by an initial edit policy that limited gender edits to supervisors, due to the sensitivity of the information. Thus the overwhelming majority of gender edits were done by supervisors themselves, who tended to be well aligned and not to reject each other's edits.

At any rate, the instability of film titles shown by rejection rates should not be discarded as mere happenstance. Film titles are important identifiers, often used (also by CENTRIC) as a core index variable for cross-referencing film information proceeding from different data sources. In other words, we often recognize two works as the same because they share the same title.

As each database uses its own unique identifier for professionals and works, it is impossible to reference them across sources outright. Yet using titles for index variables runs into further complications: not only may there be films with the same titles released in the same period (this is particularly true for small independent films), but titles of the same works might be reported by different sources with differences small (for instance using apostrophes instead of accented letters) and large (for instance, changing the title completely from preproduction phase to the final release).

In general, index variables are the ones showing the higher instability in our dataset. Most of the crew modification edits regarded alignment of IDs of professionals or works, amounting to 78% of total modifications (see Table 3). This means the system failed to probabilistically attribute entries to the same professional (most notably because the professional's name was reported differently across sources, or because filmographies showed gaps – for instance for early-stage professionals) or professionals to the same work.⁴

FIELD MODIFIED	% OF TOTAL
Prof ID	64%
Film ID	14%
Gender	9%
Film Title	5%
Name	4%
Job/profession roles	4%

TABLE 3

Percentage of edits by field modified.

Source: CENTRIC/CineAF. Università Cattolica del Sacro Cuore, Milan.

While this, of course, points to the need to strengthen CENTRIC's probabilistic match algorithms, it also points to more significant underlying problems in index variables – and with them, of the stability of online data about Italian cinema. The work of alignment recorded in our databases is driven by the lack of unique, stable identifiers across the multiple available databases regarding both works and professionals. This significantly raises the operational cost of complex, big-data-driven analyses of the cinema industry.

Ethical and Methodological Questions Relating to Gender Attribution

As mentioned earlier, collecting data on gender identities that fall beyond the binary is a complex question for big-data research. If someone might interject that the number of people who fall outside the dimorphic genders is too small to matter, the answer would be obvious: it is precisely this exclusionary feedback loop that leads to such individuals being erased from data in the first place (Lindqvist et al. 2020: 332). In other words, we must be aware that data proceeding from a binary gender dichotomy may be trading in accuracy (and even discrimination) for efficiency. Additionally, until we find and implement an appropriate way to gather information on gender-variant people, we will not know the scale of that inaccuracy. For example, in the data we collected during the summer of 2021 from the Ministry of Culture on distributed feature films between 2017 and 2020, we made use of the gender binary to check for any progress made in the employment situation of women since the introduction of gender parity incentives through the new law. This meant that we may have failed to account for filmmakers such as Giuseppe Sciarra who, as we were in the process of data collection, came out publicly as non-binary.¹² On the other hand, if anonymization processes are less than perfect, the collection of data on gender-variant identities runs the serious risk of revealing information about people's gender without their consent. To speak of gender variant

¹² “Giuseppe Sciarra – ‘Non mi sento né uomo né donna’” 2014. <https://cinemaitaliano.info> (last accessed 26-03-23). In this particular case, our data collection was limited to feature films, so Giuseppe Sciarra would have not been part of our dataset since their recent output consists in short films. Sciarra's case is indicative, however, of the types of exclusions we are likely to operate if we continue to use the gender binary in data research.

film professionals in this context complicates the principle of “making the invisible visible”, if the terms under which that visibility is performed are not constantly negotiated with the people who are being made visible. In any case, it is important to remember that quantitative work on groups that have been systematically excluded involves “carrying a huge imbalance of power”, including the power to misgender or out people against their will (Matias 2014).

This is one of the challenges we came up against when deciding on our own strategy for data collection. Because we are partly dealing with historical archival data about thousands of people, a significant percentage of whom are no longer alive or traceable, we made the decision to assign gender based on first names and following the gender binary. This is far from ideal. Although we tried to guard against some of the problematic aspects of this method (as we will describe momentarily), it would be disingenuous to claim it is impossible to check with each and every person of a larger dataset to make sure their gender is represented correctly: a more honest explanation is that it is *too slow*, and therefore more labour – and resource – intensive. Especially in an academic environment that demands fast turnover of research results – both as a success fetish and as a material condition resulting from precarious, short-term research contracts – it is difficult to advocate for research methods that appear less than efficient. Such dilemmas bump up against a basic premise of feminist ethics – the fact that it takes more time.¹³

To return to our project, in order to counteract some of the problematic aspects of our chosen method, we make it clear that gender is assigned by the system and in reference to the name, not the person. To the extent that we have been able to manually check a sample of the dataset, the automatically-assigned gender was cross-referenced with official accounts of the film professionals, in magazine or newspaper articles, and, wherever possible, interviews or personal websites in which we hear their own voice, in order to determine how they self-identify publicly (for instance, through the use of pronouns or adjectives, which are gendered in the Italian language).

Although still imperfect, this method allows us to establish that these professionals were socially perceived, or presented as, women, and were

¹³ This is discussed, for instance, by the US-based filmmakers collectives Ethnocine and Rhiza in Episode 1 of their podcast, “Bad Feminists Making Films.” <http://badfeministmakingfilms.com/podcast> (last accessed 26-03-23).

therefore subject to the same biases and systemic exclusions as other women of their time. By assigning a gender based on the traditionally accepted gendered meanings of first names, in other words, we understand “woman” as an active category of social signification. Gender therefore functions not as a reflection of the gender identity of the individual, but as a term that condenses the norms and regulations that a patriarchal society applied to that individual.¹⁴

What this means, as already noted, is that we don’t have data on gender categories outside the binary. To make sure that this is understood as a lack in our research, we intend to note this absence in our text and figures. In the next steps of the project, we will integrate our approach to data collection with case studies that provide deep and textured stories of individual creators, to restore some of the context that accompanies our data.

Finally, and most importantly, the participatory nature of CENTRICwebi means that the professionals themselves will be able to make modifications and integrations to the collected data, including data on their gender, across their entire filmography or for specific productions, thanks to the dataset’s molecular approach (Fanchi, Barotsi and Tarantino forthcoming)

Many questions remain open: in order to render the CENTRIC/CineAF database open and participatory, one of the fundamental next steps is to establish a set of ethics guidelines relating to access and verification. This will undoubtedly involve complex decision-making. But, as with previous stages of the project, navigating those challenges as an interdisciplinary research team becomes part of the knowledge production process, and sharing them with other researchers in contexts such as this one is, as I hope you will agree, crucial.

¹⁴ Natalie Wreyford and Shelley Cobb follow a comparable logic, in that they suggest they attempted ‘to “read” a person’s gender identity the way a potential employer would’ (2017). Nancy López (2014) has developed the terms “street race,” “street gender,” and “street race-gender” as tools for use in the US census, with the capacity to separate between a person’s self-identity and the way other people, and therefore institutions and decision-makers, perceive them based on societal bias and contextual norms. A question she suggests might concisely account for the distinction between gender and ascribed gender as a census question might look like this: “If you were walking down the street, how would other Americans who do not know you personally identify your gender?”

Conclusions

CENTRIC defines a field of knowledge oriented towards providing solutions to the issue of the gender gap in the specific historical moment and framework in which the researchers found themselves operating. In this sense, as any other research project, it is physiologically limited and susceptible to additions, modifications, extensions, and changes. Additionally to those planned improvements that have already been mentioned, concerning, for instance, the question of automatic gender attribution or the sharing of ethical and political assumptions of knowledge, there are others with which researchers are now faced. For instance, the focus on the off-screen dimension, that characterised the early stages of CENTRIC's work, prompted by the need to compensate for the absence of data on gender balance in crews, requires an extension to the dimensions of onscreen representation and access to the cultural experience of spectatorship for all people. This expansion from the offscreen to the onscreen and in-front-of the screen, as well as the challenge of applying intersectional parameters to the discrimination mechanisms in screen industries, requires a never-ending redefinition of the perimeters and coordinates of CENTRIC's 'knowledge ecology', also by creating synergies with other projects that cover complementary areas.¹⁵

We are convinced that, in this way, it is possible to bring 'knowledge ecologies' to life, constantly evolving to reflect the needs and challenges of the present, and capable of offering an effective and powerful tool for the promotion of inclusiveness and cultural pluralism.

¹⁵ For example, the University of Bologna has developed a series of tools to assess gender balance in representations, with a focus on the imagery of TV series (Rocchi et al. 2023, Rocchi 2023).

BIBLIOGRAPHY

- Conor, Bridget, Rosalind Gill and Stephanie Taylor (2015). "Gender and Creative Labour." *The Sociological Review* 63(1 suppl): 1-22. <https://doi.org/10.1111/1467-954X.12237>.
- Cullen, Pauline and Mary P. Murphy (2018). "Leading the Debate for the Business Case for Gender Equality, Perilous for Whom?" *Gender, Work & Organization* 25(2): 110-26. <https://doi.org/10.1111/gwao.12199>.
- D'Ignazio, Catherine and Lauren F. Klein (2020). *Data Feminism*. Cambridge: The MIT Press.
- Eikhof, Doris Ruth, Jack Newsinger, Daria Luchinskaya and Daniela Aidley (2019). "And ... Action? Gender, Knowledge and Inequalities in the UK Screen Industries." *Gender, Work & Organization* 26(6): 840-59.
- Fanchi, Mariagrazia (2020). "Equality Illusions. Finanziamento pubblico del cinema e politiche di 'genere'." In *Il sistema dell'impegno nel cinema italiano contemporaneo*, edited by Dom Holdaway and Dalila Missero, 187-203. Milano: Mimesis.
- Fanchi, Mariagrazia (2021). "Equilibri ed equilibrismi. Bilanci di genere nella produzione cinematografica e audiovisiva italiana. 2017-2020." *8 e 1/2. Numeri, visioni e prospettive del cinema italiano* 60(12): 68-73.
- Fanchi, Mariagrazia and Matteo Tarantino (2021). "Has Digital Transformation Impacted Gender Imbalance in Italian Cinema? A Data Analysis of Creative Clusters 2004-2016." In *Digital Transformation in the Cultural and Creative Industries*, edited by Marta Massi, Marilena Vecco and Yi Lin, 162-84. London and New York: Routledge.
- Fanchi, Mariagrazia and Matteo Tarantino (2019). "Donne e reti creative nel cinema italiano (2004-2016)." *Economia della Cultura* 4: 523-30.
- Fanchi, Mariagrazia, Matteo Tarantino and Rosa Barotsi (2023). "Heaven Can Wait? Gender (Im)Balance in Contemporary Italian Film Crews." *Comunicazioni sociali* 1: 98-113. https://doi.org/10.26350/001200_000178
- Graham, John and Alessandro Gandini (eds) (2017). *Collaborative Production in the Creative Industries*. London: University of Westminster Press.
- Guyan, Kevin (2022). "Constructing a Queer Population? Asking about Sexual Orientation in Scotland's 2022 Census." *Journal of Gender Studies* 31(6): 782-792. <https://doi.org/10.1080/09589236.2020.1866513>.
- Hartley, John, Wen Wen and Henry Siling Li (2015). *Creative Economy and Culture. Challenges, Changes and Futures for the Creative Industries*. London: Sage.
- Jansson, Marie and Louise Wallenberg (2020). "Experiencing Male Dominance in Swedish Film Production." In *Women in the International Film Industry*, edited by Susan Liddy, 163-78. Cham: Palgrave Macmillan.

- Lindqvist, Anna, Marie Gustafsson Sendén and Emma A. Renström (2021). “What Is Gender, Anyway: A Review of the Options for Operationalising Gender.” *Psychology & Sexuality* 12(4): 332-44. <https://doi.org/10.1080/19419899.2020.1729844>.
- López, Nancy (2014). “What’s Your “Street Race-Gender”? Why We Need Separate Questions on Hispanic Origin and Race for the 2020 Census.” *Robert Wood Johnson Foundation*. https://www.rwjf.org/en/insights/blog/2014/11/what_s_your_street.html (last accessed 26-03-23).
- Lutter, Marc (2015). “Do Women Suffer from Network Closure? The Moderating Effect of Social Capital on Gender Inequality in a Project-Based Labor Market, 1929 to 2010.” *American Sociological Review* 80(2): 329-58. <https://doi.org/10.1177/0003122414568788>.
- Montesi, Barbara (2002). “Donne, professioni, cittadinanza.” *Storia e problemi contemporanei* 31(9): 147-66.
- Morgan, George and Pariece Nelligan (2015). “Labile Labour – Gender, Flexibility and Creative Work.” *The Sociological Review* 63(1 suppl): 66-83. <https://doi.org/10.1111/1467-954X.12241>.
- Nathan, Matias J. (2014). “How to Ethically and Responsibly Identify Gender in Large Datasets – MediaShift.” *MediaShift*. <http://mediashift.org/2014/11/how-to-ethically-and-responsibly-identify-gender-in-large-datasets/> (last accessed 26-03-23).
- Permanyer, Iñaki (2015). “Why Call It ‘Equality’ when It Should Be ‘Achievement’? A Proposal to Un-Correct the ‘Corrected Gender Gaps’ in the EU Gender Equality Index.” *Journal of European Social Policy* 25(4): 414-30. <https://doi.org/10.1177/0958928715594545>.
- Rocchi, Marta, Lorenzo Cattani and Guglielmo Pescatore (2023). “Gender Equality in European Netflix TV Series Production (2014-2019).” *Mediascapes journal* 21(1): 188-204.
- Rocchi, Marta (2023). “Women’s Labour in TV Series Production: A Comparative Analysis of Italian Generalist TV and Pay Platforms (2016-2019).” *Comunicazioni sociali* 1: 69-81. https://doi.org/10.26350/001200_000176.
- Verhoeven, Deb, Katarzyna Musial, Stuart Palmer, Sarah Taylor, Shaukat Abidi, Vejune Zemaityte and Lachlan Simpson (2020). “Controlling for Openness in the Male-Dominated Collaborative Networks of the Global Film Industry.” *PLoS ONE* 15(6): 1-23. <https://doi.org/10.1371/journal.pone.0234460>.
- Wreyford, Natalie and Shelley Cobb (2017). “Data and Responsibility: Toward a Feminist Methodology for Producing Historical Data on Women in the Contemporary UK Film Industry.” *Feminist Media Histories* 3(3): 107-32. <https://doi.org/10.1525/fmh.2017.3.3.107>.

CONSTRUCTING AN OPEN, PARTICIPATORY DATABASE ON GENDER (IN)EQUALITY IN THE ITALIAN FILM INDUSTRY: METHODOLOGICAL CHALLENGES



© The Author(s) 2023 <https://doi.org/10.21428/93b7ef64.9fa071ea>.

In Giorgio Avezzù and Marta Rocchi (eds). *Audiovisual Data: Data-Driven Perspectives for Media Studies. 13th Media Mutations International Conference*. Bologna: Media Mutations Publishing. <https://doi.org/10.21428/93b7ef64.ddf0a578>.