Datafication Genealogies beyond Algorithmic Fairness: Making Up Racialised Subjects

Ana Valdivia*
ana.valdivia@oii.ox.ac.uk
Oxford Internet Institute, University of Oxford
Oxford, UK

Martina Tazzioli*
martina.tazzioli@gold.ac.uk
Politics and International Relations, Goldsmiths University
of London
London, UK

ABSTRACT

A growing scholarship has discussed how datafication is grounded on algorithmic discrimination. However, these debates only marginally address how racialised classification or race categories are enforced through quantification and neglect its political and historical conceptualisation. In this work, we argue that literature partially fails to show that datafication reinforces racial profiling beyond the creation of racial categories as features. This article casts a new light on datafication by retracing its genealogy focusing on identification procedures in the colony and at the border. Such a genealogy foregrounds how datafication enforces racialised profiles by showing that it is part of a longer historical trajectory of modes of racialising individuals beyond algorithms and racial categories. Building on archival material, it develops this argument through two case studies. First, it focuses on the study of datafication of colonised bodies through biometrics by Francis Galton during the 19th-century. Second, it takes into account police identification procedures about unauthorised migrants, enforced by the French police at the Italian border in the 20th-century. These two cases show that although race categories as variables have been historically used to translate individuals into data, datafication processes as such also produce racialised profiles. A genealogical approach highlights continuities as well as quantitative and qualitative shifts between analogue and digital datafication. The article concludes arguing that datafication mechanisms have historically enforced legal and political measures by states in the name of science and objectivity and debates around algorithmic fairness should bring this key aspect back to the core of their critiques.

CCS CONCEPTS

• **Social and professional topics** → **History of computing**; • **Applied computing** → *Sociology*.

KEYWORDS

datafication, genealogies, racialised subjects, classification, borders

*Both authors contributed equally to this research.

1 INTRODUCTION

Our movements, bodies and actions are objects of multiple data extraction processes, enacted both by state and private actors; or better, they are transformed into data, they are datafied. Mayer-Schonberger and Cukier coined the term "datafication" arguing that "to datafy a phenomenon is to put it in a quantified format so it can be tabulated and analysed" [63]. To be object of datafication are not the individuals as such but, rather, their bodies and conducts enacted in a certain capacity: that is, we are datafied as political subjects, as workers, as welfare recipients, as commuters, as patients, as lovers, as newborns, as prisoners, and as border crossers. Yet, datafication is not only about translating social phenomena and people into data, it is also about crafting racialised subjects. Far from being a neutral and objective mechanism [57, 69], datafication has historically served for constructing, recording and classifying individuals through racialised profiles [28, 59, 61, 64, 77]. As a result, a burgeoning literature has stressed that datafication and algorithmic ecosystems raise questions of fairness and, thus, of discrimination across different contexts [8, 14, 17]. This is because in current debates datafication is associated with algorithmic-driven processes. Yet, this paper argues that, first, datafication per se should not be narrowed to digitalisation and thus, to algorithmic discrimination: an exclusive focus on algorithmic fairness de-historicises datafication as a recent phenomenon and conceptualises social injustices in terms of automated discrimination [16, 33]. Second, it contends that critiques of racial categories in algorithmic fairness [9, 27] should move forward by considering datafication as an historical mechanism that also reproduces racial profiling. In this paper we show that datafication has served to categorise, racialise and, thus, discriminate subjects beyond algorithmic systems and racial variables. By speaking of racialised mechanisms, we build on Alexander Wehelye's definition of racialisation which departs from biological understanding and designates it as "a conglomerate of socio-political relations that discipline humanity into full humans, not-quite-humans, and nonhuman" [92, p. 3]. Therefore, in this paper we do not understand racialisation as narrowed to race features fixed once for all: subjects who are racialised more than others and the racialised hierarchies of subordination and labour exploitation change over time and according to the context [25].

The production of racialised subjects through datafication does not necessarily require an algorithmic process or a racial feature. For this reason, we engage in a genealogy of datafication by retracing how current datafication mechanisms are part of a longer historical trajectory of racialising individuals through categorisation and, thus, reiterating forms of injustice and oppression beyond algorithmic processes. Building on critical works that mobilise the genealogical approach as a method to examine machine learning datasets from an historical and political perspective [27, 28], we retrace a genealogy of datafication, with a focus on the colony and the border, to unveil the racialised mechanisms of datafication. Genealogy is not synonymous with history, as it sheds light on the partial continuities as well as on the shifts and discontinuities between datafication processes in the present and in the past. Genealogy, as Michel Foucault fleshed out, is not about positing a linear historical development but, on the contrary, it consists in "disturbing what was previously considered immobile; it fragments what was thought unified", revealing "the heterogeneity of what was imagined consistent with itself" [35, p. 147]. That is, genealogy unfolds the instability and the contingent character of power relations and, at once, the possibilities of disrupting, altering those specific configurations of power and knowledge. Therefore, a genealogical approach investigates the continuities as much as the shifts that occurred between analogue datafication, digital datafication and racial categories. In this paper we do not assume analogue and digital as in clear-cut opposition to each other. On the contrary, we stress the partial mutual imbrications between the two, as "two modes of mediations" [38, p. 229], without however downplaying the fact that "the distinction analogue/digital is the media historical and theoretical guiding difference of the second half of the 20th-century" [79]. The analogue exists when there are similarities and proportions and the digital is "the capacity to divide things and make distinctions between them" [38, p. 229].[1] The genealogy of datafication processes we retrace in this paper equips us with a critical-analytical lens for understanding our present and, more specifically, the current border regime. A genealogical approach shows that datafication intrinsically can never be fair, as it is historically rooted in discriminatory and racialising functioning.

Drawing on Ian Hacking's "historical ontology" which investigates how kinds of people, classifications[2] and concepts have historically come into being, this paper argues that datafication is ultimately about making up racialised people [54].[3] "What is constructed is not only a certain classification" [53, p. 27]: more than that, subjects are socially crafted as a result of datafication. The emergence of concepts and of classification systems contribute to bringing into being specific kinds of people [24]. However, "ways of classifying human beings interact with the human beings who are classified" [53, p. 31]. That is, following Ian Hacking, the point is

not only about how certain kinds of subjects are produced through datafication and classification but also to grasp how these racialised subjects are the result of specific interactions with institutions, policies and laws [53, p. 27]. Indeed, each classification reflects given socio-economic rationales and desirable subjects (e.g. classifying criminals, ranking nationalities, or distinguishing between deserving and undeserving migrants). A similar point has been raised by Science and Technology Studies (STS) scholars who stress that the performativity of classification and categories depends on networks of practices: indeed, the realities generated through classification "are only real in particular networks or systems of circulation. This means, counter-intuitively, that realities are not real outside the chains of practices that perform them" [62, p. 242].

This paper retraces a genealogy of datafication by looking at the ways in which people's mobility has been datafied in the colony and at the border before the digitalisation era. Specifically, we focus on two case studies that foreground how datafication historically has been used to produce racialised subjects as well as racialised hierarchies of mobility: the British empire profiling through biometrics and statistical methods in the colony and police identification practices at the French-Italian border. The reason why we focus on the colony and the border is because these are contexts where the production of racialised profiles through datafication emerges blatantly. In the colonies datafication through biometrics has been used since the very beginning for creating racialised profiles of deception and risk [82]. The border is the site per excellence at which datafication is enacted by state authorities to identify people and to multiply hierarchies of (un)desired mobility. In addition to this, our twofold focus on the British colonies and the French-Italian border is particularly indicative due to the longstanding colonial histories and legacies of France and the UK. Indeed, the act of classifying and, relatedly, racialised undesired subjects, was key in both French and British colonial and postcolonial contexts. As Simone Browne has put it, "it is at the border —territorial, epidermal, and digital— a site where certain bodies are cast out and made out of place, that a critical biometric consciousness can be forged" [16, p. 129]. In a similar vein, we suggest that the colony and the border are emblematic contexts for retracing how historically datafication has been used for producing racialised subjects through identification procedures. The socio-legal production of "migrants" and of migration categories is inherently a racialising procedure given that "ideas of 'race' closely and easily articulate with ideas of 'nationhood', racism is — and has historically been — central to the construction of the figure of 'the migrant'" [80, p. 2]. Through such a genealogy, the paper unfolds that an exclusive attention to automated discrimination, algorithmic fairness and race categories overshadows how datafication is enshrined/incorporated in historical mechanisms for profiling and dividing populations, as well as for producing, which also discriminates.

The article builds on archival material we have collected at the National Archives in London (UK) and at the Departmental archives of Maritime Alps in Nice (France). Through a formal petition to the University College London (UCL) Library Special Collections, we asked to provide accreditation and visit The National Archives to view Galton archives. This collection contains more than 5,000

---

[1]Thus, by defining the digital in terms of making discrete units, Galloways challenges the idea artificial intelligence and computers are the essence of it: "the logic gate and the computer are merely the latest in a long stream of digital technologies that would begin with the integers, the alphabet, or even the atom, the synapse, the gene" [38, p. 229].

[2]Throughout the paper we use the term "classification" to encompass both scientific classifications and social science ones.

[3]Hacking defines historical ontology as an approach "concerned with objects or their effects which do not exist in any recognizable form until they are objects of scientific study" [54, p. 11].

items about Galton's work on eugenics, statistics and biometrics.[4] The Departmental Archive of Maritime Alps in Nice is public and, therefore, no special authorisation was needed. We selected these two specific case studies because Galton's work unveils the colonial legacy of the datafication of racialised bodies and at the French-Italian border there is a longstanding history of migrant crossings and, as we illustrate, of datafication of unauthorised mobility. Far from providing an exhaustive genealogy (of datafication), first archives are based on a substantial "piecemeal partiality" [84, p. 43], as they are always the result of selective knowledge production; and it is precisely this partiality that is of interest to us, given that "archival records can be read both for what they say and for their silence" [18, p. 189]. The archival records of the French police reveal that many migrants' passages remained undetected. Second, archives are not just repositories; rather, they are aspirations and artefacts [2]. Galton's archive clearly highlights this point, as it does not fully correspond to what was implemented on the ground but, rather, it also showcases aspirations and scientific projects — in this case of profiling through biometrics.

The contribution of this paper is as follows: we analyse the literature on datafication and racialisation within fairness and migration which contextualise our discussion. We introduce the concept of the racialisation mechanism of datafication which exposes that critical debates should move beyond the idea that algorithmic-driven datafication or racial categories discriminate. We then expose our two archival research on British colonies and the French-Italian border. The first encounter with Galton's collection shows how colonised bodies were datafied in the 19th-century during the birth of eugenics and biometrics. The second encounter at the French-Italian frontier's archives unveils the making up of undesirable border crossers through datafication in the mid 20th-century, when border controls became systematic and datafication was fully incorporated in practices of mobility policing. The paper moves on by discussing what nowadays persists in datafication procedures and what has changed. It contributes to the FAccT community by illustrating the importance of situating digital datafication processes within a longer history that also considers analogue datafication. Indeed, the racialised mechanisms of datafication cannot merely be reduced to algorithmic-driven procedures nor to racial categories embedded in datasets. For this reason, a critical analysis of datafication implies moving beyond the debate on debiasing algorithms and scrutinising how datafication mechanisms are used for enforcing specific legal-political exclusionary measures. Thus, a genealogical approach to datafication unveils the historical and colonial roots of this mechanism and should consider past forms of racialisation that made up undesirable subjects.

2 THE RACIALISED MECHANISMS OF DATAFICATION

The forms of discrimination and exclusion enforced through machine learning have gained traction in literature on algorithmic fairness. Much has been said about how algorithms and datasets reproduce historical and current social and political inequalities [17,

19, 28]. Critical scholarship in this field has examined how automated discrimination is enforced through data, biases and algorithmic codes. Genealogies of datasets in machine learning have been proposed to conceptualise them as infrastructure and analyse how they have been historically created [27, 28, 61]. In this article we engage with this critical and cross-disciplinary literature on datafication, genealogies, machine learning and racial categories that conceptualises racialised mechanisms that underpin datafication and that are strengthened through it. By this, we mean modes of classification of human beings which are predicated on racial features and proxies. Yet, far from being stable these features change over time. Therefore, datafication of human beings — although not all forms of datafication — invisibilises and justifies discriminatory and exclusionary profiling that appear as value-neutral. This paper intervenes in debates about the datafication of mobility showing that the (re)production of hierarchies of mobility in the colony and at the border is paradigmatic of how datafication has been used to enforce racialising processes. In a similar way, we contend that a focus on the border highlights modes of identification through datafication and shows that this latter is not narrowed to matters of algorithmic fairness and discrimination: it is rather about making up racialised subjects. In this respect, we draw on literature that has demonstrated how "race is produced by invisible statistical procedures" [26, p. 2] as well as on scholarship that has retraced the racism that has historically underpinned anthropometry [30, 66].

2.1 The constitutive unfairness of datafication

Datafication as a political mechanism unveils forms of discrimination [75, 78]. However, an exclusive focus on algorithmic fairness ends up de-historicising datafication. In fact, some works have been done to conceptualise norms and values embedded in datasets from an historical perspective proposing a genealogical approach [27, 28]. As Braun and Hummel have remarked, "justice in connection with datafication relates to, but ultimately encompasses more than, solely fairness" [14, p. 1]. The problem with that is twofold. First, a de-historicised approach prevents us from seeing that the racialised and discriminatory effects of datafication in the age of digitalisation do not constitute radical break or novelty with the past: rather, what is specific of it is the increasing invisibilisation of racialised procedures [26]. Second, it overshadows the partial continuities and the differences between the past and the present; presentism invisibilises that the classification systems and categories used do have specific political-historical origins. More precisely, in relation to datafication, as Hacking has pointed out, it is worth noticing that "many of the modern categories by which we think about people and their activities were put in place by an attempt to collect numerical data" [53, p. 181]. Thus, we concur with Beer about "the need to contextualise our understanding of big data within the history of social statistics. That is to say that we need to place big data within the genealogy of social data of various types" [6, p. 1].

Yet, key elements of datafication, i.e. features and categories, are also systematically used for enforcing discriminatory criteria, although these are often invisibilised. The use of racial categories to feed algorithmic systems has been largely discussed in fairness scholarship [9, 58]. On the one hand, variables representing racial

---

[4]More details about Galton's archive can be found at: https://www.ucl.ac.uk/library/special-collections/a-z/galton (Last accessed January 30, 2023).

features are used to measure and evaluate algorithmic discrimination. On the other hand, the use of these variables perpetuates and reifies the concept of race. Therefore, Benthall and Haynes proposed the construction of race-like features through social and spatial segregation to avoid the use of race as such. Yet Denton et al. criticised this approach by arguing that their solution could potentially neglect that race "does exist in the world" and "is social constructed" [27, p. 503]. While we concur with their critical analysis of racial categories in algorithmic fairness frameworks, we suggest that it is essential to also take into account the racialisation of datafication. Historical and archival investigation on the datafication mechanism used in the colony and at the border showcases the relevance of bringing datafication at the core of a critique towards algorithmic fairness. Throughout the paper we use the term racialisation instead of race, drawing on Hanna et al. who stressed that race is not fixed once for all and it is rather socially constructed [27]. In this respect it is worth clarifying the fact that despite this paper draws on critical scholarship on racial statistics and categories, its main focus is datafication.

2.2 The making of datafied colonial bodies and border crossings

Classification serves to multiply racialised bordering mechanisms and hierarchies of mobility. By focusing on the socio-political context it is possible to grasp how individuals at times manage to dodge, twist or hack modes of identification through datafication [37, 93]. Datafication procedures have been gaining increasing traction in scholarly debates about border controls and migration [4, 7, 63, 70]. This is because datafication of movements plays a key role in migration governance [60]. Indeed, the economy of knowledge production about refugees and migrants is sustained and shaped by datafication processes, that is by turning migrants' bodies and mobility into data: "what is known, negotiated and targeted as migration is mediated by a plethora of data practices, including registering, enumerating, counting and estimating to storing, cleaning, imputing, extrapolating and anticipating" [76, p. 579]. National frontiers are sites where unruly mobility is subjected to scrutiny, containment and selection also by identification through datafication. However, scholarship on datafication at the border tends to draw attention to the nexus between datafication and discrimination [15, 64] and to warn against the obfuscation of discriminatory procedures enhanced through algorithmic-driven systems [91]. To be of interest to us is the border, conceived not merely as a physical site but also as a multiplier of inequalities and of hierarchies of mobility [71]. As Simone Browne has pointed out, "it is at the border — territorial, epidermal, and digital — a site where certain bodies are cast out and made out of place, that a critical biometric consciousness" could be developed [16, p. 129]. This paper intervenes in debates on the datafication of mobility situating this within a broader colonial genealogy of datafication of undesired racialised subjects, showing that datafication procedures at the border have been historically used for multiplying racialised hierarchies of migrants. Hence, the datafication of migrants' mobility at the border is part of a broader history of the use of anthropometry and racialised classifications on colonised populations [83].

Classification and identification through datafication enacts data-subjects. This takes place both by extracting data from individuals and using this for creating racialised profiles, and by individualising procedures that assign a given identity to a single person, combining biometric and biographic features [86]. Methodologically, this entails, starting "with the analysis of data practices as objects of research, rather than existing entities and phenomena that are datafied" [76, p. 582]. Instead of asking how colonised subjects or migrants are datafied in the colony and at the border, it is a question of shifting attention towards how individuals have been enacted as "colonised subjects" or "migrants" through identification and classification, which rely on datafying procedures such as fingerprints [82]. Yet, this does not mean that one process follows the other - for instance that migrants are brought into being exclusively through data-driven practices. Rather, it is a matter of registering the mutual interdependence between data practices and subject-making processes, that in turn depends on the historical-political origins of categories [36, 54]. Colonised subjects and migrants are partly enacted as such due to identification through datafication, that is because they are labelled, governed and known as "colonial subjects" or "migrants". At the same time, not everything is the result of datafication and which subjects fit in some categories (such as refugees) change over time taxonomy is adopted and tailored according to specific socio-political contexts, and as scholars have demonstrated it is used for excluding most from international protection — for instance by distinguishing between "genuine" refugees and migrants — and to multiply hierarchies of undeservingness [22]; likewise, they are re-adapted and re-crafted in response to migrants' tactics to dodge identification or to twist to their advantage classification systems, in order to obtain for instance the right to stay, as it is the case with the category of vulnerability.

However, by speaking of colonial genealogy of datafication we should be careful in not assuming a substantial continuity between the use of data in the colonies and in the present. That is, if on the one side datafication is grounded in a longstanding history, shaped by colonial knowledge, on the other it is key to attend the partially different functions of datafication through history. Moreover, and relatedly, this paper contends that a critique of datafication should not be exclusively focused on its colonial legacies and, rather, it should encompass how datafication processes are used, as Gray aptly pointed out, for extracting value by multiplying differences [51]. For instance, today datafication at the border is grounded on the biometric theory developed in the 19th-century and reiterates racialised hierarchies of mobility. However, far from being a matter of linear continuity, datafication processes are used to enforce migration laws and policies that target, select and contain border crossers according to criteria different from the past. Second, the core role of private actors in migration governmentality has unfolded the central nexus between datafication and value extraction processes [89].

3 THE COLONIAL LEGACY OF DATAFYING BODIES

Biometrics has been implemented over the last three centuries to identify and verify the identity of subjects. From inked fingerprints to modern facial recognition, body traits are extracted and transformed into data as trustworthy elements of human identity. Despite the attention that biometric systems have recently attracted in current debates on surveillance and migration control, scientific methods have been developed since the 19th-century to datafy and extract patterns of colonised and marginalised subjects by states and empires [65]. In this section we unveil the colonial legacy of biometrics and its datafication by analysing archival material that exposes how anthropomorphic measures were used as a racialisation mechanism.

Through colonisation, Britain extracted resources and imposed ways of governing in other territories, for instance in the Raj[5] [82]. Analogue datafication played a predominant role in this historical context to control and "civilise" the colonial population. The foundations of modern statistics, a theory that is inherited today within machine learning and pattern recognition, were laid in this colonial context [49]. The datafication of human body traces, such as fingerprints and facial profiles, were developed during this era to classify "human races" or identify colonised subjects or criminals. Francis Galton, a Victorian scholar, played a key role in this scenario by fueling biometrics, statistics and anthropomorphic measurements for the making of a "civic" society.

Galton's career can be splitted into two phases: (1) geography and (2) human heredity [48, p. 3]. During the first part of his career, Galton travelled to "wild countries"[6] in Africa [46]. As a colonial explorer, he visited slave markets and even felt seduced to purchase Circassian female slaves [48, p. 35]. However, the second period of his life is better well-known. Highly influenced by his cousin, Charles Darwin, Galton was obsessed with the idea of datafying and quantifying human pedigrees and characteristics for classification or identification processes. He coined and founded the field of eugenics "feeling that its principles ought to become one of the dominant motives in a civilised nation" [44, p. 322]. His main goal was to investigate the "making of civic worth in man" by understanding how good heritage traits were transmitted through generations [43].

Datafication, identification and classification mechanisms played an important role in Galton's work. He studied "the improvement of the human race" through data gathered by the philanthropist and sociologist Charles Booth [42]. Booth analysed poverty rates in London by classifying streets and social classes based on urban visible characteristics [12]. He proposed different dubious associations between social class and colour categories to produce a map of poverty based on these observations.[7] Questionably, "the lowest class, vicious and semi-criminals" were grouped into the black category — not surprisingly — while "the upper-middle, upper class and wealthy" were grouped into the yellow one. Galton used

Booth's datafication of socio-economic subjects and its racialised classification to prove that the inheritance of the social class to which parents belong is inherited following the statistical Normal distribution. Under the assumption that better societies are made of "talent" individuals and that "the brains of the nation lie in the higher of our classes" [42, p. 661], Galton claimed that the improvement of human race relied on "granting diplomas to a select class of young men and women, by encouraging their intermarriages, by hastening the time of marriage of women of that high class, and by provision for rearing children healthily" [42, p. 663]. In other words, Galton advocated for the promotion of marriages between upper class individuals — and "the younger the woman the better" — as they pass on the "talent" genes to their offsprings. He justified his theory through the use of data, science and statistical methods now widely used in machine learning. However, neither Galton nor Booth realised at that time that what it was demonstrated was a well-known sociological theory: privileged classes have more access to resources such as education, rather than having "talent" genes or being "smarter" [21].

Galton and his disciple Karl Pearson, established the scientific grounds of biometrics and modern statistics. Motivated by the study of natural inheritance and human characteristics, Galton collected large amounts of biometric data in the Anthropometric Laboratory that they founded at the UCL. Galton was profoundly inspired by William Herschel, a civil servant at the Raj who collected fingerprints of colonised subjects to avoid personation [56] and Henry Faulds who is considered the founder of fingerprint identification and collected fingerprints of Japanese and other nationalities for ethnology classification [34].

Through meticulous and manual analysis of thousands of fingerprints, Galton aimed at indexing and describing patterns on the fingerprints of subjects. Through a deep examination on thousands of fingerprints, Galton proposed to classify fingerprints taking into account symmetries, slopes, whorls and loops for the only reason of classifying human races through these symbols (see Figure 1). Following these classification guidelines, Galton transformed inked fingerprints in data and tables that categorised fingerprints through patterns and codes which could be considered a manual datafication process of fingerprint indexing. In the words of Dongus: "Galton's system is an early form of pattern recognition" [29] and without a doubt, his work could be considered the forerunner of today's pattern recognition in biometrics.

Galton was interested in classifying human races through these patterns. In *Fingerprints*, he wrote a whole chapter on "Races and Classes". He fingerprinted and manually extracted patterns of "English, pure Welsh, Hebrew, and *Negro*[8]; also some Basques" individuals which we had access to at the archives [40, p. 192]. There, we also observed that he collected fingerprints of other human categories such as *idiots*[9] (see Figure 2).[10] After printing 2,082 individuals and quantifying the number of patterns on each supposed race, Galton

---

[5]Term that refers to the British colonising rule in India from 1858 to 1947.

[6]"Wild countries" is an expression used by Galton in his notes and it is even used in the title of his book. We, the authors, reproduce this expression to show and emphasise the pejorative vocabulary used at that time by scientists.

[7]See: Charles Booth's map at https://booth.lse.ac.uk/map/16/-0.0173/51.4994/100/1 (Last accessed December, 14 2022).

---

[8]Note that this word was explicitly used by Francis Galton during his work. We do not comply with this expression, as we consider it a deeply offensive term.

[9]Idem.

[10]While we are unsure under which circumstances and ethical guidelines Galton gathered these fingerprints, our sole intention to reproduce this material is for research purposes and of public interest. By showing these fingerprints, we bring historical evidence that the datafication mechanism of biometrics is rooted in a colonial and racist past.

Figure 1: Pattern codes to index fingerprints proposed by Galton [39]. Although it may seem a neutral categorisation of fingerprints, Galton aimed to classify human races based on these patterns. Source: Galton, Perason and Penrose, UCL Library Special Collections at The National Archives, Collection: Galton, Perason and Penrose (London, UK).

Figure 2: Galton's booklets containing fingerprints of "Negroes, Jews and Idiots". Galton aimed at classifying *human races* through patterns on fingerprints. The fourth image shows fingerprints of a Jewish subject and the patterns manually identified by Galton. Source: Galton, Perason and Penrose, UCL Library Special Collections at The National Archives, (London, UK).

concluded that "there is no peculiar pattern which characterises persons of any of the above races" [40, pp. 192-193].

Influenced by Bertillon, a French police officer who developed anthropological identification methods for criminals, Galton also studied the use of fingerprints for identification and search purposes. Arguing that individuals in the colonies were illiterate[11], deceitful[12] or similar looking[13], he advocated for the use of fingerprints to enhance the efficiency of identification processes in the British colonies: "beginning with the simplest requirement, of being assured that a particular person is really the man he professes to be, it has become recognised in India that the impression in ink of one or more fingers is an admirable criterion of identity, being cheap, easy and most trustworthy" [41, p. 119]. Galton also explored secondary means of anthropomorphic identification such as human profiles. Absorbed by the idea of datafying the human body, he proposed a formula based on anthropological measurement of the nose, lips and the curve of the chin. In *Nature*, Galton claimed that the use of numerical values is more appropriate than "vague adjectives" [45, p. 127] (see Figure 3). Thus, the use of this datafication mechanism, Galton argued, could enhance police investigations and criminal identifications through a numerical catalogue of criminal profiles — rather than qualitative descriptions.

This archival research showcases that racialised mechanisms of datafication procedures through fingerprints began at the colony to rule the colonised. In the words of Simon A. Cole "the British contribution to the development of fingerprinting occurred not at home but abroad" [20]. Galton, the scientist that found the roots of

[11]"The need for rapid means of identification is greatly felt in these two countries [India and Egypt]. The natives are too illiterate for the common use of signatures. Alphabetical registers are of little service, owing to the paucity of different names, and in Egypt, owing to the various ways in which a man fairly describes himself" [41, p. 118].

[12]"In India and in many of our Colonies the absence of satisfactory means for identifying persons of other races is seriously felt. The natives are mostly unable to sign; their features are not readily distinguished by Europeans; and in many cases they are characterised by a strange amount of litigiousness, wiliness and unveracity" [40, p. 149].

[13]Herschel while visiting Galton's biometric laboratory: "The uniformity in the colour of hair, eyes, and complexion of the Indian races renders identification far from easy, and the difficulty of recording the description of an individual, so that he may be afterwards recognised, is very great" [40, p. 150].

modern statistics and was tempted to buy female slaves during his travels, strongly advocated for the use of biometrics as means of a trustworthy identification mechanism to be imposed on colonised subjects deemed to be "deceitful" and "wiliness". This practice was then extended to other undesirable subjects at home such as prisoners. Moreover, this archive brings evidence of the connection between science, racism and the British empire through the datafication of human bodies that began at the colony and was extended to the birth of eugenics. In the 19th-century, datafication of bodies and other socio-economic features were instrumentalised by scientists and states by imposing their despotic theories and rules to govern undesirable and marginalised subjects [50]. In Koopman's words: "data became tethered to conceptions of discoverable fact and demonstrable knowledge" [61, p. 178]. An effect that undoubtedly resonates with nowadays large-scale biometric databases for migration control in the European border regime.

4 GENEALOGIES OF IDENTIFICATION TECHNOLOGIES AT THE BORDER

The French-Italian border has historically been a key crossing point and since 1861, when the frontier between France and Italy was established, it progressively became a highly policed frontier. Illegalised border crossers from different nationalities started to be systematically identified and classified by the French police at the Italian border in the late 1920s and in a more robust way in the late 1930s with the arrival of Jewish escapees from Eastern Europe. Here we draw on archival records stored in the Archive of Maritime Alps in Nice to retrace how the unauthorised border crossers were datafied at the border, after being stopped by the French police. The unauthorised border crossers at that time were Italian fugitives (communists, socialists, partisans), Italian workers without permit to expatriate, Jewish refugees (1938-1941) and, in the 1950s and 1960s, political dissidents from Romania, Bulgaria, Hungary
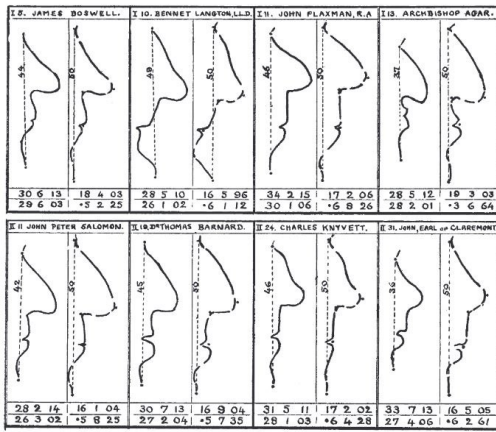
Figure 3: Galton challenged the accuracy of descriptive methods to identify criminals. In 1910, he introduced a method to quantify human profiles: "The replacement in all scientific work by numerical values, in the place of vague adjectives, is again of first class importance". Source: [45, pp. 127–129].

Poland and Yugoslavia [55]. Later on, in the late 1970s, the first migrants from North African countries and from the Middle East had been found by French police as well as by both Italian and French mountain rescuers [72]. Importantly, the possibilities of tracing a genealogy of datafication depends on the data that French authorities collect and extracted from unauthorised border crossers and, therefore, is a partial and incomplete one. Indeed, as it is the case at any frontier, many migrants were not detected by the police and, therefore their passage does not appear in any archive [84]; and, at the same time, it is a genealogy of datafication that unavoidably relies on modes of "seeing like a state" [81], since it is based on state-practices of datafying mobility at the border. In this respect, for the purpose of this paper, it is worth noticing that police practices of racialisation through classification do have a consolidated colonial genealogy in France - both in the colonies and in France, towards former colonised populations, such as Algerians [5, 11]. Hence, a genealogical approach foregrounds historical partial continuities between racialisation through datafication at the border and in the urban (post)colonial space [73].

As illustrated in the picture below, the datafication of unauthorised border crossers conducted by the French police in the last century was quite similar to the current procedure. Indeed, the identification form filled in by French police officers included personal data of migrants, (name, nationality, gender, age, among others), their declared final destination and reasons for coming to France and their inked fingerprints. Nowadays fingerprints at the border are usually taken with digital scanners, and not on paper. Usually when migrants are pushed back to Italy they are fingerprinted by the Italian police but not by the French authorities. This is because these latter adopt a practice of not-recording, in order not to be responsible for their asylum application, and they are only given an expulsion order on paper, called *refus d'entrée*, that contains basic personal information — such as name, nationality, age, gender, documents held and the reason why entrance to France was denied [87].

In the 1950s, French authorities produced both monthly statistics related to the unauthorised entry of people from Yugoslavia and from Eastern Europe and detailed files related to individual cases — which often included migrants' stories, their routes and information about smuggling networks used. It can be argued that the lack of digitalised identification technology — such as digital biometric systems — rendered necessary a more exhaustive verbalisation than nowadays when comparable detailed information about migrants can be found only in asylum application files.
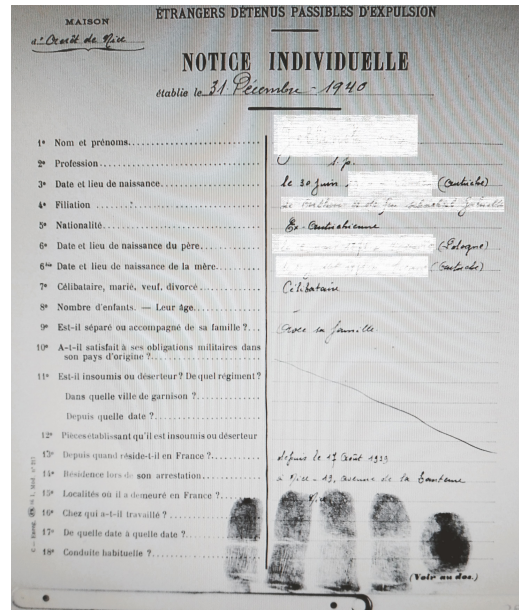

Figure 4: Form used in 1940 by the French police at the Italian border to identify and register people who crossed without authorization. The file contains information about nationality, job, date and place of birth of the parents, whether or not the person travels with family, working place, place of residence, fingerprint and documentation that can prove if the person is a deserter or a political disobedient. Source: Departmental Archives of Maritime Alps (Nice, France).

A close reading of missives and letters exchanged between local and national authorities and police officers, stored in the archives enables analysing "along the archival grains" [84]. More precisely, it allows unveiling what happened behind the scenes and in between the folds of states' diplomacy and inter-state bilateral agreements. By drawing attention to archival records that report how fugitives were classified, it becomes possible to reconstruct broader inter-state histories and geopolitical contexts. In one of the many letters exchanged in 1938 between French authorities about the increasing number of Jewish who enter the French territory from Italy, the Ministry of the Interior warned the Prefect of Maritime Alps about the importance of "assuring an effective surveillance of the French-Italian border". In order to enforce it, the French government "restricted the access to the territory to a certain number of crossing points, on the basis of geographical conditions" and ordered to surveil both roads and mountain paths. What was the

Figure 5: Example of datafication process at the French-Italian border in 1959: alongside individual identification forms, the French police used to make a list of migrants readmitted to Italy and of those readmitted to France by Italian authorities. Source: Departmental Archives of Maritime Alps (Nice, France).

reason of such governmental worry? Between 1938 and 1940 Italian Jewish as well as Jewish people who were coming from other countries (Germany, Austria and Poland, in particular), were trying to reach France and claim asylum there. In order to further hamper the legal entry of fugitive Jewish into France, at the start of 1938 the French government imposed the obligation for German and Austrian citizens to get a visa, together with their passport. Later that year, the same obligation was enforced on Italian citizens. Thus, first, France did not simply start to classify and criminalise migrants: the making of illegalised migrants was the result of decrees and laws enforced by the French government for regaining control over bodies out of place. Second, the increasing governmental anxiety towards Jewish migration did not turn into a mere sharp division between natives (French citizens) and migrants (foreigners).

Rather, a hierarchy of undesired mobility and presence was created by differentiating the legal requirements that people had to comply with in order to enter France legally: while Austrian, German and Italians had to obtain a French visa in addition to the passport, for other European nationals — like British, Danish, Portuguese, Swedish and Dutch — only the passport was required, and for others — like Swiss and Belgium citizens — an ID card was sufficient to enter the French territory. Later on, in the 1960s, French authorities could not push back Yugoslavs citizens because they destroyed their passports before crossing the border. In this regard, it is worth reporting what a missive of the French police from 1966 said: "it seems that many people who enter France through the crossing point along the coast, from Ventimiglia, while they wait

to fill in the form at the custom, and taking advantage of the usual long queues, seize the opportunity to get rid of their passport by throwing it into the public toilet of the custom office [. . .]. And yet, it seems difficult to constantly monitor that public toilet; the only solution turns out to be shutting down the public toilet in the custom office, but this would raise complaints".

Nevertheless, datafication of undesired mobility at the border should not lead us to conclude that state authorities conduct exhaustive controls and registration practices. On the contrary, both nowadays and in the past this is characterised by partial non-recording and non-registration [74]. This is because of migrants who manage to eschew border checks as well as because of states dodging legal and political responsibilities, as it is the case today with the Dublin Regulation, that determines which member state is in charge of examining an asylum application. In order to establish this, member states are obliged to fingerprint migrants upon entry and to store their biometric data in a European database called Eurodac. The goal of such data storing is not only to determine the identity of the person here and now but also to be able to do it in the future, irrespective of the actual physical presence at that moment, the state responsible to assess their asylum claim. Eurodac stores fingerprints of border crossers by classifying them into five categories: (1) asylum-seekers, (2) unauthorised entry, (3) unauthorised stay, (4) and (5) terrorists and criminals [31]. Overall, the amount of information, stories and data collected by national authorities at the French-Italian border in the last century could be seen as an antecedent of contemporary digitalised migration databases. Similarly, a genealogy of datafication at the border suggests that the difference between analogue and digital datafication is not a matter of a neat watershed [23]: in fact, migrants were fingerprinted in the last century as they are now, while what is noteworthy is the use of analogue and digital systems to perform the same task (extract biometric data). This pushes us to question analyses that speak of datafication of mobility at the border as a recent phenomenon. Indeed, we argue, it is key not to replicate a presentist view on datafication: as the historian Adam McKeown has pointed out, the emergence of border controls in the 19th-century went in parallel with, and has been consolidated through, the development of "photography, fingerprinting and anthropometric measurement" [68]. The datafication of different unauthorised migrants at the French-Italian frontier as much as at other borders, has been a crucial component for the formation and consolidation of the European border regime. The archival records of the French police we found did not classify unauthorised border crossers on the basis of race categories. And, yet, as we illustrate in this paper, racialisation through datafication takes place (also) by classifying individuals for determining their right to access the territory and stay — on the basis of their nationality, reasons for coming and job.

5 DISCUSSION

Identification and classification through datafication have historically been necessary to nation states, as John Torpey has retraced, to "develop the capacity to embrace their own citizens" and to expropriate from people "the legitimate means of movement" [88]. As scholars have shown, the census became the primary form of datafication to classify and identify populations over centuries [52, 61].

Identification and classification through datafication have consolidated as mechanisms for reproducing "structural violence" [47] and racialised hierarchies among citizens, as well as between citizens and foreigners [67]. Datafication has been historically enforced by nation states to identify undesirable and deceitful subjects and to make them reliable and knowledgeable [41]. This is nowadays strengthened through the use of large-scale systems implemented at the border to control migration and to block people on the move [10, 85].

As Chun argues "there are many similarities between 20th-century eugenics and 21st-century data analytics" [19, p. 66]. At the border, we have observed that datafication procedures have not substantially changed over time. For instance, as illustrated earlier in the paper, identification procedures used by the French police at the French-Italian border in the last century are partly similar to the ones adopted nowadays, if we consider the data collected from migrants. Moreover, fingerprints were taken at that time as they are now, with the difference that today border authorities might use digital systems. Mirroring Galton's archive at the colony brings us another example of quantitative continuity in the narrative. The alleged trustworthiness of fingerprints in contrast to the deceitful narrative of colonised subjects still underpins the ways in which authorities justify the use of biometrics at the border to control migration nowadays. The hierarchy of truthfulness imposes science, and more specifically biometrics, over the narrative of the colonised or the migrant [4].

Overall, despite datafication contributes to the racialisation of subjects, both states and private actors posit the neutrality and objectivity of datafying procedures. Yet, datafication of bodies has drastically evolved since Galton as a result of technological innovation in computing and, in particular, in biometrics. Since the development of architectural processing units, such as GPUs, that undoubtedly accelerated the computational power allowing the storage of millions of data traces and the design of algorithms to process this amount of data, states have been investing large amounts of public resources on these innovative technologies at the border for the sake of efficiency and security [89].

As Galloway illustrates "the digital organises technologies, bodies, and societies" [38, p. 232]. Devices that digitalise fingerprints, biometric databases that store more than 5.8 million fingerprints [32] and *efficient*[14] biometric systems that automatically match similar fingerprints through deep learning algorithms have been a paradigm shift since Galton, Herschel and Faulds. However, is it only a matter of quantitative shift or has the advent of big data generated also a qualitative shift in datafication? [6] Scholars have stressed that "the turn to data collection and algorithmic decision-making is therefore not simply a question of quantity — being able to process more information at a faster speed. It is also a qualitative shift that shapes reality and political subjects" [70, p. 8]. A focus on identification practices at the border foregrounds the apparently marginal qualitative shifts that digital datafication has triggered.

First, we contend, the quantitative shift we discussed above has led to some qualitative shifts, which concern the goals and the effects of classification: datafication and algorithmic-driven systems have facilitated the production of profiles of risk, based on the collection of huge amounts of data, abstracting from individuation. In other words, datafication has fostered abstraction processes to create categories and profiles of risk from the combination of individual-based features that stem from multiple data extraction procedures [1, 3]. For instance, datafication and biometric systems are nowadays key components in knowledge production about migration [76]; and the knowledge that they contributes to generate is mainly in terms of categories of risk.

A case in point is represented by the European Travel Information and Authorisation System (ETIAS) which is planned to be implemented in 2024, officially "created to identify security, irregular migration or high epidemic risks posed by visa-exempt visitors travelling to the Schengen State".[15] This system is set to work on the basis of different risk indicators that will be used to screen travellers and decide whether they could be granted the ETIAS authorisation. The risk indicators will include "age range, sex, nationality, country and city of residence or birth, level of education achievement, as well as current employer or occupation". These risk indicators will enact modes of proxy discrimination, that is by using categories that indirectly could be the markers of race (such as nationality), gender (sexual orientation), ethnicity (such as city of residence or nationality) and class (such as work/occupation).[16] Hence, algorithmic datafication enforces new modes of discrimination based on factors that work as indirect indicators and markers, and that are the result of recombinant identities. Second, algorithmic datafication is eminently future-oriented: this is particularly glaring in the field of migration governance, where migrants' mobility is datafied at the border not only for identification purposes but also for generating digital traces of the person and for keeping track of it in the future. By drawing attention to continuities and shifts between analogue and digital datafication algorithmic discrimination and racial variables appear as part of a broader spectrum of racialising classifications and exclusionary mechanisms enforced through datafication. Relatedly, a genealogical approach showcases the pitfalls of analyses which are exclusively centred on algorithmic (un)fairness, given that datafication as such is used — and has been used in history — by states for enacting and justifying hierarchies of political and racial subordination.

6 CONCLUSION

Classification criteria "become more visible, they break down or become objects of contention" [13, p. 2]. Along these lines, this paper has drawn attention to the racialisation mechanisms that datafication has historically generated. Indeed, datafication processes did not start with the rise of digitalisation, artificial intelligence or big data although it is not a matter of a full continuity between past and present. On the contrary, a genealogical approach has highlighted

---

[14]Biometric efficiency is arguable from a critical perspective. The implementation and functioning of large-scale biometric systems in real life involve a large mobilisation of resources that brings into question the effective efficiency of the whole procedure. For instance, the EU has postponed several times the implementation of a new biometric system at the border due to technical and logistic reasons. See: https://www.biometricupdate.com/202211/eu-border-biometric-checks-in-uk-likely-to-be-delayed (Last accessed April 12, 2023).

[15]See: https://home-affairs.ec.europa.eu/policies/schengen-borders-and-visa/smart-borders/european-travel-information-authorisation-system_en (Last accessed February 02, 2023).

[16]See: https://etias.com/articles/what-will-be-the-etias-screening-rules (Last accessed February 02, 2023).

partial continuities and shifts between analogue and digital datafi-cation and how this latter contributes to make up racialised subjects and profiling. As our archival research has shown, colonial subjects and migrants have been historically racialised by datafying their names, date of births, gender, nationalities and even fingerprints or face profiles. While the British empire extracted fingerprints to avoid personification of colonised subjects in the Riaj, the French police datafied unauthorised border crossers. Since the exclusion-ary processes it brings into being are not narrowed to algorithmic unfairness, racial categories or automated discrimination, it is key to expose the multiple racialising mechanisms that datafication enforces in the name of science, objectivity, and value-neutral fea-tures. Racialisation through datafication is historically enshrined in the very functioning of anthropometry [30].

Alongside unpacking how racialisation is enacted through datafi-cation, our genealogical approach invites the FAccT community to move beyond a techno-focused critique of algorithmic (un)fairness and to situate the analysis on bias and discrimination within a broader account of the exclusionary laws and policies that datafica-tion is used for by states. Indeed, a genealogy in the colony and at the border foregrounds that racialising mechanisms, not narrowed to race-based features, are enshrined in classification; this implies that there cannot be a non-discriminatory datafication about human mobility given that datafication is used for multiplying historical hierarchies of access to movement and rights. Thus, this paper does not offer policy recommendations, nor does it endorse a problem-solving approach. On the contrary, it suggests that instead of fo-cusing (exclusively) on criticising algorithmic bias to discuss how injustices are produced through socio-technical systems, critical scholars should reverse the standpoint by scrutinising how specific datafication mechanisms are put in place to enforce exclusionary legal-political measures and that are rooted in history. In other words, in the place of asking "how to design fair algorithms at the border?" or "how to unbias biometrics?" [90], we should rather ask "which racialising legal-political measures does datafication strengthen?".

ACKNOWLEDGMENTS

FUNDING DISCLOSURE

REFERENCES

[1] Louise Amoore. 2020. *Cloud ethics: Algorithms and the attributes of ourselves and others*. Duke University Press, Durham, USA.

[2] Arjun Appadurai. 2003. Archive and aspiration. *Information is alive* (2003), 14–25.

[3] Claudia Aradau and Tobias Blanke. 2022. *Algorithmic reason: The new government of self and other (p. 288)*. Oxford University Press, Oxford, UK.

[4] Claudia Aradau and Sarah Perret. 2022. The politics of (non-)knowledge at Europe's borders: Errors, fakes, and subjectivity. *Review of International Studies* 48, 3 (2022), 405–424. https://doi.org/10.1017/S0260210522000080

[5] Francoise De Barros. 2005. Des "Français musulmans d'Algérie" aux "immigrés". *Actes de la recherche en sciences sociales* 4 (2005), 26–53.

[6] David Beer. 2016. How should we do the history of Big Data? *Big Data & Society* 3, 1 (2016), 2053951716646135. https://doi.org/10.1177/2053951716646135

[7] Rocco Bellanova and Georgios Glouftsios. 2022. Controlling the Schengen Infor-mation System (SIS II): The infrastructural politics of fragility and maintenance. *Geopolitics* 27, 1 (2022), 160–184.

[8] Ruha Benjamin. 2019. *Race after Technology: Abolitionist Tools for the New Jim Code*. Polity, Oxford, UK.

[9] Sebastian Benthall and Bruce D. Haynes. 2019. Racial Categories in Machine Learning. In *Proceedings of the Conference on Fairness, Accountability, and Trans-parency* (Atlanta, GA, USA) *(FAT* '19)*. Association for Computing Machinery, 289–298. https://doi.org/10.1145/3287560.3287575

[10] Didier Bigo. 2020. The socio-genesis of a guild of 'digital technologies' justifying transnational interoperable databases in the name of security and border pur-poses: a reframing of the field of security professionals? *International journal of migration and border studies* 6, 1-2 (2020), 74–92.

[11] Emmanuel Blanchard. 2018. *Histoire de l'immigration algérienne en France*. la Découverte, Paris, France.

[12] Charles Booth. 1903. *Life and Labour of the People in London*. Macmillan and Co, Limited, London, UK.

[13] Geoffrey C Bowker and Susan Leigh Star. 2000. *Sorting things out: Classification and its consequences*. MIT press, Cambridge, USA.

[14] Matthias Braun and Patrik Hummel. 2022. Data justice and data solidarity. *Patterns* 3, 3 (2022), 100427.

[15] Denis Broeders and Huub Dijstelbloem. 2015. The datafication of mobility and migration management: The mediating state and its consequences. In *Digitizing identities*. Routledge, London, UK, 242–260.

[16] Simone Browne. 2015. *Dark matters: On the surveillance of blackness*. Duke University Press, Durham, USA. https://doi.org/10.2307/j.ctv11cw89p

[17] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, Vol. 81. Proceedings of Machine Learning Research, 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

[18] Dipesh Chakrabarty. 1983. *Conditions for knowledge of working-class conditions: employers, government and the jute workers of Calcutta, 1890-1940*. Oxford Uni-versity Press, Oxford, UK. 179–232 pages. Selected subaltern studies.

[19] Wendy Hui Kyong Chun. 2021. *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. MIT Press, Massachusetts, USA.

[20] Simon A. Cole. 2005. Imprint of the Raj: How Fingerprinting was Born in Colonial India. *Technology and Culture* 46, 1 (2005), 252–253.

[21] Patricia Hill Collins. 2019. *Intersectionality as critical social theory*. Duke Univer-sity Press, Durham, USA.

[22] Heaven Crawley and Dimitris Skleparis. 2018. Refugees, migrants, neither, both: Categorical fetishism and the politics of bounding in Europe's 'migration crisis'. *Journal of Ethnic and Migration Studies* 44, 1 (2018), 48–64.

[23] Xavier Crettiez and Pierre Piazza. 2006. *Du papier à la biométrie: identifier les individus*. Presses de Sciences Po, Paris, France. https://doi.org/10.3917/scpo.crett.2006.01

[24] Arnold I. Davidson. 2001. *The emergence of sexuality: Historical epistemology and the formation of concepts*. Harvard University Press, Cambridge, USA.

[25] Angela Y. Davis. 2011. *Abolition democracy: Beyond empire, prisons, and torture*. Seven Stories Press, New York, USA.

[26] Abigail Nieves Delgado. 2022. Race and statistics in facial recognition: Pro-ducing types, physical attributes, and genealogies. *Social Studies of Science* 03063127221127666 (2022). https://doi.org/10.1177/03063127221127666

[27] Emily Denton, Alex Hanna, Razvan Amironesei, Andrew Smart, and Hilary Nicole. 2020. Bringing the People Back In: Contesting Benchmark Machine Learning Datasets. In Proceedings ICML Workshop on Participatory Approaches to Machine Learning.

[28] Emily Denton, Alex Hanna, Razvan Amironesei, Andrew Smart, and Hilary Nicole. 2021. On the genealogy of machine learning datasets: A critical history of ImageNet. *Big Data & Society* 8, 2 (2021), 20539517211035955. https://doi.org/10.1177/20539517211035955

[29] Ariana Dongus. 2019. Galton's Utopia. *Spheres: Journal for Digital Cultures* 5 (2019), 1–16.

[30] Kevin Donnelly. 2020. We Have Always Been Biased. *Public* 30, 60 (2020), 20–33.

[31] eu LISA. 2021. *Eurodac - 2020 statistics*. Technical Report. https://www.eulisa.europa.eu/Publications/Reports/Eurodac

[32] eu LISA. 2022. *Eurodac Annual Report 2021 Factsheet*. Technical Report. https://www.eulisa.europa.eu/Publications/Reports/Eurodac

[33] Virginia Eubanks. 2018. *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press, New York, USA.

[34] Henry Faulds. 1880. On the skin-furrows of the hand. *Nature* 22, 574 (1880), 605–605.

[35] Michel Foucault. 1978. Nietzsche, genealogy, history. (1978), 139–164.

[36] Michel Foucault. 2008. *he History of Sexuality. Volume 1 the Will to Knowledge*. Penguin, Camberwell, UK.

[37] Anja Franck and Darshan Vigneswaran. 2021. Hacking migration control: Repurposing and reprogramming deportability. *Security Dialogue* (2021), 0967010621996938. https://doi.org/10.1177/0967010621996938

[38] Alexander R Galloway. 2022. Golden Age of Analog. *Critical Inquiry* 48, 2 (2022), 211–232.

[39] Francis Galton. 1891. III. Method of indexing finger-marks. *Proceedings of the Royal Society of London* 49, 296-301 (1891), 540–548.

[40] Francis Galton. 1892. *Fingerprints*. Macmillan and Co. Limited, London, UK.

[41] Francis Galton. 1900. Identification offices in India and Egypt. *The Nineteenth century and after: a monthly review* 48, 281 (1900), 118–126.

[42] Francis Galton. 1901. The possible improvement of the human breed under the existing conditions of law and sentiment. *Nature* 64 (1901), 659–665.

[43] Francis Galton. 1904. Eugenics: Its definition, scope, and aims. *Amer. J. Sociology* 10, 1 (1904), 1–25.

[44] Francis Galton. 1908. *Memories of my life*. Methuen, London, UK.

[45] Francis Galton. 1910. Numeralised profiles for classification and recognition. *Nature* 83, 2109 (1910), 127–130.

[46] Francis Galton. 1971. *Francis Galton's Art of travel (1872): A reprint of The art of travel; or, Shifts and contrivances available in wild countries*. Stackpole Books, Harrisburg, USA.

[47] Johan Galtung. 1969. Violence, peace, and peace research. *Journal of peace research* 6, 3 (1969), 167–191.

[48] Nicholas W. Gillham. 2001. *A life of Sir Francis Galton: From African exploration to the birth of eugenics*. Oxford University Press, Oxford, UK.

[49] Lawrence Goldman. 2022. *Victorians and Numbers: Statistics and Society in Nineteenth Century Britain*. Oxford University Press, Oxford, UK.

[50] Stephen Jay Gould. 1996. *Mismeasure of man*. W. W. Norton & Company, New York, USA.

[51] Catriona Gray. 2023. More than Extraction: Rethinking Data's Colonial Political Economy. *International Political Sociology* 17, 2 (2023), 1–20.

[52] Ian Hacking. 1990. *The Taming of Chance (Ideas in Context)*. Cambridge University Press, Cambridge, UK. https://doi.org/10.1017/CBO9780511819766

[53] Ian Hacking. 1999. *The social construction of what?* Harvard University Press, Cambridge, USA.

[54] Ian Hacking. 2004. *Historical Ontology*. Harvard University Press, Cambridge, USA.

[55] Philippe Hanus. 2020. 'Along Footpaths over Snow-Covered Mountains…' Historical Perspectives on Migration Journeys across the France–Italy Border (1945-1960). *Journal of Alpine Research Revue de géographie alpine* 108-2 (2020). https://doi.org/10.4000/rga.7121

[56] Willian Herschel. 1880. Skin Furrows of the Hand. *Nature* 23 (1880), 76. https://doi.org/10.1038/023076b0

[57] Sheila Jasanoff. 2017. Virtual, visible, and actionable: Data assemblages and the sightlines of justice. *Big Data & Society* 4 (2017), 2. https://doi.org/10.1177/2053951717724477

[58] Atoosa Kasirzadeh and Andrew Smart. 2021. The Use and Misuse of Counterfactuals in Ethical Machine Learning. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Virtual Event, Canada) *(FAccT'21)*. Association for Computing Machinery, New York, NY, USA, 228–236. https://doi.org/10.1145/3442188.3445886

[59] Yarden Katz. 2020. *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*. Columbia University Press, New York, USA.

[60] Rob Kitchin. 2014. Big Data, new epistemologies and paradigm shifts. *Big Data & Society* 1, 1 (2014), 2053951714528481. https://doi.org/10.1177/2053951714528481

[61] Colin Koopman. 2019. *How We Became Our Data: A Genealogy of the Informational Person*. The University of Chicago Press, Chicago, USA.

[62] John Law. 2009. Seeing like a survey. *Cultural sociology* 3, 2 (2009), 239–256.

[63] Matthias Leese, Simon Noori, and Stephan Scheel. 2022. Data Matters: The Politics and Practices of Digital Border and Migration Management. *Geopolitics* 27, 1 (2022), 5–25. https://doi.org/10.1080/14650045.2021.1940538

[64] Koen Leurs and Tamara Shepherd. 2017. Datafication & Discrimination. *The datafied society* (2017), 211–234.

[65] Mark Maguire. 2009. The birth of biometric security. *Anthropology today* 25, 2 (2009), 9–14. https://doi.org/10.1111/j.1467-8322.2009.00654.x

[66] Mark Maguire. 2012. Biopower, racialization and new security technology. *Social Identities* 18, 5 (2012), 593–607.

[67] Amade M'charek. 2008. Silent witness, articulate collective: DNA evidence and the inference of visible traits. *Bioethics* 22, 9 (2008), 519–528.

[68] Adam McKeown. 2008. *Melancholy order: Asian migration and the globalization of borders*. Columbia University Press, New York, USA.

[69] Dan McQuillan. 2022. *Resisting AI: an anti-fascist approach to artificial intelligence*. Bristol University Press, Bristol, UK.

[70] Philippa Metcalfe and Lina Dencik. 2019. *The politics of big borders: Data (in)justice and the governance of refugees*. First Monday. https://arianadongus.com/projects/galtons-utopia-data-accumulation-in-biometric-capitalism/#

[71] Sandro Mezzadra and Brett Neilson. 2013. *Border as Method, or, the Multiplication of Labor*. Duke University Press, Durham, USA.

[72] Sandro Rinauro. 2009. *Il cammino della speranza: l'emigrazione clandestina degli italiani nel secondo dopoguerra*. Einaudi.

[73] Clifford Rosenberg. 2006. *Policing Paris: The origins of modern immigration control between the wars*. Cornell University Press, Ithaca, USA.

[74] Katerina Rozakou. 2017. Nonrecording the "European refugee crisis" in Greece: Navigating through irregular bureaucracy. *Focaal* 77 (2017), 36–49.

[75] Evelyn Ruppert, Engin Isin, and Didier Bigo. 2017. Data politics. *Big Data & Society* 4 (2017), 2. https://doi.org/10.1177/2053951717717749

[76] Stephan Scheel, Evelyn Ruppert, and Funda Ustek-Spilda. 2019. Enacting migration through data practices. *Environment and Planning D: Society and Space* 37, 4 (2019), 579–588.

[77] Morgan Klaus Scheuerman, Alex Hanna, and Emily Denton. 2021. Do datasets have politics? Disciplinary values in computer vision dataset development. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–37.

[78] Morgan Klaus Scheuerman, Alex Hanna, and Emily Denton. 2021. Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 5. Association for Computing Machinery, 1–37. https://doi.org/10.1145/3476058

[79] Jens Schröter. 2020. Analogue/Digital-Opposition or Continuum? *MediArXiv* (2020). https://doi.org/10.33767/osf.io/x7eq3

[80] Cornelia Schweppe and Nandita Sharma. 2015. Borders–transborders–no borders: problematizing the "figure of the migrant". *Transnational Social Review* 5, 1 (2015), 2–6.

[81] James C. Scott. 1998. Seeing like a state. In *How certain schemes to improve the human condition have failed*, Yale University Press and U. K. London (Eds.).

[82] Chandak Sengoopta. 2003. *Imprint of the Raj: how fingerprinting was born in colonial India*. Macmillan, London, UK.

[83] Ann Laura Stoler. 2010. *Carnal knowledge and imperial power: Race and the intimate in colonial rule*. University of California Press, Berkeley, USA.

[84] Ann Laura Stoler. 2010. *In Along the Archival Grain*. Princeton University Press, Princeton, USA.

[85] Javier Sánchez-Monedero. 2018. *The datafication of borders and management of refugees in the context of Europe*. Technical Report. https://orca.cardiff.ac.uk/id/eprint/128361/

[86] Martina Tazzioli. 2019. *The making of migration: The biopolitics of mobility at Europe's borders*. Sage, London, UK.

[87] Martina Tazzioli. 2020. Disjointed knowledges, obfuscated visibility. Border controls at the French-Italian Alpine border. *Political Geography* 79 (2020), 102155.

[88] John C. Torpey. 2018. *The invention of the passport: Surveillance, citizenship and the state*. Cambridge University Press, Cambridge, UK.

[89] Ana Valdivia, Claudia Aradau, Tobias Blanke, and Sarah Perret. 2022. Neither opaque nor transparent: A transdisciplinary methodology to investigate datafication at the EU borders. *Big Data & Society* 9 (2022), 2. https://doi.org/10.1177/20539517221124586

[90] Ana Valdivia, Júlia Corbera Serrajòrdia, and Aneta Swianiewicz. 2022. There is an elephant in the room: Towards a critique on the use of fairness in biometrics. *AI and Ethics* (2022), 1–16.

[91] Niovi Vavoula. 2020. Interoperability of EU information systems: The deathblow to the rights to privacy and personal data protection of third-country nationals? *European public law* 26, 1 (2020), 131–165.

[92] Alexander G. Weheliye. 2014. *Habeas viscus: Racializing assemblages, biopolitics, and black feminist theories of the human*. Duke University Press, Durham, USA.

[93] Saskia Witteborn. 2022. Digitalization, Digitization and Datafication: The "Three D" Transformation of Forced Migration Management. *Communication, Culture and Critique* 15, 2 (2022), 157–175. https://doi.org/10.1093/ccc/tcac007