Check for updates

Editorial: Migration studies and the digital: datafication, implications and methodological approaches

Lorella Viola[1]*, Machteld Venken[1] and Jaap Verheul[2,3]

[1]Centre for Contemporary and Digital History (C[2]DH), University of Luxembourg, Esch-sur-Alzette, Luxembourg, [2]Department of History and Art History, Utrecht University, Utrecht, Netherlands, [3]Department of Modern Languages and Cultures, Radboud University, Nijmegen, Netherlands

Editorial on the Research Topic
Migration studies and the digital: datafication, implications and methodological approaches

Migration is a complex and multifaceted phenomenon that shapes societies and individuals around the world. As the digital era unfolds, the field of Migration Studies has witnessed a significant shift in focus, with more scholars increasingly mobilizing digital archival sources and employing Digital Humanities methodologies for their analysis (Pötzschke and Rinken, 2022). This Research Topic brings together a collection of contributions that explore new directions within the evolving field of Digital Migration Studies, offering an overview of innovative approaches and results that highlight the richness and potential of the intersection between Migration Studies and the digital realm.

Advancements in the field of digital humanities have paved the way for the development of analytical tools that enable researchers to trace migration patterns with unprecedented depth and accuracy. These tools facilitate a wide range of data-driven techniques, including word vector models, discourse-driven topic modeling, place name extraction, visualization techniques (e.g., "Flow toward Europe"), automated text recognition of handwritten sources, network analysis, and classifiers for genre detection (Clavadetscher et al., 2022). These methodological advancements empower scholars to delve into migration studies with a fresher perspective, uncovering intricate patterns, and shedding light on the forces and dynamics that shape human movement.

The use of digital datasets and technologies for example brings forth new dimensions to the field of Migration Studies, unveiling a multitude of research directions. One of the key advantages offered by digital collections is the ability to seamlessly integrate macro and micro perspectives. Distant reading techniques, for instance, enable researchers to map and visualize larger migration patterns, facilitating the identification of continuities and changes over extended periods. By merging the macro and micro lenses, comprehensive overviews of demographic migration patterns can be constructed, while simultaneously revealing unexpected connections between departure and arrival locations. Digitized data allows for the assessment of both grand historical forces and the intricate pathways of individual lives, painting a holistic picture of migration experiences (see Venken and Sauer in this collection and their discussion of complex migration trajectories).

Furthermore, the utilization of digital and digitized collections enables scholars to uncover the lived experiences of immigrants (Al Kalak, 2022; see Viola in this collection and her analysis of return migration as a cognitive experience). Through the mapping and recording of narratives, stories, and voices of underrepresented individuals, digital methodologies provide a platform to give life to previously marginalized perspectives. Additionally, digital technologies play a crucial role in the study of public media as carriers of meaning, particularly in relation to migrants. Mass migration and mass media have long been intertwined, shaping the expectations of migrants and constructing their roles within host societies. Ethnic media, including newspapers and social media networks, offer unique insights into the lived experiences of migration, reflecting the realities and challenges faced by immigrant communities (see Winarnita, 2019; Dedecek Gertz, Viola). Just like immigrant newspapers at the turn of the twentieth century, digital-born social media networks and internet connectivity play a significant role in maintaining connections among diaspora groups across borders, facilitating communication networks and aiding migrants in their journeys and settlement processes (see Dedecek Gertz in this collection and her analysis of migrants' digital traces and Viola in this collection on how diasporic newspapers can unlock new perspectives on migration studies). These networks become particularly crucial during moments of transition and crisis, such as the death of a family member (see Sánchez-Querubín in this collection and her conceptual analysis of "technological haunting" as a concept in migration and transnational death studies).

While the integration of digital heritage collections and analytical methodologies presents exciting possibilities, this Research Topic also addresses the challenges inherent in their usage within the study of migration experiences and patterns. One such challenge lies in the diversity of data, as researchers must navigate the connections between databases, media sources, stories, and multimedia of different origins and metadata structures (see Arthur and Smith in this collection and their discussion of the tensions between technological advances and methodological shifts in migration research). Meaningfully tracing the life trajectories of migrants necessitates bridging archival collections that often reflect the needs and power structures of nation-states, treating migrants as exceptions rather than central figures. Governments, both in countries of origin and destination, often create and utilize data structures to monitor and govern migration, making it a digital challenge to connect administrative waypoints in migrant journeys (see Faassen and Hoekstra in this collection and their analysis of the dispersed and fragmented nature of migration heritage involving at least two countries).

In addition, the categorization of migrants introduces complexities and challenges in data collection and analysis. The layers of identity and the positive or negative connotations associated with terms such as migrants, fugitives, asylum seekers, aliens, foreigners, radicals, or terrorists can influence the creation of digital collections and the institutional attempts to create rubrics. These categories shape our understanding of migration experiences and impact the representation and interpretation of data (see Viola in this collection and her analysis of how returned migrants were linguistically framed in the press).

Furthermore, the silences within the cultural archive pose significant challenges. Inequality of archival representation and knowledge production is a persistent issue, as certain migrant groups, such as female immigrants or intersecting marginalized identities, remain underrepresented. Digital methodologies hold the potential to address these gaps, but doing so requires a constant application of source criticism and methodological scrutiny to ensure accurate and inclusive representation (see Arthur and Smith; Faassen and Hoekstra). However, it is essential to navigate the ethical questions surrounding information gathering, particularly when vulnerable groups are involved, to uphold the principles of informed consent, privacy, and the responsible use of data (see Dedecek Gertz in this collection and her discussion of the ethical issues surrounding digital migration research).

But digital scholarship also opens opportunities to revisit old Research Topics in historical migration studies -such as return migration -from new angles and perspectives. Beyond socio-economic indicators, migration statistics, or post factum interviews, newly rediscovered digitized sources, for example immigrant newspapers, allow scholars to access the inner voice of migrants and explore unanswered questions about the mass migratory movements of the past century. Digital approaches can indeed open access to a more authentic voice of the migratory experience, enabling researchers to understand it as a profoundly changing psychological experience and a crisis of identity (see Viola in this collection and her analysis of historical return migration as a complex, discursive process).

The contributions to this Research Topic illuminate the dynamic landscape of Migration Studies in the digital age, offering valuable insights into new research frontiers, analytical methodologies, and the challenges faced by scholars in this emerging field. By exploring digital archival sources and employing Digital Humanities approaches, the authors of this collection contribute to a deeper understanding of migration, fostering interdisciplinary dialogue, and encouraging critical reflections on the implications and applications of the digital in the study of human mobility.

Ultimately, the Research Topic "*Migration studies and the digital*" propels the field of Migration Studies forward.

Author contributions

LV: Conceptualization, Writing – review & editing. MV: Writing – review & editing. JV: Writing – original draft.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Al Kalak, M. (2022). *Ithaca. Interconnecting Histories and Archives for Migrant Agency*. (2022) Available online at: https://ithacahorizon.eu/ (accessed November 30, 2023).

Clavadetscher, S., Mahrer, S., and Salvisberg, S. (2022). Mapping forced academic migration. *Stu. Digital History Hermeneut.* 107, 105–124. doi: 10.1515/9783110744828-006

Pötzschke, S., and Rinken, S. (2022). *Migration Research in a Digitized World: Using Innovative Technology to Tackle Methodological Challenges*. Cham: Springer.

Winarnita, M. (2019). Introduction to this special issue on multimedia, mobility and the digital southeast asian family's polymedia experiences. *Migr. Mob. Displ.* 4, 1–5. doi: 10.18357/mmd41201 918967

Migrant visibility: Digitization and heritage policies

Marijke van Faassen*† and Rik Hoekstra†

Huygens Institute, KNAW, Amsterdam, The Hague, Netherlands

Digitization and digital methods have had a big impact on migration history and history in general. The dispersed and fragmented nature of migration heritage that involves at least two countries and many cultural heritage institutions make it clear that migration history can be much improved by using digital means to connect collections. This makes it possible to overcome the biases that policy have introduced in private and public collections alike by selection and perspective. Digital methods are not immune to these biases and may even introduce new distortions because they often change heritage contextualizations. In this article, Van Faassen and Hoekstra argue that therefore they should be embedded in source criticism methodology. They use the example of post-world War II Dutch-Australian emigration to show how a migrant registration system can be used as a structural device to connect migrant heritage. They use methods from computer vision to assess the information distribution of the registration system. Together, connecting collections and information assessments give an encompassing view of the migrant visibility and invisibility in the heritage collections and perspectives for scholars to become aware of heritage biases.

KEYWORDS

migrant history, digital methods, source criticism, computer vision, connecting heritage

Introduction

In the last two decades migration studies showed a more pronounced interest in what they call the "invisibility" of migrants. The topic was put on the agenda by developments within the multidisciplinary field of migration studies itself, but also by debates within the archive and heritage sciences, that were exploring the new responsibilities and possibilities coming with the multimedia and digital era. The volume *Europe's invisible migrants* (2003) for instance, that protaganized migrants of decolonization who were till then euphemistically categorized as repatriates by their governments and understudied by academia, was inspired by the new approach suggested by Lucassen and Lucassen in their landmark work of 1999 (Lucassen and Lucassen, 1999). They argued that migration history had developed too many typologies, like forced and voluntary migration, that had evolved into fixed dichotomies and divided migratory experience and scholarship alike and that these conceptual walls had to be broken down to open up the field again (Smith, 2003, p. 18). From 1993 academia picked up that certain migrant communities experienced themselves as being invisible because as a result of strict assimilation policies little tangible heritage of them had been retained with which they could

distinguish themselves (Walcker-Birckhead, 1988; Willems, 2001, 2003; Coté, 2010, p. 122; Peters, 2010; Horne, 2011).

The cultural heritage field was criticized in those years that it supported and contributed to this invisibility as it only followed the official archives bringing just official stories and histories that affirmed the importance of hegemonistic groups in society, usually the white majority. To remediate these false and lopsided views, critical scholars argued that archives need to be decolonized (Stoler, 2002; Jeurgens and Karabinos, 2020). Other researchers and cultural heritage professionals have made ongoing efforts to find and tell alternative minority stories. They also strove to identify cultural heritage in the archive by creating forms of alternative access. Özden Yalim's "Passing on 'invisible' histories" (Yalim, 2008), for instance, put her finger on the blind spots in the Dutch women's movement and its archive Atria for the role of immigrant women in the labor force. In the 1998 preparations of the commemoration of the 100-year *National Exhibition on Women's Labour (1898)*, immigrant women were first completely left out. She describes the methods that were used to remediate this: collecting, preserving and reconstructing the cultural heritage of several women immigrant groups by using oral history and storytelling techniques and new media, so that the immigrant women were "given a voice," which put them at the center of collaborating with experts and academics. In the United States and Oceania there have been similar experiments of co-creation with first nation or indigenous minority groups with heritages with forms of expression that do not fit the customary archive format (McKemmish, 2017).

Scholars who engage in combining digital humanities and migration or heritage studies also point at a shortage of sources as a cause for a certain blindness to study certain migrant groups. In the newly launched *Journal of Digital History*, Oberbichler and Pfanzelter (2021) justly point out that return migration—defined as cross-border migration to the country of origin—is a too frequently neglected topic in migration studies and migration history, as return migration is always part of every migratory movement. In their opinion national or even regional approaches have proven to be fruitful for studying the historical developments of migration, despite its transnational nature, "where archival material is scarce and sources often are lacking." They experimented with corpus building from German language newspapers using text mining and machine learning techniques to address the lack of sources for studying return migration from the Americas to Europe between 1850 and 1950.

Although we agree that national and regional approaches can be fruitful in migration studies, we have a different opinion about their statement on the scarcity of sources and archival material. Precisely because of the transnational character of migration, migration history and migrant heritage is by its very nature *dispersed* over different collections in different countries, but not necessarily absent. As the examples

on invisibility studies above show, there are often issues of ethnicity at stake, paradoxically enough even for what most researchers see as dominant white migrant groups. Upon closer inspection the more important common denominator or cause of experienced invisibility seems to be the policy at the time: ways of categorization that concealed certain groups, not only in the public debates of the time but especially in the archival heritage. This problem was discussed in 2013 in Schrover and Moloney's study on *Gender, Migration and Categorization*. They observed that scholars tend to follow the categorizations that policy makers use, often as a result of the source material that is available and organized according to these categorizations (Schrover and Moloney, 2013, p. 9). Therefore, we argue, we should not only engage in preserving heritage by co-creating or corpus building to make migrants visible but also take one step back and ask ourselves how to get the best out of the existing archival collections. More specifically our research questions are who and what is actually visible in dispersed archival migrant collections? How can digital methods help in discerning what is actually visible and what is not and what does this mean with respect to source criticism?

In this article we use the *Migrant Mobilities and Connection* project in which we have been working on methodologies for structurally connecting the information for Dutch-Australian migrants from 1945 to 1992 as a use case. We work on an aggregative representation that includes the many different perspectives that together constitute the migrant experience. Processes of selection and dispersion of information and heritage accumulate to make collections silos of information when they are digitized and datafied. In our case migration registration systems can play this pivotal role in connecting different types of information, ranging from personal files and stories to policy files and propaganda and including archive files, private collections, books, photos, moving images and sound, as well as digital communication such as Facebook and Instagram groups (Arthur et al., 2018). We argue that for exploring dispersed collections and for analysis, a serial resource is vital to interconnect dispersed cultural heritage and provide them with a logical context and make it possible to move beyond the impressionism of isolated case studies and cherry-picking.

In the following sections, we first elaborate on the discourse on the decolonization of the archives in order to explore the impact of digitization and the use of digital methods and the way scholars have used them and what that means for what is visible and invisible for them. Then we explain the way we try to connect different collections regarding migrants and migrant policy in our project by using a data backbone. Finally, we elaborate on the digital methods we used to explore the impact of national and bilateral policies on the nature and content of the different collections and what this means for the visibility of migrants in heritage collections.

Historians and the use of digital collections

The "archival turn" of the recent years, brought a lot of attention to reconfiguring and decolonizing the archive by paying attention to the ideological biases that are part of the archive that explicitly and implicitly capture policies and their ideologies and the hegemonistic master narratives that they supported. It made the archive itself an object of study and is for a part, a reinterpretation based on a combination of distant and close reading, but it has also been proposed that digital methods of virtual reordering could reconfigure the archives in such a way that alternative narratives cold be supported (Gilliland et al., 2017; Ketelaar, 2017; Jeurgens and Karabinos, 2020; Hoekstra et al., 2021, p. 8). We noticed that generally the "archival turn" just paid attention to the single archive. For example, writing about the changed relation between archivists and historians, Blouin and Rosenberg (2011, p. 4) write: "*The most common understanding of an archive would describe it as a body of records generated by the activities of a specific individual or organization and commonly located (although not always) in a repository housing similar or related collections. The Boston City Archives holds records generated by the bureaucracy of the city of Boston. The Archives Nationales in Paris and the U.S. National Archives in Washington hold essential records of the modern French and American states.*"

The extent and richness of digitized collections and the challenges in exploring them with new methodologies reinforces the tendency to concentrate on a single collection. However, this contrasts with how historians usually work. Chassanoff (2013, p. 461) notices that "*Historians typically consult a large number of institutions during the archival research process. Archival institutions may include public or university libraries, academic special collections/repositories, state or local historical societies, museums, and state or government archives.*"

The reason that historians consult different archives is that they want to contextualize their findings. Archives were always created for a specific purpose, usually related to administrative procedures that involve different actors, even if they are from a single government (Upward et al., 2018). Because of this, different archives and even archive collections contain different information, say the policy files that set out the policy lines and the administrative files that recorded the implementation of that same policy. This is the case for archives (and other heritage collections) that come from for instance different departments of the same organization, but of course even more if they were created by different organizations for different purposes. The separation of collections leads to a fragmentation of the policy and executive files that were once part of a connected reality. To get a fuller grasp of past realities, it is vital to reconstruct politics and ideas or ideologies and recontextualise documents and collections by connecting them. This also makes it possible to come to a fuller awareness of the ideological biases and the necessary scope for decolonization and other assessments of past policies and ideologies.

Historians have always seen the need for combining heritage collections but in the case of migration history it is obvious that heritage is distributed over many different collections in different countries. In practice it was always impossible to connect this dispersed heritage, because of the practical issue that they are physically distributed over different institutions with different policies and even over different countries, like in the case of the migrant heritage. Of course, historians have known this for a long time, but without connected collections it is really hard and very time consuming to get an idea about all collections relevant for a particular historical phenomenon. Interviewing historians about their use of digitized archives, Coburn (2021, p. 404) cites as one of the main perceived advantages "the improved availability of relevant materials, describing this as the ability to forestall travel to archive sites for research purposes." He continues, that aside from traveling it is next to impossible to connect a myriad of collections in another way than using digital methods. If chosen carefully, digital methods enable us to make links between collections in a way that makes it possible to explore a much wider variety of heritage materials in a structured way. Coburn points out that there is concern about the selectivity of digitized archives and the naiveté of (other) historians in dealing with them (Coburn, 2021). He argues that in his experience historians are actually well aware of the pitfalls both of digital selectivity and the loss of context that occurs when archival items are located by search engines. This is indeed an important issue that researchers have to take into account. He also cites historians who try to reconstruct this context. As we have argued above, a full context reconstruction should include at least an awareness of all heritage holdings involved.

Of course, there have been numerous efforts to connect previously separate historical sources into a connected resource. Indeed, the linked data movement strives to enable linking data and there are large meta collections like Europeana (cf. Hoekstra et al., 2021). However, as we have argued elsewhere collections should not be just linked but structurally connected into a new dataset instead, to prevent digital selectivity, processing and decontextualization from introducing other, more subtle biases. We have called such a connected set a datascope (Hoekstra and Koolen, 2019). It needs a structural device to serve as a backbone to connect the constituent datasets and collections. The purpose of our connecting efforts is to construct a datascope for Dutch-Australian emigration using the registration cards as a structural device.

Writing about biases of dominant culture in the archives and the ways digital methods can contribute to solving this, Blouin and Rosenberg (2011, p. 4, cf. Hoekstra et al., 2021) write about the awareness of "*the importance of 'authority'… in conveying a sense of the past as well as an understanding of its*

documentary residues. This led [them] to such issues as the role of identity and experience as 'authorities' in forming both historical understanding and the structures of archival collections; the activism of archivists themselves in these processes; and the forms and often contested natures of archival and historical sources."

In other words, the archives and other heritage collections reflect policies, usually official policies, both past and current. The influence of policy and policy informed decisions on the archives is manifold and multilevel. The first level is the formation and division of archives and archive collections themselves—which agents were involved in governance and therefore in the creation of an archive and what aspect and perspective do they represent? The second level is the composition of the different collections—archive collections are usually the residues of administrations. What was its purpose and what do they contain and how was it ordered? The third level is about the nature and content of specific administrative devices, such as registration systems—what does it register, what is left out?

These represent very different facets of a governance system and the way it kept and handed down its records. Digitization and datafication make it possible to explore and connect much larger portions of an archive, but first they also introduce new policy induced biases of selection and organization. Second, they tend to obscure traces of policies because they distort old context and introduce new ones, they tend to obscure the policies that determined the original context. On a final point, researchers have to keep in mind that digitization brings the possibility to connect archival and other heritage materials, but most of heritage is not digitized and probably never will be. The last European survey of 2017 estimated that 10 percent of European heritage collections had been digitized and 40 percent would never have to. The same report remarks that even if 80 percent of the cultural heritage institutions in 2017 had digital collections, less than half of them had a digitization strategy document whatsoever (Nauta et al., 2017, p. 28, 5, 15–16; Nyirubugara, 2012, p. 81–88). The parts that have been digitized, are only digitized partially. Often there is a selection of the "most interesting" or most used parts or just the inventory and some highlights are digitized. For example, the Dutch National Archives write that "the most frequently consulted archives of the past 20 years have been selected for digitisation" (Nationaal Archief, 2022) and the National Archives of Australia state they "[d]igitise the collection with particular emphasis on [...] high-use information" (National Archives of Australia, 2022, also cf. Thylstrup, 2019; Hauswedell et al., 2020).

This is even more the case for datafication, as datafying puts structural requirements on the contents of the documents. For instance, a form-based registration can be converted relatively easily, but it is really hard to fit in annotations that do not follow the structure of forms. Moreover, there are many documents in which information is not structured in such a way that they can be reduced to a data structure easily, for instance because they contain a policy argument or an analysis instead of data. In this way, much of this type of documents are not datafied and they are seldom included in data-based analysis. All this has even more consequences for dealing with (partially) digital research. Some choose to circumvent these confinements of digital research by concentrating on a single digital collection or digital corpus. While this is a legitimate object for research, migrant history teaches us that such an approach is too limited. Others have introduced alternative collections by using social media and oral history to better represent the unheard voices (for example Leurs, 2021). Social media only became available with the advance of the digital. However, it has often been pointed out that social media echo the views and preferences of the public debate and policy issues at stake (cf. Tufekci, 2017). We would argue that similar considerations are true for oral history. While these sources are valuable contributions, they should still be critically examined using the same criteria of source criticism.

This illustrates our point that it is important to not just decolonize the archive, but to structurally connect policy with what may appear cultural heritage. Researchers often know that collections are institution artifacts, but fail to take into account that they also change dynamically under the interaction with both society and policy changes (but see Upward et al., 2018). The characteristics and changes of archives influence who is visible in the archival collections, because like policy, archives do not treat everyone equally. Digitization often unconsciously enhances these processes of selection. Digital methods may be used to make visible what was previously invisible, but they also can result in affirming old policy preferences if archivists and researchers are not aware that they are hidden in the archive, and often old policy is lost in time.

Our project *Migrant Mobilities and Connection* was conceived as a combined digital-analog project in which we have strived to incorporate these different aspects from the beginning. Its point of departure was the awareness that migration history and migrant heritage is by its very nature dispersed over different collections in different countries. Therefore, a study that is based on just one collection, let alone one corpus, for us was never a satisfying option because it leaves out the information and perspectives that are represented in other collections. We strove to find ways to integrate digital and analog material and allow for many different facets of the history of migrants, instead of telling a single story (Faassen and Hoekstra, forthcoming). Migrant, Mobilities and Connection, therefore, has a 2-fold aim: first digitally connect the cultural heritage of Dutch-Australian migrants that is dispersed over many collections in many institutions in two countries. As the archival heritage of the emigration policy of the Netherlands is not digitized and still hidden behind very generic inventory terminology which is—because of the complexity of the Dutch emigration governance system—dispersed over more than eight governmental ministries and even more private organizations,
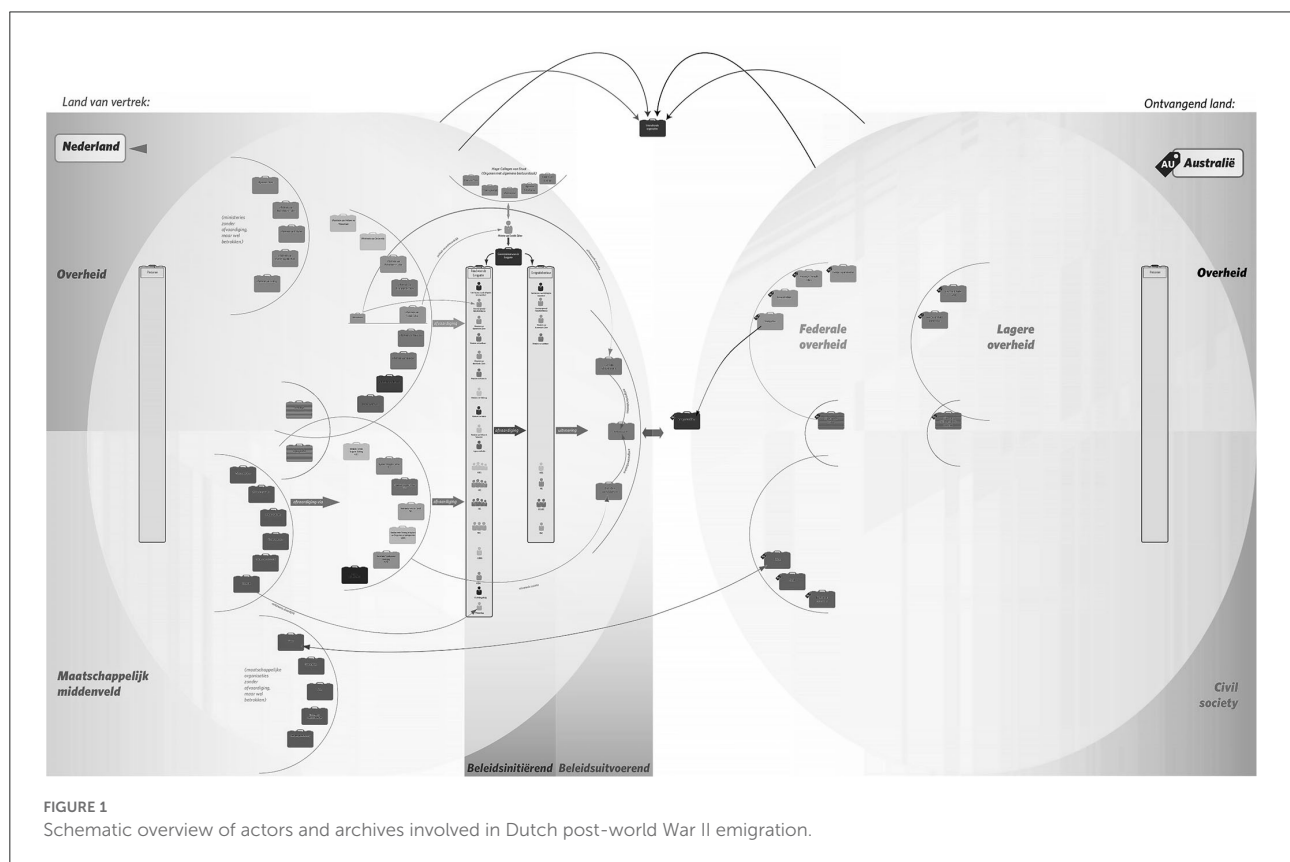
the first effort was to digitally identify the historical actors (and thus record creators) involved and to make a summary description of their 367 migration-related collections (see Figure 1, Faassen, 2014b).

Although the documents themselves are still not digitally available, at least their existence can be found online and their interdependence is made clear. The next step was to find a connecting device to link the institutional to the personal in both countries. As a core for the Migrant-project, we chose to digitize the migrant registration cards that were made by the Dutch emigration services in the Netherlands and traveled to Australia with the migrant application files, which served as input for the Australian immigration authorities. Subsequently these cards were repurposed by mainly the emigration attaches that were positioned at the Dutch consulates in Australia to support the migrants in their new home country (Faassen and Oprel, 2020). In doing so we hope to have established a new resource that is easily accessible for a larger public (the core business of Huygens Institute, see also Arthur et al., 2018) and that can be supplemented by other public or private collections, interviews etc. This resource can also facilitate our second aim: to start answering our main research question: how are policy and migrant agency related with respect to the whole migration experience?

Card index systems and methodology: Computer vision to remove blind spots

The card files of migrant cards are part of the collection of the Ministry of Foreign Affairs (NL_HaNA, 2.05.159). There are 51,525 cards that correspond to 100,000 images. They represent circa 180,000 migrants, or 80–90 percent of all migrants to Australia, as the cards contained data about migrant units, often but not always families, with personal data about the migrants. The cards contain a number of data about the migrants like personal data such as name, birthdate, occupation, place of origin and the dates and means of migration. These were compiled before migration. The cards contain additional unstructured data about interactions with the consulates. Because they comprise such a large share of the emigrants and a lot of their context, the cards enable us to connect all sorts of digital heritage (Faassen and Hoekstra, forthcoming). They also make it possible to make informed samples.

From a research point of view, it is necessary to assess the cards. We have images of the cards and a very summary index table with core data. In digitizing, the archive lost the connection between the index and the images. Also, the original

FIGURE 1
Schematic overview of actors and archives involved in Dutch post-world War II emigration.

order was disturbed in many places. We cannot read all cards because there are far too many and the cards contain a mix of typed and handwritten information that made experts in text recognition for the most part shy away. Developments in Handwritten Text Recognition (HTR) now make it possible to recognize manuscript and mixed manuscript and handwritten text. In fact, one of the authors is involved in other projects that employ HTR for a large scale archive resource and compare and develop recognition methologies (Koolen et al., 2020). However, the cards are mostly structured models for which recognized text is not sufficient and therefore that require much more additional work on structuring (cf. Tames, 2022).

We did a general assessment of the cards with a one percent sample, taking each 100th card (consisting of 2 images and possible follow-up cards as far as they were localizable), structuring and analyzing the information on the cards. The sample confirmed that there is a wealth of information about migrants on the cards about the Dutch-Australian migrant population. Even a correctly drawn sample results in a simplification, or a small world representation of a larger world (McElreath, 2020, p. 19–46). Historians often complement this simplification by taking cases from the collection and studying these in depth to get insight into the variation, an established method in history. For the emigration cards, however, this posed a few different problems, especially selecting the cases. It is very hard to select the largest cards as some historians suggested, as we have no physical or visual access to the cards. Moreover, there may be a reason that files get big that would lead to an unconscious selection bias, also known as cherry-picking, as it is unclear for what reason migrants would get more attention. Notwithstanding the samples, the cards effectively were closed for research and we had to find a solution to complement the sampling. This solution had to be based on digital methods to be able to deal with the size of the registration system. We devised a method with different steps (Hoekstra and Koolen, 2019; Hoekstra, 2021).

We first manually reconstructed the relation between images and the index table. Then, we devised a way to measure the information density on the cards, using a simple form of computer vision. The mixed script on the cards may be too difficult to transcribe using a combination of OCR and handwritten text recognition (HTR), but it is possible to measure the amount of writing on the cards, using the script edges that can
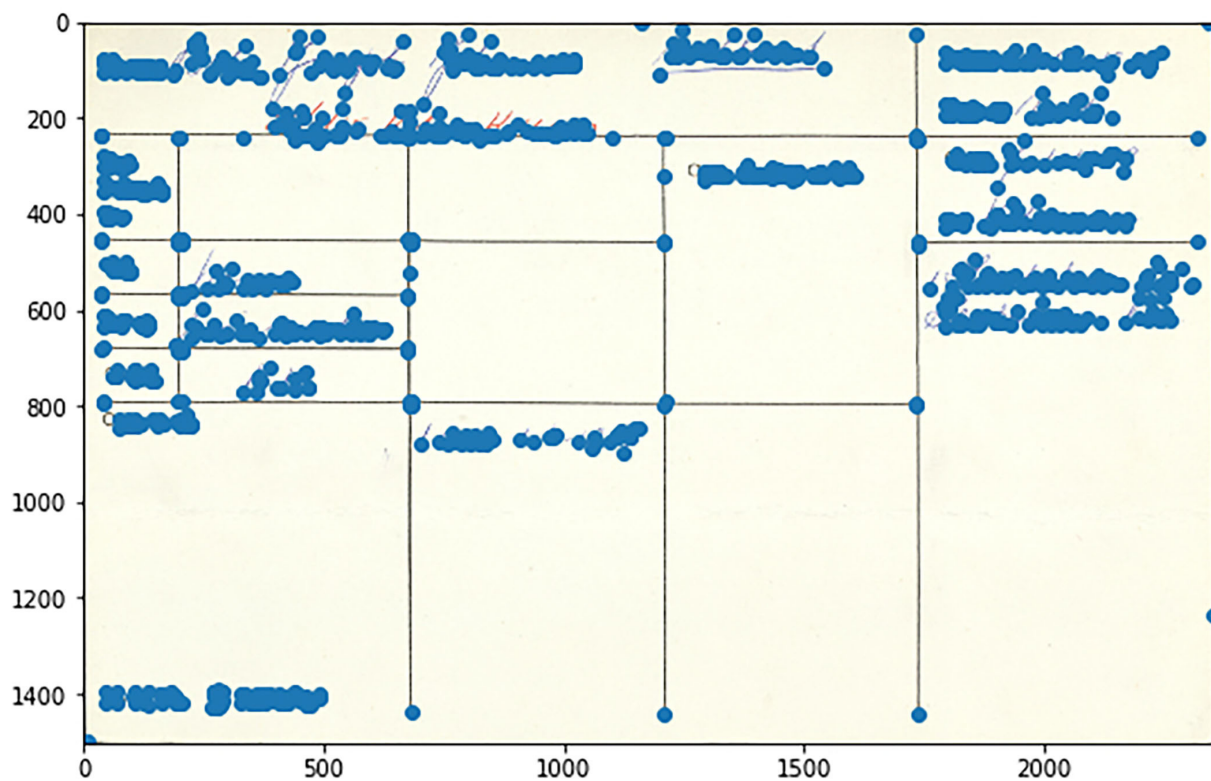

FIGURE 2
Script edges on a migrant registration card—(source—emigrant registration cards).
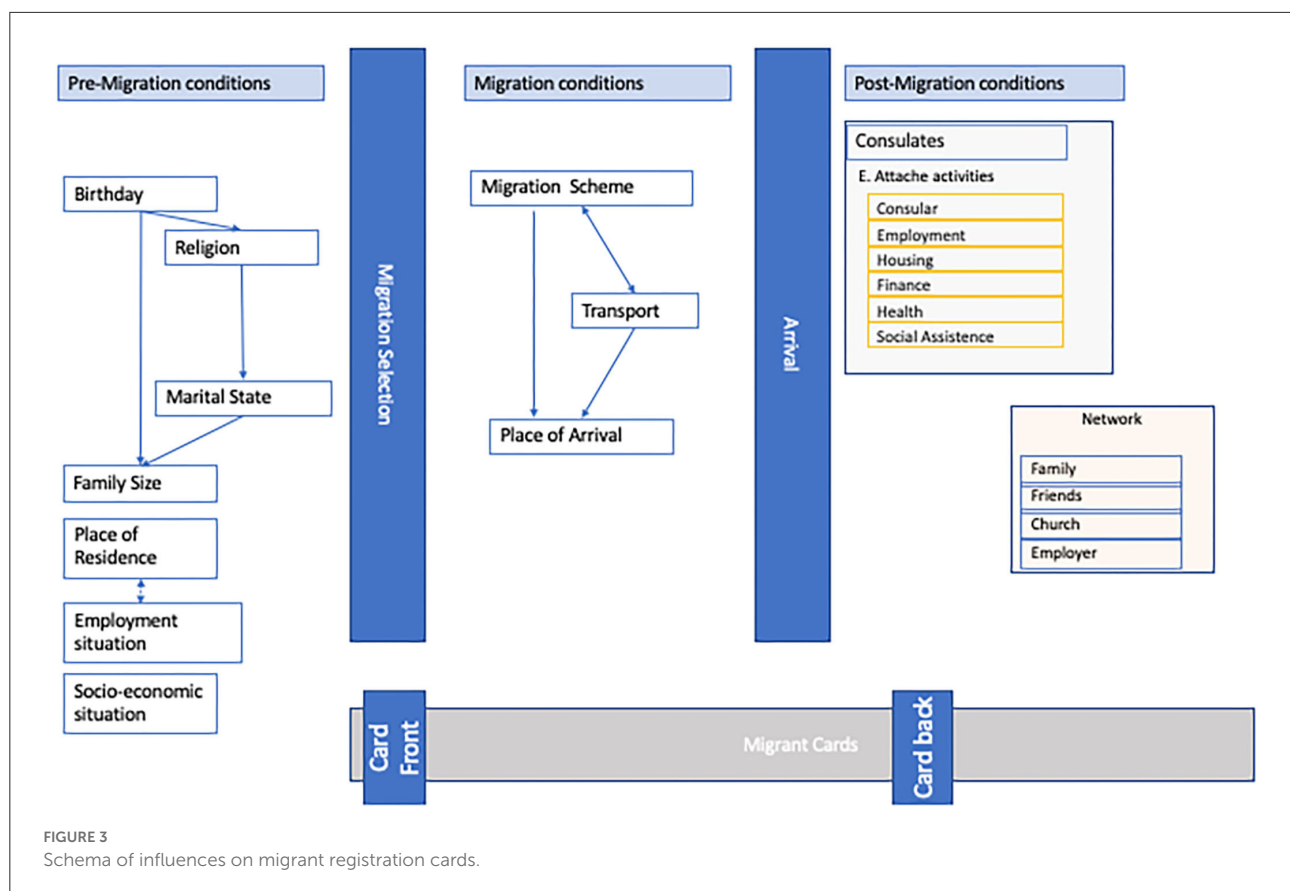
FIGURE 3
Schema of influences on migrant registration cards.

be measured using software (see Figure 2). In this way we do not know *what* is written, but *how much* is written, giving a measure of the information density on the cards. The information density can then be related to what we do know about the content of the cards.
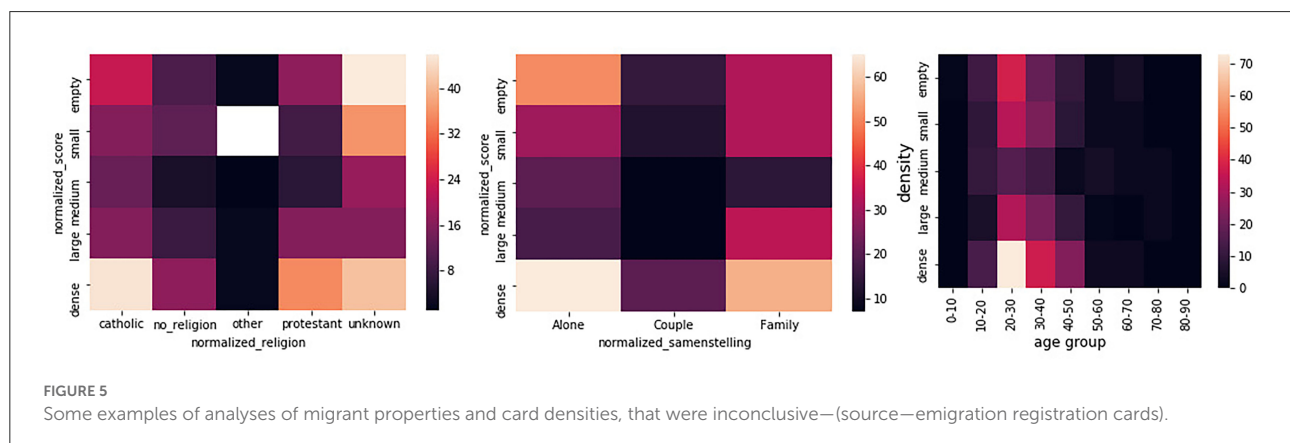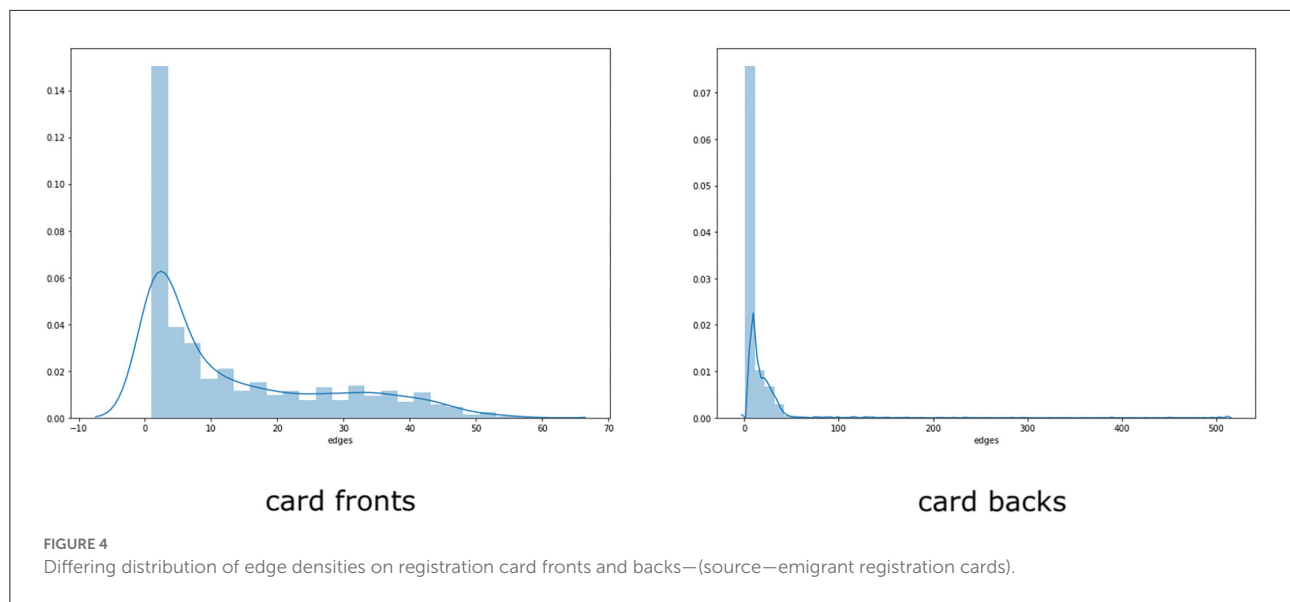
What we know about the cards is analyzed in the sample analysis. We distinguished different stages and influences on the cards, that may be depicted in the schema in Figure 3.

The different stages in the scheme are all visible on the cards. As we wrote above, the front of the cards was (primarily) filled in before migration, and the back of the cards after migration. The card fronts are therefore better structured. There is also a time lag with an average of 2.5 years and a median of 1 year between the date of emigration on the front of the cards and the first dates on the back of the cards. Although the cards sometimes contain information about the travel themselves and the first time after arrival, this is not structural. In the information distribution this translates to a different characteristic for the card fronts compared to the card backs (Figure 4).

Because the cards contain partial reflections of the lives of the migrants, it is obvious to assume that some properties of the migrants would determine the information on the cards. Of course, the cards represent the perspective of the registering authorities, that is the Dutch Emigration service (NED) in the Netherlands and consulate personnel, mainly the emigration attaches and the social work officials in Australia. We have studied the possible relations between all the variables, ranging from age to family composition, religion, place of residence in the Netherlands primary to migration and the migrant scheme (some examples in Figure 5).

Only the place of residence and the migrant scheme showed conclusive influence on the information distribution. This allows for one conclusion, but also leads to a further research question. The conclusion is, that in selecting the largest files (that is the cards with the highest information density) from the registration card files, will not introduce a selection bias for variables such as age, family composition or religion. If we want to study migrant lives, the distribution of information density does not reflect any sub groups among the migrants. Of course, there is a selection bias because cards with a lot of information reflect the most eventful lives, for whatever reason, but this constitutes a point of further study and it is a good idea to compare them with (a selection of) less information dense files.

FIGURE 4
Differing distribution of edge densities on registration card fronts and backs—(source—emigrant registration cards).

FIGURE 5
Some examples of analyses of migrant properties and card densities, that were inconclusive—(source—emigration registration cards).
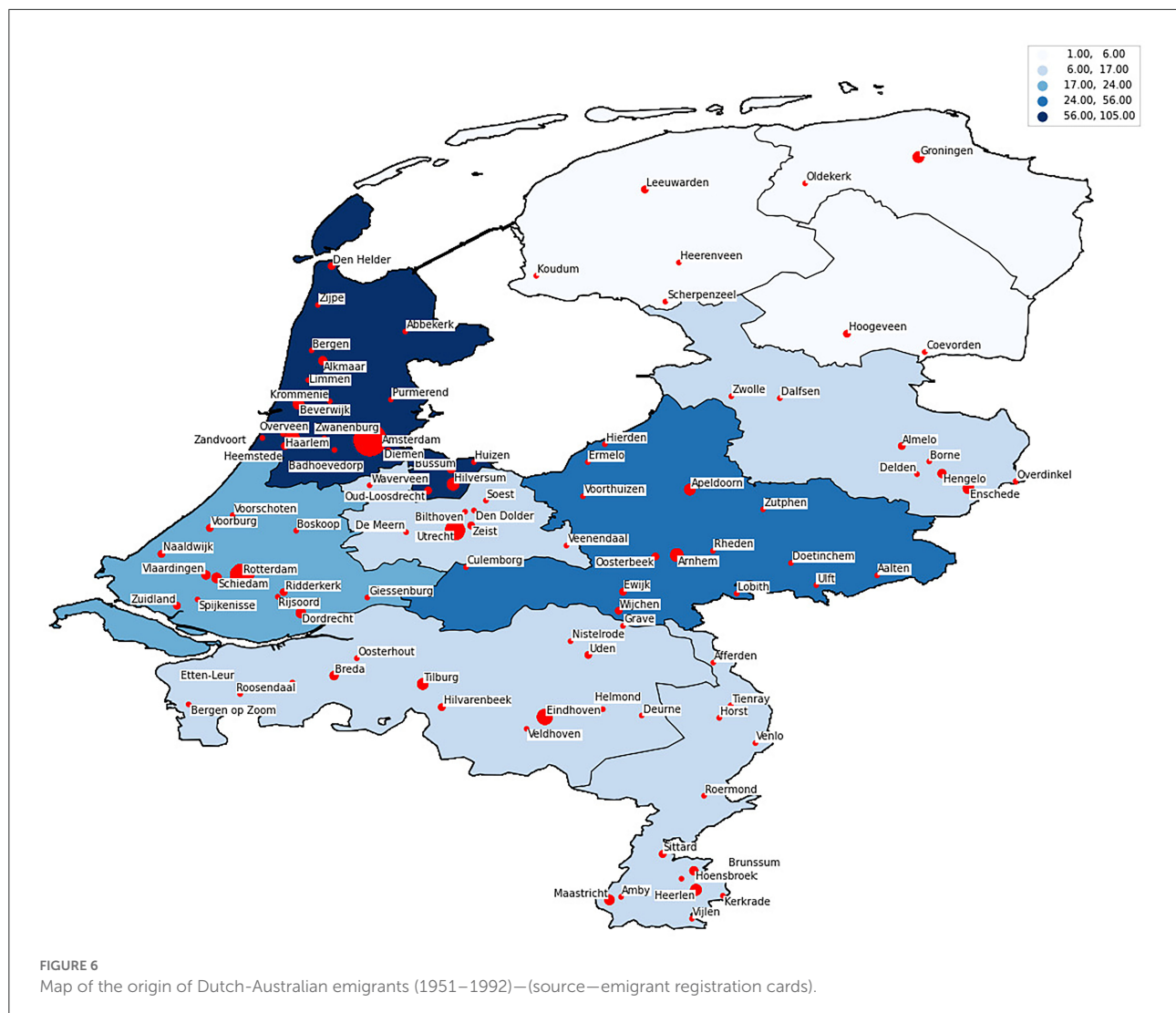
Connecting policy to the registration system

The further research questions that this raised, is why the place of residence and the migration scheme did have a marked influence on the cards (Figure 6). To be able to assess that, we first have to consider the intention of the instrument of the cards for the authorities. It was a type of monitoring device that was used at two different stages of emigration. The first was to record and streamline the emigration process itself. That stage, however, ended with the arrival of migrants in Australia and was primarily recorded on the card fronts. The second stage that was recorded on the card backs (and possible follow-up card) started well after arrival. The question is why the Dutch authorities would care about the Dutch who had left. This is only obvious for the strict consular activities in which official intervention of either a consul or

the ambassador was required, such as passport prolongation and remigration. But the range of activities employed by the consulates was much broader and had to do with the well-being of the migrants. This reveals that there was an active policy by the authorities aimed at making emigration a success (also compare Devereaux, 1996; Torpey, 1998; Shoemaker, 2008).

The two influences of the migration scheme mentioned above reflect different sides of the policy. The variation in place of residence suggests that there was a conscious policy on the part of the authorities to stimulate migration in parts of the country. Although this is known from historiography, the focus has been primarily on the agrarian sector, as post-war emigration policy is understood as a solution for the "small farmers problem" in the Netherlands. This seems in line with the fact that the province of Zeeland is absent and the Northern provinces of Groningen, Friesland and Drenthe are underrepresented in

FIGURE 6
Map of the origin of Dutch-Australian emigrants (1951–1992)—(source—emigrant registration cards).

the sample the map is based on. However, as emigration was meant to be complementary to the industrialization policies (Faassen, 2014a), this also calls for further research, because it seems likely that there also was a relation with post-war changes in the industrial situation of the Netherlands and the closing of the Limburg coal mines in the early 1960s.

The effects of the different migrant schemes are much more subtle. Most migrants traveled under a migration scheme, that is an agreement of the Dutch and Australian governments that subsidized the passage on the migration ships or planes. The most important was the Netherlands-Australian Migration Agreement, that was officially operative from 1951, but there were more schemes (Faassen, 2014a, 165–6). The schemas implied an involvement of both the Dutch and the Australian authorities that found an expression in many areas.

In the NAMA case, Australian co-subsidizing of the passage required migrants to work in Australian government service for 2 years. The schema also included migrant selection with

both Australian and Dutch involvement (Schrover and van Faassen, 2010). On the other side, it also implied that the Dutch authorities wanted to make migration a success. Return migration was always sizable, but the contemporary files in the archives were marked secret as return was seen as failed migration, a label that still dominates historiography. To take away reasons to return, the Dutch authorities invested in social officers that resided at the Dutch consulates in Australia, partly in response to bottom-up pressure from the civil society organizations, who had the majority in the Dutch emigration governance system [(Faassen, 2014a), Ch. 2]. They supported the emigrants by intervening in their affairs, providing assistance in all sorts of social matters. They also were the ones who (predominantly) filled out the backs of the cards. In our sample, we classified several types of events they noted in categories, such as finance, housing, health, labor, social issues, consular/administrative. In combining close reading (sample) and distant reading (edging) of the cards it is possible to reveal
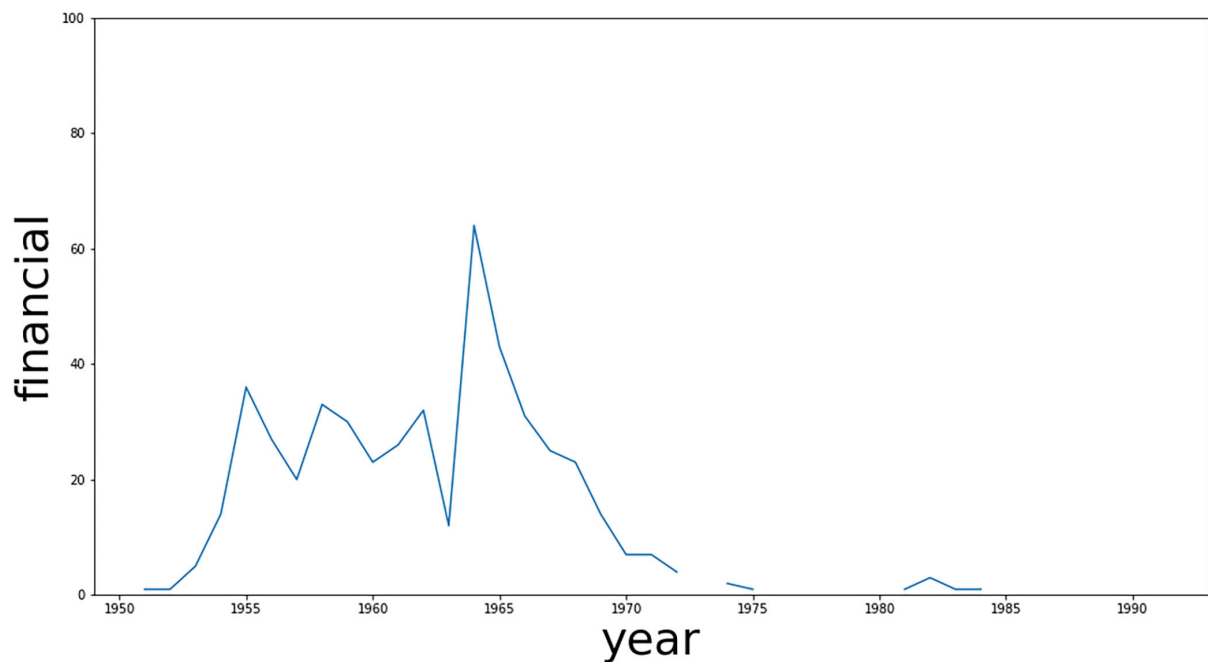
FIGURE 7
Financial events/year—(source—emigrant registration cards).

how policy and agency come together for the events *finance* (Figure 7) and *housing* (Figure 8).

An example of a direct influence of a migration schema is visible in the visualization of financial events. From the graph, it would seem that there were many migrants with financial issues in the mid-1960. Upon closer inspection, however, it appears that these mostly stem from migrants that migrated under the so-called youth program (*Jongeren Programma, JP*) who could only stay for one or 2 years in Australia and had to save with the consulates for their return fare. The sums they saved were notated in succession on the registration cards. Archival research to find an explanation for this phenomenon revealed that this temporary migration, embedded in youth programs, was a deliberate policy of both governments, aiming at increasing the "emigratability" of the Dutch population (when departure figures went down after 1956) by introducing young people to perspectives abroad for a longer period of time. After their return to the Netherlands, the youth could function as "goodwill ambassadors" for emigration, as they were expected to supply emigration supporting information (Faassen and Hoekstra, 2015). Later on, the Youth Programs were succeeded by Working Holiday Schemes (NL-HaNa, 2.15.68, inv.nr. 1351-1353 and 1355-1357) (Figure 8).

One of the other prominent problems in the Dutch-Australian migration was housing. There is a wealth of studies on the first years after arrival especially from the migrant perspective, when emigrants were housed in camps (or, formally,

reception centers) like Bonegilla, Scheyville, Wacol etc. or hostels and families were separated from the moment the man had to start working, for instance to fulfill the NAMA-scheme requirements (Peters, 2001, Ch.4, Walcker-Birckhead, 1988, p. 190–206; Eysbertse, 1997). In contrast with the graph on finance, the housing graph doesn't show specific patterns that prompt analysis. In general, it follows the pattern of the departure peaks and tops, which is rather obvious. However, analyzing the housing events per migration scheme show more variation (Figure 9).

The first thing that stands out is that the Youth Program (JP) showed the least housing events, in contrast with the two largest and simultaneously run schemes NAMA and the Netherlands Government Agency Scheme (NGAS), which have the highest percentages of housing events. This is still rather obvious as migrants under the Youth programs usually traveled alone, or at least without a family. For the other schemes (which are often affiliated with churches and their social networks) it is known that most of the time sponsorships of private persons (including housing) were required. Close reading of the housing events on the cards of the sample however reveals an intriguing shift around 1960–1961. Complaints about the migrant camps and hostels then make way for questions about *Building Societies* and *Housing Committees*. Together with the downward trend in housing events/problems after 1960 in the graph in Figure 8, this can be interpreted as a possible starting point for further analysis on supposed governmental policy on housing. Archival
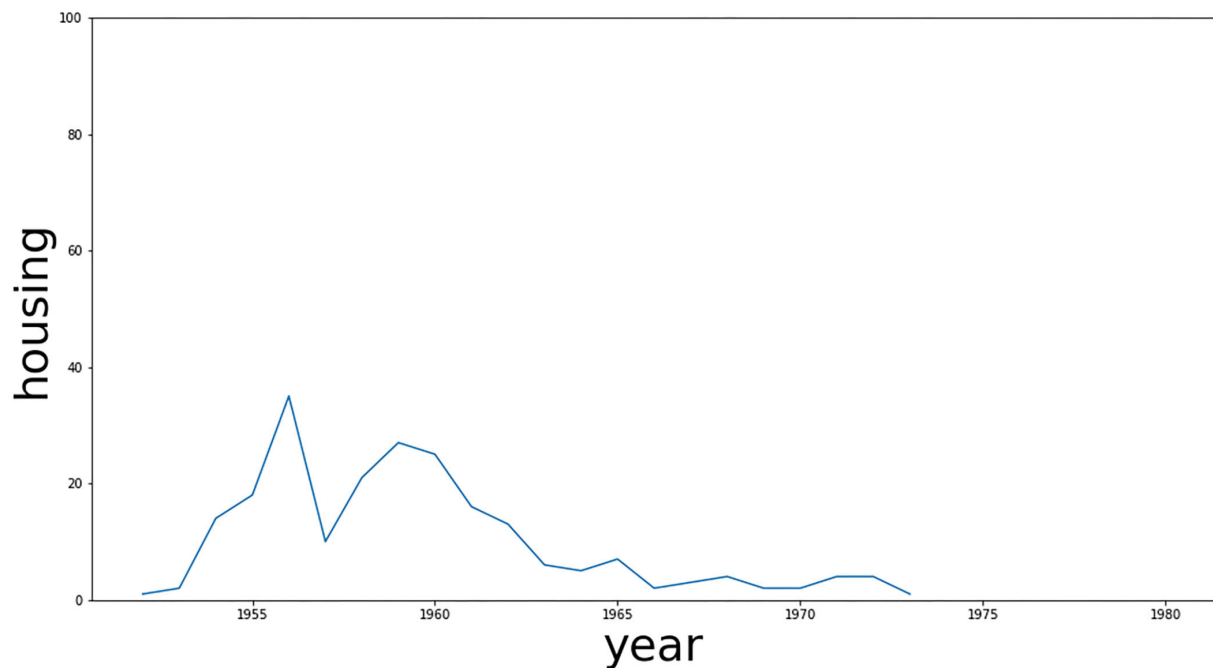
FIGURE 8
Housing events per year—(source—emigrant registration cards).

research in both countries indeed reveals the other side of the camp-stories in migrant studies.

In 1948–1949 the Australian Government organized a conference on housing for "Australians and migrants" in which a policy was formulated to improve the supply of building materials and secure a greater output of houses. One of the suggestions to the Australian states was to order more "prefabricated houses" (NAA, A445/202/3/34). Archival research in the Netherlands shows a twofold response to this idea. The Utrecht Building Company Bredero that already developed the idea of prefabs during the Second World War (Clark, 2002, p. 24–25) took the momentum (1950) to establish an Australian holding to build prefab-houses for Dutch migrants who were employed at the Snowy Mountains Hydro Electric Schemes, which ultimately resulted in the now famous Dutch Australian multinational *Lendlease Corporation*, led by Dick Dusseldorp (Clark, 2002, p. 5, 23–25; Harfield and Prior, 2010; Schlesinger, 2018; Hoekstra and van Faassen, 2022). This initiative was supported by the Dutch government who developed in the early 1950s a policy of financing and sending prefab houses to overseas emigration destinations, in addition to their policy to solve their own domestic housing issues. In researching the specific files, it becomes clear that this policy was based on reports from Dutch emigration attaches abroad. Thus, we can conclude that the Dutch emigration attaches in Australia converted the complaints recorded on the emigrant cards into a more general policy

issue on housing, leading the Dutch government to react with formulating new policy lines (NL-HaNA, 2.15.68, inv.nr. 842, esp. 1385).

Further research on the housing policy leads to the conclusion that the questions about the Building Societies (and Housing Committees) on the cards from the 1960's onwards must in turn be understood as a reflection of a follow up of the "prefab phase" of the housing-policy-abroad, developed by the Dutch government from 1954 and implemented from 1959 till 1975. In order to create a new incentive for emigration to Australia, the Dutch Government designed a policy in four consecutive steps in which money was provided by the Development Loan Fund, Australian Banks and later by Dutch institutional investors. They lend money to so-called Dutch(-Australian) Building Societies, which in turn made it possible for Dutch migrants to borrow money on low interest rates to buy newly built houses (NL-HaNa, 2.15.68, inv.nrs 1262–1265). There even seems to be a closer connection with the Dutch initiated real estate business in Australia, although the exact relations require further investigation.

Both examples leads to the conclusion that the card system was not only a one way monitoring or surveillance device, but that it was a constant form of systemic interaction with input and feedback loops between migrants' experiences abroad (e.g., complaints about housing in camps and hostels to the emigration attaches) and policymaking in the homeland (sending prefab houses for migrants, followed by financial

	NAMA	LP	other	JP	NESS	NGAS
social assistance	30%	39%	34%	27%	17%	32%
employment	25%	23%	18%	24%	12%	18%
financial	13%	8%	17%	33%	4%	26%
consular tasks	13%	20%	15%	9%	54%	10%
housing	8%	5%	5%	1%	4%	7%
administration	7%	4%	8%	5%	8%	2%
health	3%	1%	2%	0%		3%
unknown	1%		1%	0%	2%	2%

FIGURE 9
Event-per-scheme in percent—(source—emigrant registration cards).

incentives for building activities in Australia), in order to reduce the rate of return migration.

Scope of the migrant registration system

Although we now have a rather good insight in the intention and the focus of this registration system as a policy instrument, this still leaves the question about the exact scope of the registration, what is hidden or missing. Summarizing our previous institutional research, we know that the phase of selection consisted of a pre-selection process (Figure 3), in the Netherlands carried out by the private organizations mentioned above. These organizations also left archival collections and card systems. The question arises if and how these index systems are related to "our" registration system. While the mix of manuscript and typewriting makes analyzing the information on the cards still difficult, the card models (prototypes and pre-printed) can be read by OCR. In this way of distant reading, we could get a complete overview of all card models

in the registration system. Once again, we complemented this by close reading cards and connecting them to policy files. This mixed methods approach of close and distant reading allowed us to conclude that the file reference numbers on "our" registration system (ref. nr. A/920/41988) formed the linchpin between the pre-selection files (ref. nr. A/920) kept in the private organizations in the Netherlands and the application files (ref.nr. 41988), now preserved in the National Archives Australia (Faassen and Oprel, 2020, Figure 10). Our registration system literally contains the master cards (*stamkaart*) on the migrants' application files.

Combining the fact that there are several information models of the master card and that not all master cards actually have a reference file number, revealed that cards also could be filled in Australia after arrival. These migrants somehow have been able to "avoid" the selection procedures. This can be explained partially by departures in the late 1940s, early 1950s (before this selection procedure was perfected or because they departed from other places like the ex-servicemen serving in the Netherlands East Indies) but also by migrants who were not the primary target of the policymakers, the more expat-like

FIGURE 10
File references codes and subsystems of the migrant registration card index.
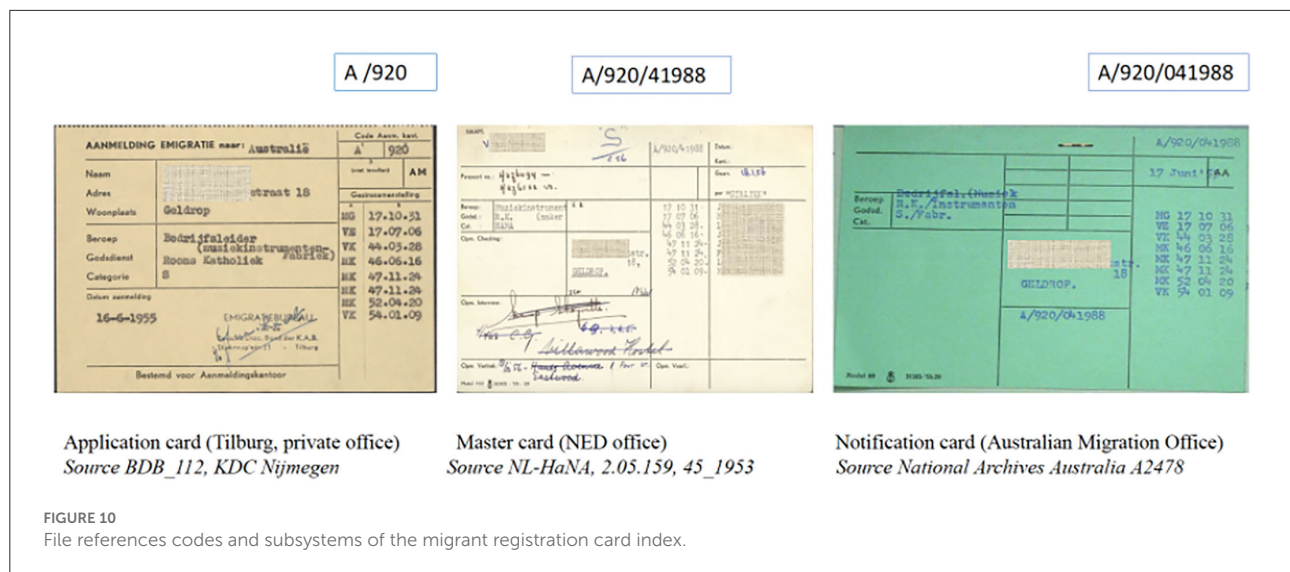
migrants from companies or industry. Thus, there has been a hybridity in using the migrant registration system. Another conclusion that can be drawn from this exercise is that pre-selection also means that there are dropouts before migration, in other words, aspiring migrants who never left for whatever reasons. Summing up: the master cards registration system gives information on those who went, but this information can range from very summary to very extensive, it can contain more hidden information on non-targeted migrants or on return migrants and it contains no information on migrants who did not pass the selection procedures.

In the next section, we will elaborate how until today these policies have resulted in making some migrants invisible, and others iconic in the archival collections and the storytelling and historiography based on these collections.

From iconic to invisible migrants

In many ways, Adri Zevenbergen has become one of the most iconic migrants of the Dutch-Australian migration. This was the intention of the authorities, both the Dutch and the Australian, that singled her out as the 100,000th migrant and carefully orchestrated and documented her migration (NL-HaNA, 2.15.68, inv.nr. 1400). She got a lot of media coverage in both countries and both the Dutch and the Australian national archives preserved a lot of (identical!) photos from Adri Zevenbergen. They show her in the Dutch village of Abbebroek, with a windmill and a sixteenth century church, where she lived among people "who still wore clogs" according to the captions—to show that migration to Australia means entering a modernized world, to the migration travel that started with a trunk imprinted with "100,000 migrant," and the ship Johan van Oldenbarnevelt where she had a pleasant travel to her arrival in Australia. There she was offered the keys to a brand-new house

in the suburbs of Geelong, near Melbourne. Her husband was offered a job on the Shell refineries in Australia. Her selection as the 100.000th migrant was in line with the governmental policies to stimulate emigration. As was the case with the 50.000th emigrant, again a woman was selected, to persuade Dutch "housewives," who—in the eyes of the emigration authorities were more vulnerable to homesickness—to emigrate as well.

A propaganda subject, her migration was anything but typical, if only because most Dutch-Australian migrants had trouble finding suitable work and housing, as was explained above. Nowadays, in both Australia and the Netherlands Zevenbergen has disappeared from the collective memory, leaving the Dutch Australian community confused about their invisibility. But when this emigration wave was commemorated during a public diplomacy visit of the Dutch King Willem Alexander and his wife in 2016, the now partially digitized archives showed Zevenbergen once again as a typical Dutch-Australian migrant. For a Dutch documentary in 2018 even her thoroughly Australianised son Addo, who did not speak Dutch anymore, was interviewed and reflected on his mother's celebrity (Omroep Max, Documentary Vaarwel Nederland, 2018, destination Australia 7649080; images NED-fotoarchief, 0833 + 0834. NAA: A12111, 1/1958/4/39 barcode 7529953, 1/1958/4/70 barcode 7529984, 1/1958/4/45, 1/1959/13/22; A2478, Zevenbergen C).

On the same voyage as Adri Zevenbergen and her family on the migrant ship Johan van Oldenbarnevelt were a group of 37 Moluccans who had earlier migrated to The Netherlands as stowaways, regretting their previous choice of Indonesian citizenship after the decolonization wars (Eijl, 2012). They were evicted from The Netherlands and forcefully repatriated to Indonesia in a separate section on the ship that had been constructed especially for them (NL-HaNA, 2.15.68, inv.nr. 1400). In the Netherlands, there was a lot of protest against the eviction of these young men who were former Dutch after

all. Like Zevenbergen, these 37 stowaways were emigrants, and they traveled on the same ship, though not to the same destination. The stories of emigrants and stowaways normally do not come together because they are in different ministerial collections and even if emigration and politics of eviction coincided, their connection is only revealed by connecting the separate collections.

These two stories of the passengers on a same journey aboard the Johan van Oldenbarnevelt illustrate some of the determinants for the visibility of migrants and the role of cultural heritage institutions and digitization. Cultural heritage institutions focus on storytelling and claim to tell the story of historical events and groups of people by highlighting the story of individuals. While this makes a historical phenomenon come alive, it is from the perspective of an individual who comes to stand for the whole phenomenon. In this way, iconic migrants are created. Digitization usually reinforces this, as the cultural heritage objects of these iconic migrants are prioritized in digitization efforts.

While most migrant cultural heritage materials are not digitized, archives tend to prioritize digitisation of the kind of registrations systems we used in our project and which can be found worldwide (Faassen and Oprel, 2020). They give an overview of many migrant names and are usually systematic in contrast to policy and individual case files that are patchy and often disorganized. Australian archival policy also tends to prioritize the digitization of passenger lists, which also consist of long lists of names of migrants, disembarkation schemes and sometimes even information on individual migrants or migrants' groups, traveling under the same scheme. These lists can be very helpful in giving invisible migrants (or their children) more grip on their own history, as we figured out in several pilots done by interns on the project. When migrants cannot be found in our registration system but the ship they traveled on is known, their families who stayed in the Netherlands can at least imagine the events from the journey through the eyes of fellow travelers with more extensive information in our card index system.

As we said above, the schema of the process of filling out cards makes it possible to assess who were not in the cards either. A very special and often forgotten category by researchers are those who were in the selection process for migration but never left for whatever reason. Fortunately, in some cases the personal selection files have survived in other archives such as the Catholic Documentation Centre (KDC) in Nijmegen or the International Institute for Social History in Amsterdam, making it possible to link those who did emigrate and compare them with those who stayed (Kosterman, 2021).

The objectives of the consulates also left their traces in the registration. As they were primarily focused on consular work on the one hand and social assistance on the other, these are the interactions that are most prominent in the registrations. Therefore, they contain lots of information about housing and employment, also the policy targets of the Dutch government.

There is very little direct information about health issues, except when they led migrants and their families into social trouble such as poverty or were reasons to remigrate. On the other hand, situations where things got out of hand because migrants socially derailed were common. In some cases, migrants who appeared in the Australian newspapers because they committed a crime, already were trouble in the consular offices. Also, if interventions in family matters threatened or were necessary, these are also reflected on the cards, such as cases when one of a migrant family suffered from mental disorders, causing trouble to his or her family or in cases of (suspicions of) child molesting (Faassen and Oprel, 2020). These traces can be followed to mental hospital registers in Australia where names of Dutch migrants can be found (Faassen, 2014b).

The consular work itself left trails too, usually if migrants or remigrants needed passport renewal in order to travel. This sometimes (lightly) documents the migrants who would not appear in the registration because they were sent by their companies. For other collections, similar remarks can be made, but they all tend to highlight different aspects of migrants, therefore connecting them leads to a more complete view of migrants. But because they are dispersed, this is only possible with a cross collection and cross institutional approach. As it is, institutional policies, archiving practices and institutional collection policies such as the destruction of executive files all contribute to the presence or absence and visibility or invisibility of migrants and migrant groups in the archives. And, as we said before, storytelling approaches and selective digitization policies tend to single out the most visible migrants and unconsciously accentuate past policies.

On the other hand, the creation of alternative collections by using social media, creates even more possibilities to better represent the unheard voices and make migrants visible in the existing heritage when combined with connected collections. On social media there are often calls from emigrants who are looking for their family or lost friends. In a small pilot we found out for several cases that the registration system does contain rather elaborate cards and sometimes even a whole file (with photos at the time of immigration) in the Australian immigration files. Perhaps social media would yield more information, but that depends on whether the question reaches the right people. In addition, it is doubtful whether these contain information from the archives.

A last Dutch-Australian emigrant may serve to highlight some other aspects of iconicity. This is Dick Dusseldorp (1918–2000), mentioned above. He was an industrial tycoon who became rich, famous and powerful in Australia, first as a builder and later as the founder and president of the Lend-Lease company that still is a world-wide building and financing concern. He is still famous in Australia, but unknown in the Netherlands. He is an icon but not an important figure in most of the collections of either the Australian or Dutch archives. As he was sent to Australia by his Utrecht Building Company Bredero for the prefab housing project, his migration was not recorded

in the registration system in the first instance. However, he was registered once his passport had to be renewed. Due to his public appearances and the high-profile construction activities, his story can be reconstructed and added to the data backbone using digitized press archives, such as newspapers and even YouTube films.

Conclusion

Digitization and digital methods can make important contributions to the analysis of collections and to expose and even overcome policy biases. However, this potential is complicated when digitisation and digital methods tend to decontextualise archive material and contribute new distortions and biases. Moreover, if digital methods are applied within the confinements of a single collection, they cannot contribute a more encompassing picture that results from combining the many facets in other public and private heritage collections. A more balanced use depends on the connection of different collections and using digital methods as an extension of methods of source criticism. Only if they are employed in this way, it becomes possible to assess the pervasiveness of policy in holdings of archives, other collecting institutions and private resources alike. Migration history is par excellence suited to show the way. The dispersed nature of migration heritage not only implies that it is important to connect collections to get an integral view of migration history and more grip on the variety of migrants, but it also makes it clear that there are fundamentally different views that are codified in collections.

We exemplified this with our study of Dutch-Australian migration by using a registration system as a connecting device between collections and by scrutinizing the impact of policy on this system. Our analysis depended on the realization that the registration cards were an instrument of the registering authorities and their own policies were much more influential than the characteristics of the migrants. While computer vision was instrumental in measuring the information distribution in the card system, further analysis was impossible without close reading policy files. This illustrates that for historical research computer assisted methods can provide an important extension of the methodology, but that they are most effective if combined with established methods.

The assessment of the information in the registration cards was established to prevent selection biases in further sampling migrant lives and connecting them to dispersed cultural heritage materials. On the basis of that assessment, we now know that further sampling will not give a selection bias toward specific groups of migrants, but we should include the policies of the emigration authorities in our analysis. This underlines Schrover and Moloneys' statement that it is necessary to acknowledge the interaction between policy and migrants' choices. We can even go one step ahead: only the whole assessment of information revealed that states do use other, more hidden or implicit forms

of categorization, like the migration schemes based on economic principles. Thus, researchers should also take into account that there are micro forces such as seemingly "neutral migration schemes" that can influence migrant's life course experiences. Furthermore, in combining close reading (sample) and distant reading (edging) of the cards we are able to conclude that the card system was not only a one-way monitoring device, but that it was a constant form of systemic interaction with input and feedback loops between migrants' experiences abroad and policymaking in the homeland.

Next to that we have presented some iconic migrants as illustrations of what makes migrants visible and invisible in digital cultural heritage collections. Our main point is that visibility is largely determined by a chain of policies on the part of archive creators and curators, both analog and digital. Connecting collections from dispersed institutions add different perspectives to the larger view on migrants. In this way connected digital resources transcend collection limitations and provide more possibilities to make migrants visible, either directly or by providing context.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://github.com/HoekR/MIGRANT/blob/master/results/exploring_data_integration/notebooks/Profiles.ipynb.

Author contributions

MF and RH are collaborators on the project Migrant: they both conducted the research and they both wrote an equal share of this article (MF on the migration history, RH on the digital methodology: both on source criticism). Both authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Arthur, P., Ensor, J., Faassen, M. van., Hoekstra, R., and Peters, N. (2018). Migrating people, migrating data. Digital approaches to migrant heritage. *J. Jap. Assoc. Dig. Human.* 3–1, 98–113. doi: 10.17928/jjadh.3.1_98

Blouin, Jr., Francis, X., and Rosenberg, W. G. (2011). "On the Intersections of Archives and History," in *Processing the Past: Contesting Authorities in History and the Archives.* Oxford Academic. doi: 10.1093/acprof:oso/9780199740543.003.0001 (accessed August 5, 2022)

Chassanoff, A. (2013). Historians and the use of primary source materials in the digital age. *Am. Arch.* 76, 458–480. doi: 10.17723/aarc.76.2.lh76217m2m376n28

Clark, L. (2002). *Finding a Common Interest. The Story of Dick Dusseldorp and Lend Lease.* Cambridge: Cambridge University Press.

Coburn, J. (2021). Defending the digital: awareness of digital selectivity in historical research practice. *J. Librariansh. Inf. Sci.* 53, 398–410. doi: 10.1177/0961000620918647

Coté, J. (2010). The indisch Dutch in post-was Australia. Low Countries *J. Soc. Econ. Hist.* 7, 103–125. doi: 10.18352/tseg.378

Devereaux, S. (1996). The city and the sessions paper: "Public Justice" in London, 1770-1800. *J. Br. Stud.* 35, 466–503. doi: 10.1086/386119

Eijl, C. van. (2012). *Tussenland. Illegaal in Nederland 1945-2000.* Verloren: Hilversum.

Eysbertse, D. M. (1997). *Where Waters Meet. Bonegilla, the Dutch Migrant Experience in Australia.* Erasmus Foundation North Brighton (VIC): Melbourne, VIC.

Faassen, M. van. (2014a). *Polder en Emigratie. Het Nederlandse Emigratiebestel in Internationaal Perspectief 1945-1967.* The Hague: Huygens ING.

Faassen, M. van. (2014b). *Emigratie 1945-1967.* Available online at: http://resources.huygens.knaw.nl/emigratie (accessed March 30, 2022).

Faassen, M. van., and Hoekstra, R. (2015). "Tijdelijke migratie als bewust beleid. De Nederlands-Australische Youth Programs en Working Holiday Schemes," in Presentation CGM (The Hague). Available online at: https://www.researchgate.net/profile/Marijke-Van-Faassen (accessed August 5, 2022).

Faassen, M. van., and Hoekstra, R. (forthcoming). "Storytelling, identity and digitizing heritage," in *Digitising Heritage. TransOceanic Encounters Between Australia and Europe*, eds C. Wergin and S. Affeldt (Heidelberg: KEMTE, Heidelberg University Publishing).

Faassen, M. van., and Oprel, M. (2020). "Paper trails to private lives. The performative power of card indexes through time and space," in *Information and Power in History: Towards a Global Approach*, eds I. Nijenhuis, M. van Faassen, R. Sluijter, J. Gijsenbergh, and W. de Jong (Routledge), 254–274.

Gilliland, A. J., McKemmish, S., and Lau, A. J. (2017). *Research in the Archival Multiverse.* Clayton: Monash University Publishing.

Harfield, S., and Prior, J. H. (2010). "A bright new suburbia? GJ Dusseldorp and the development of the Kingsdene Estate," in *Green Fields, Brown Fields,New Fields: Proceedings of the 10th Australasian Urban History, Planning History Conference*, eds D. Nichols, A. Hurlimann, C. Mouat and S. Pascoe (Melbourne, VIC: University of Melbourne), 7–10, 207–220.

Hauswedell, T., Nyhan, J., Beals, M. H., Terras, M., and Bell, E. (2020). Of global reach yet of situated contexts: an examination of the implicit and explicit selection criteria that shape digital archives of historical newspapers. *Arch. Sci.* 20, 139–165. doi: 10.1007/s10502-020-09332-1

Hoekstra, R. (2021). *Summary Analysis Notebook.* Available online at: https://github.com/HoekR/MIGRANT/blob/master/results/exploring_data_integration/notebooks/Profiles.ipynb (accessed August 10, 2022).

Hoekstra, R., and Koolen, M. (2019). Data scopes for digital history research. *Historical Methods J. Quan. Interdiscip. History* 52, 79–94. doi: 10.1080/01615440.2018.1484676

Hoekstra, R., Koolen, M., and van Faassen, M. (2021). Vested authorities, emergent brokers and user archivists: power and legitimacy in information provision. *J. Comput. Cult. Herit.* doi: 10.1145/3484481

Hoekstra, R., and van Faassen, M. (2022). "Hoog-Catharijne en het Sydney Opera House," in *Nog meer Wereldgeschiedenis van Nederland*, eds L. Heerma van Voss, N. Bouras, M. t Hart, M. van der Heijden, and L. Lucassen (Amsterdam: Ambo-Anthos).

Horne, S. (2011). *The Invisible Immigrants: Dutch Migrants in South Australia.* An arts Internship Project for the Migration Museum. Adelaide: University of Adelaide. Available online at: https://apo.org.au/node/74461 (accessed August 09, 2022).

Jeurgens, C., Karabinos, M. (2020). Paradoxes of curating colonial memory. *Arch. Sci.* 20, 199–220. doi: 10.1007/s10502-020-09334-z

Ketelaar, E. (2017). "Archival turns and returns," in *2017 Research in the Archival Multiverse*, eds A. J. Gilliland, S. McKemmish, and A. J. Lau (Clayton: Monash University Publishing), 228–269.

Koolen, M., Hoekstra, R., Nijenhuis, I., Sluijter, R., van Gelder, E., van Koert, R., et al. (2020). "Modelling resolutions of the dutch states general for digital historical research," in *Collect and Connect: Archives and Collections in a Digital Age 2020. Proceedings of the International Conference Collect and Connect: Archives and Collections in a Digital Age* (Leiden). Available online at: http://ceur-ws.org/Vol-2810/paper4.pdf (accessed July 5, 2022).

Kosterman, F. (2021). "Op zoek naar een beter leven. Gecombineerde datasets brengen levensgeschiedenissen van Nederlandse migranten naar Australië in kaart," in *Impressie, Tijdschrift voor Katholiek Erfgoed*, 18–22. Available online at: https://www.ru.nl/kdc/weten/impressie-nieuwsbrief-kdc/ (accessed August 5, 2022).

Leurs, K (2021). "Personal digital archives: doing historical migration research in the digital era," in *Doing Historical Research on Migration in the Digital Age: Theories, Concepts and Methods. Luxemburg: IMISCOE Conference 2021.*

Lucassen, J. M. W. G., and Lucassen, L. A. C. J. (1999). "Introduction," in *Migration, Migration History, History: Old Paradigms and New Perspectives* (Bern; New York, NY: Peter Lang), 9–38.

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course With Examples in R and Stan.* London: Chapman and Hall/CRC.

McKemmish, S. (2017). "Recordkeeping in the continuum: an australiantradition," in *Research in the Archival Multiverse*, eds A. J. Gilliland, S. McKemmish, and A. J. Lau (Clayton: Monash University Publishing), 122–161.

Nationaal Archief (2022). *Proces van Digitalisering.* Available online at: https://www.nationaalarchief.nl/archiveren/kennisbank/proces-van-digitalisering (accessed July 5, 2022).

National Archives of Australia (2022). *Corporate Plan 2017-18 to 2020-21.* Available online at: https://www.naa.gov.au/about-us/our-organisation/accountability-and-reporting/our-corporate-plans/corporate-plan-2017-18-2020-21 (accessed July 5, 2022).

Nauta, G. J., van den Heuvel, W., and Teunisse, S. (2017). *D4.4. Report on Revision ENUMERATE Core Survey 4.* Europeana/DEN. Available online at: https://www.den.nl/uploads/5c6a8684b3862327a358fccaf100d59668eedf1f7f449.pdf (accessed March 30, 2022).

Nyirubugara, O. (2012). *Surfing the Past: Digital Learners in the History Class.* Leiden: Sidestone Press.

Oberbichler, S., and Pfanzelter, E. (2021). Topic-specific corpus building: a step towards a representative newspaper corpus on the topic of return migration using text mining methods. *J. Dig. History.* jdh001. Available online at: https://journalofdigitalhistory.org/en/article/4yxHGiqXYRbX (accessed August 9, 2022)

Omroep Max. (2018). *Documentary.* Available online at: https://www.maxvandaag.nl/programmas/tv/vaarwel-nederland/australie/POW_03876992/, 25:43 (accessed March, 2022).

Peters, N. (2001). "Milk & honey - but no gold," in *Postwar Migration to Western Australia Between 1945–1964* (Perth, WA: UWA Publishing).

Peters, N. (2010). Aanpassen and invisibility: being Dutch in post-war Australia. Low Countries *J. Soc. Econ. Hist.* 7, 82–102. doi: 10.18352/tseg.377

Schlesinger, L. (2018). *Lendlease's Dick Dusseldorp: From Nazi Labour Camp to Listed Giant. Financial Review* (september 2018). Available online at https://www.afr.com/property/residential/lendleases-dick-dusseldorp-from-nazi-labour-camp-to-listed-giant-20180702-h1252g (accessed August 8, 2022).

Schrover, M., and Moloney, D. M. (2013). *Gender, Migration and Categorization. Making Distinctions Between Migrants in Western Countries 1945-2010.* Amsterdam: Amsterdam University Press.

Schrover, M., and van Faassen, M. (2010). Invisibility and selectivity. Introduction to the special issue on Dutch overseas migration in the nineteenth and twentieth century. Low Countries *J. Soc. Econ. Hist.* 7, 3–31. doi: 10.18352/tseg.374

Shoemaker, R. B. (2008). The Old Bailey proceedings and the representation of crime and criminal justice in eighteenth-century London. *J. Br. Stud.* 47, 559–580. doi: 10.1086/587722

Smith, A. L. (2003). "Introduction," in *Europe's Invisible Migrants*, ed A. L. Smith (Amsterdam: Amsterdam AUP), 9–33.

Stoler, A. L. (2002). "Colonial archives and the arts of governance: on the content in the form," in *Refiguring the Archive*, eds C. Hamilton, V. Harris, J. Taylor, M. Pickover, G. Reid, and R. Saleh (Dordrecht: Springer), 83–102.

Tames, I. (2022). *Digitized Archives and Stateless People: New Windows on Those Usually Invisible*. Available online at: https://medium.com/@ismee.tames/digitized-archives-and-stateless-people-new-windows-on-those-usually-invisible-c85357c3c1bc (accessed July 5, 2022).

Thylstrup, N. B. (2019). *The Politics of Mass Digitization*. Massachusets: MIT Press.

Torpey, J. (1998). Coming and going: on the state monopolization of the legitimate 'means of movement'. *Soc. Theory* 16, 239–259. doi: 10.1111/0735-2751.00055

Tufekci, Z. (2017). *Twitter and Tear Gas*. New Haven: Yale University Press.

Upward, F., Reed, B., Oliver, G., and Evans, J. (2018). *Recordkeeping Informatics for a Networked Age*. Clayton, NC: Monash University Publishing.

Walcker-Birckhead, W. (1988). *Dutch identity and assimilation in* Australia: an interpretative approach [Dissertation]. Australian National University, Canberra, ACT, Australia.

Willems, W. H. (2001). *De Uittocht uit Indië 1945-1995*. Amsterdam: Bert Bakker.

Willems, W. H. (2003). "No sheltering sky: migrant identities of Dutch nationals from Indonesia," in *Recalling the Indies. Colonial Culture & Postcolonial Identities*, Eds J. Coté and L. Westerbeek (Amsterdam: Aksant), 33–61.

Yalim, Ö. (2008). "Passing on 'Invisible' Histories. New media and the Cultural Heritage of Immigrant Women in the Netherlands," in *Traveling Heritages. New Perspectives on Collecting, Preserving and Sharing Women's History*, ed S. E. Wieringa (Amsterdam: Aksant), 185–198.

Human journeys in the digital age: Advances and challenges in Digital Historical Migration Studies

Paul Longley Arthur* and Isabel Smith

School of Arts and Humanities, Edith Cowan University, Perth, WA, Australia

Accelerations in migration, mobility, and processes of globalization in recent decades have intersected with parallel developments in information and communications technology (ICT). These advances have had profound influences on historical and cultural research. With reference to a diverse range of international projects, this paper outlines major directions and opportunities in the growing field of Digital Historical Migration Studies (DHMS). The "digital turn" brings opportunities for integrating data on macro and micro scales, and finding new ways to combine and explore tensions between quantitative and qualitative materials, and between external observations of migrants and migration and self-representations by migrants. The plural and fluid nature of digital content also lends itself to multifaceted representations of migration that illustrate the complexities of lived experiences, and individual and collective identities. At the same time, digitalization in historical migration studies underscores the tensions between technological advances and methodological shifts, the need for self-reflexive approaches, the politics and power structures underlying migration data, and the ethical concerns around protecting migrants' data, privacy, and agency.

Introduction

Migration and the mass movement and mobility of people around the world has been accelerating rapidly along with the macro processes of globalization—economic, political, cultural, and social—over recent decades. Since the late twentieth century these trends have been supported and enabled by parallel developments in information and communications technology (ICT) and computing. In migration studies, these digital innovations have led to new data-driven methods and the creation of vast online resources, as well as social networks and platforms that have sustained and expanded migrant communities and diasporas. This has opened up the growing field of Digital Historical Migration Studies (DHMS). Related to but distinct from migration studies and digital history, DHMS centers upon the development of migration history research through digital technologies. This paper surveys some of the key opportunities and challenges of DHMS. Drawing on a diverse range of international examples, with

particular reference to Europe, the United States, and Australia[1], it highlights advances such as the unearthing of new collections of data, the development of new analytical tools and models, and the rich qualitative materials found in digital storytelling. It focuses on examples that demonstrate such advances, but also methodological and theoretical tensions in the field such as navigating the power dynamics between researchers and migrants, and individuals and states, as well as the ethical considerations in protecting migrants' data, agency, and privacy. Case studies generally focus on digital historical migration research, though due to the interdisciplinary nature of DHMS they also draw upon insights from neighboring disciplines such as social research, social media analysis, and data analysis. In this way the paper aims to assist migration researchers in thinking about the benefits and challenges of DHMS, potential tools for integration, and key considerations and questions for future research. Put more broadly, it aims to reflect on what the digital brings to historical migration research and what historical migration research brings to the digital.

Expanding data collections

Access to new data is one of the core drivers and foundations of DHMS. Data sets available to researchers are rapidly expanding and diversifying, going well-beyond numbers and metrics of people, places, and dates. At 2021's International Migration Research Network annual conference, researchers created, drew on, and analyzed data to investigate topics as varied as the relationships between migration policy and the subjective well-being of non-immigrant populations in Europe, the aspirations of refugee children who have arrived in Greece, and the barriers and opportunities facing economic migrants during the COVID-19 pandemic in the Czech Republic (Gheorghiev, 2021; Palaiologou et al., 2021; Tatarko et al., 2021).

Much of this research is based on born-digital data. Digital formats have dramatically expanded the possibilities for recording and tracking data in multiple, diverse forms. In particular, the prevalence of smartphones and social media has created many new opportunities for capturing previously undocumented information (Harari et al., 2016). Katja Kaufmann explains that smartphones are not only crucial lifelines and tools for migrants in managing logistics, creating connections, and maintaining relationships, but the technology also offers significant opportunities for "mobile methods" in digital migration studies — "using mobile communication technologies to study the social world" (Kaufmann, 2020, p. 168). Kaufmann refers to the new means of knowledge

---

1 The authors acknowledge that this paper is Western-centric in its focus. Future discussions of DHMS would benefit from expanding the scope to regions and frameworks beyond European and settler-colonial examples.

production enabled by smartphones, capturing rich stores of data coproduced between migrants and researchers through combinations of self-reported and automatically logged data. She cites the accessibility of customizable research apps, as well as her own research interviewing Syrian refugees, alongside which she used WhatsApp exchanges and in-person sharing of smartphone data such as photographs, screenshots, and emojis to enhance interview materials (Kaufmann, 2020).

The digitalization of analog archival materials has also laid foundations for DHMS. Digitization projects by state museums, archives, and other collecting bodies in particular have targeted materials relevant to migration histories. In Australia, an early and large-scale example of such a project was the Western Australian Museum's retrospective digitization of its "Welcome Walls" in 2012. These physical walls comprise over 400 panels with the names inscribed of more than 45,000 migrants who arrived at the ports of Fremantle and Albany in Western Australia (Western Australian Museum, 2022). Inscriptions are based on user-contributed data: registration forms detailing key information on migrants were completed by community members, a large number of whom were the migrants of interest, or relatives and direct descendants (Joseph et al., 2013). Users are able to search an online database for passenger names, arrival dates, ship names, and biographical data including family relations, and discover the physical panel number on which individual migrant names can be found. More recently, in 2021, Australia's Victoria State Government awarded an $800,000 grant to the Bonegilla Migrant Experience museum to digitize and make available its collection of migrant identification cards and other records held with the National Archives of Australia (NAA) (Brown, 2021). The museum—formerly the site of the Bonegilla Migrant Reception and Training Center, Australia's largest and longest-running reception center—has been digitizing ID cards of the more than 300,000 people who came through its doors.

Along with these discrete and targeted collections are databases bringing together larger patchworks of historical migration data. The NAA's online collections offer a consolidated space for materials including passenger, citizenship, and proof of arrival records and photographs. Similarly, the work of researchers has been transformed by the creation of Trove, an online database and discovery service hosted by the National Library of Australia (NLA) that includes digital collections from state and national museums, libraries, archives, and other organizations, as well as text-searchable content of over 700 historical Australian newspapers. The NLA also has a series of "research guides" on its website that compile themes of data for public access—notably including Australian Indigenous family history. These materials include births, deaths, and marriages records of Aboriginal people; newspaper press clippings, biographies, and autobiographies relating to missions and reserves; and the Bringing Them Home Project interviews regarding the forced removal of Aboriginal children, or Stolen

Generation. Such data speak to a different type of migration in Australia: the forced internal displacement of Aboriginal people brought about by colonization. This underscores a fundamentally different view of migration, framed through a settler-colonial framework in which migration has negative consequences for Indigenous individuals and societies[2].

Some of the largest digitization projects have occurred through partnerships and collaborations. In 2012 the NAA announced a joint initiative with Ancestry.com to create an index and digitize records of the millions of people who arrived in Western Australia by sea or plane between 1897 and 1963 (Peters et al., 2017, p. 107–114). This followed a year after the launch of a website created through a partnership between the British Library and online publisher Brightsolid (operator of findmypast.co.uk, genesreunited.co.uk, and scotlandspeople.gov.uk) that transformed genealogical research through the digitization of more than 40 million pages of historical newspapers (BBC News, 2012). In addition to institutional alliances, partnerships have formed across nations. As part of their international cultural policy, the Netherlands has pioneered a Shared Cultural Heritage Programme, aimed at better understanding the ways that Dutch culture and history have been shaped by and influence the world around it. With the 10 partner countries of Australia, Brazil, India, Indonesia, Japan, Russia, Suriname, Sri Lanka, the United States, and South Africa, this programme encourages and provides funding to digitization projects that promote shared cultural heritage through principles of international cooperation, equality and respect between partners, and involvement of the public (Kingdom of the Netherlands, 2020). In Australia, this collaboration has resulted in such outputs as a video portrait series with Dutch migrants featured on the NAA's project website, "Destination Australia: Sharing Our Post-War Migrant Stories" (National Archives of Australia, 2022).

One of the greatest challenges in combining and creating large digital collections is the integration of disparate data sets. Data are often drawn from entirely different sources, formats, languages, and jurisdictions. Some of the most extensive work in integration has been carried out around Holocaust records. The Arolsen Archives, known until 2019 as the International Tracing Service (ITS), holds the largest collection of information on Nazi victims, including documents on displaced persons, forced labor, and concentration camps (Arolsen Archives, 2022). In 1948, the ITS began collecting and storing individual case files and other documents in an internationally coordinated effort to institutionalize records in the interests of survivors and prosecuting perpetrators (Rass and Tames, 2020, p. 23). It was also early to begin systematically digitizing records. Today the Arolsen Archives works with the European Holocaust

Research Infrastructure (EHRI) in focusing on the preparation and sharing of digital data with other archives. Along with institutions such as Yad Vashem; the United States Holocaust Memorial and Museum; the NIOD Institute for War, Holocaust, and Genocide Studies; and the Wiener Library, these groups are concerned with working together to enable widespread access to and digitization of Holocaust records (Rass and Tames, 2020, p. 24). In 2021, the EHRI also announced an initiative to drive collaborations with micro-archives (Arolsen Archives, 2022). Unconstrained by national archival laws, the Arolsen Archives has made a vast amount of its data available online (Rass and Tames, 2020, p. 24). Its data form the basis for the Transnational Remembrance of Nazi Forced Labor and Migration project (TransRem), which traces the wide variety of paths taken by those migrants now defined as displaced persons following World War II. TransRem follows the journeys of individuals from their places of birth, to sites of forced labor, to the towns they returned or emigrated to, spanning Europe, the United States, Australia, New Zealand, and South America. TransRem's website uses a combination of interactive maps that plot out quantitative geographical data, and story maps that feature more detailed narrative information and images. Such archives are exemplary models in integrating and opening access to data relating to DHMS.

Evolving tools of analysis

Digital Historical Migration Studies is also advancing through developments in data analysis. Data visualizations offer particularly striking examples of the potential for analyzing and presenting migration histories. Concerned with the movement of people over time, and numbers of people from various geographical and cultural backgrounds, DHMS lends itself to visual representations of journeys and mass movement. The plethora of interactive digital maps of migration, several of which are referenced in this paper, speaks to this. Researchers are also using more abstract representations. Cruz et al. (2018) at Northeastern University created a visualization entitled *Simulated Dendrochronology of Immigration to the United States 1830–2015*, which illustrates population growth through the symbolic imagery of an aging tree trunk, with each ring of the trunk representing a decade of growth and each cell representing 100 immigrants from a region. *The Global Flow of People* (Sander et al., 2014) illustrates estimates of migration between regions of Asia, Europe, Africa, and the Americas through interactive circular plots that become animated when users select certain regions. These examples demonstrate the potential in DHMS for analyzing and sharing migration data in particularly accessible and engaging formats.

Researchers are also using digital technologies to develop their own models and software for analysis. In their work investigating a genealogical database of over 800 million

---

2 For a recent discussion on the tensions around migration studies and the recognition of settler colonialism, see Ellerman and O'Heran (2021). See also Collins (2022).

names, Otterstrom and Bunker (2013, p. 544–569) developed a conceptual model to identify the connections between historical migration patterns and intergenerational family networks across three case studies in the United States. The authors demonstrate the relative pull of different cities in attracting migrants away from their generational hinterlands across different historical periods. Otterstrom and Bunker illustrate, for example, the incredible pull of small mining towns during the gold rush—on par with that of large cities—to attract faraway migrants. Others are developing corpus linguistics models. Viola and Verheul's (2020) innovative interdisciplinary work combines quantitative and qualitative methodologies to unearth patterns in historical public discourses around migrants and migration over time. Their analysis of the United Kingdom's *Times Digital Archive* between the years 1900 and 2000 merges a computational linguistics method with a discourse historical approach to look beyond whether a word has changed in meaning over time, and explore potential mechanisms involved in the construction of collective understandings (Viola and Verheul, 2020, p. 2). Their inquiry reveals the tendency for discussions around *emigration* to be framed through positive words such as "promoting" and "relief," while discussions around *immigration* are often framed through negative terms such as "exclusion," "undesired," and "restricting;" furthermore, over the course of 100 years, discourses around immigration have focused increasingly on ethnic minorities rather than on larger national groups (Viola and Verheul, 2020, p. 14). In another study, the authors examine a collection of Italian ethnic newspapers published in the United States between 1898 and 1920 (Viola and Verheul, 2019). Looking at the words of the migrants themselves, they trace the complex negotiations of identities and narratives by Italian migrant communities and individuals, balancing their attachment to homeland and heritage on the one hand, and immersion in their new society on the other (Viola and Verheul, 2019, p. 940). The authors argue that their findings challenge and transcend the traditional binary view of migrant "integration vs. isolation" by demonstrating how ethnic media helped migrants create "a united identity that could exert political force and negotiate inclusion" within the United States (Viola and Verheul, 2019, p. 935). This work demonstrates the potential for new, cross-disciplinary digital methods to analyse very large sets of data across long periods of time—a data-driven methodology for the longue durée approach in historical studies. It also highlights the potential of looking to long-neglected sources and underrepresented languages in historical and migration research.

Digital storytelling

New forms of migrant data are also being collected, analyzed, and presented through digital storytelling. This approach offers the field of DHMS more explicitly narrative-based qualitative material, centered on migrants' voices and subjective experiences. It was popularized in the 1990s by Joe Lambert, who founded the Center for Digital Storytelling in California, now known as StoryCenter (Lambert, 2013). Lambert developed a seven-step approach for individuals to create their own stories, from conception through production and sharing. While the term *digital storytelling* has since grown to include a broader range of methods and outputs, it generally refers to any projects that combine the art of storytelling with digital multimedia, including video, photographs, text, narration, and music (Trimboli, 2020, p. 5–6).

Digital storytelling has become a particularly common mode for recording contemporary migrant stories in Australia. The Australian Centre for the Moving Image (ACMI), based in Melbourne, Victoria, carried out a digital storytelling programme between 2002 and 2016, coproducing hundreds of stories with members of the Victorian and wider Australian community (Simondson, 2009). Many of these are now available on ACMI's YouTube channel. Australian Centre for the Moving Image identify particular themes in these stories, for instance, the experiences of Indigenous people, veterans, and young people (Australian Centre for the Moving Image, 2017). Migration also stands out as a theme. In her analysis of the collection, Daniella Trimboli estimates that approximately half of the stories are concerned with "culturally diverse" community narratives of migration and ethnic identity (Trimboli, 2020, p. 6). The NAA has also run a series of projects that record historical migrant stories through digital storytelling. "Destination Australia," a digital collection and website by the NAA, features video portraits of individuals who have migrated to Australia, as well as over 22,000 photographs taken by government photographers between 1946 and 1999 to record people arriving and living in Australia (National Archives of Australia, 2022). The NAA's 2014 exhibition *A Ticket to Paradise?*, which continues to appear today around Australia as a touring exhibition, combines Australian Bureau of Statistics archival data with personal testimony, photographs, and memorabilia to "tell the story of Australia's development through migration" (Gibson Group, 2019). Alongside object displays, thematic text panels, and large photographic displays, it features an interactive globe with five touch screens, where visitors can navigate an animated map showing yearly migration patterns to Australia, as well as launch a series of migrants' first-person stories. Significantly, *A Ticket to Paradise?* also allows audiences to contribute their own stories, either through facilitated workshops or a custom-built iPad application available in the physical exhibition space. "Destination Australia" also offers users the opportunity to contribute their own stories by uploading narrative text, images, and captions, and tagging locations and themes.

Independent researchers, artists, and creatives with lived experiences of migration are also developing their own

innovative and experimental expressions of digital storytelling. Matt Huynh, a Vietnamese Australian visual artist and storyteller, has created a series of interactive comics, illustrations, and animations detailing various experiences of migration in Australia and beyond. His 2019 work *Cabramatta* is an autobiographical interactive comic about growing up in a suburb that was both home to a community of Vietnam War refugees and the country's heroin capital (Huynh, 2019). *The Boat*, an interactive graphic novel released in 2015, details the story of Mai, a 16-year-old girl sent alone to Australia from Vietnam following the fall of Saigon (Huynh, 2015).

Agency, power, and bias

As the above examples demonstrate, the digital turn has opened up opportunities for migrants to record their own stories, in their own words. Moreover, the plural and fluid nature of digital content, ever-evolving and refreshing, lends itself to multifaceted representations of migration that illustrate the complexities of lived experiences and identities[3]. Social media and smartphones in particular have created a plethora of new discourse and material around migration. Georgiou and Leurs (2022, p. 668–689) argue that smartphones operate as "personal digital archives" where migrants can author and curate their own experiences and subjectivities, offering voice, agency, and autonomy in the face of external representations constructed by conventional Eurocentric media and migration research. These multifaceted digital forms offer modalities where migrants can tell their own stories against the grain of mainstream representations. Kaufmann cites the potential for the coproduction of smartphone data between researcher and migrants, allowing migrants to act as subjects rather than objects of study (Kaufmann, 2020).

Within DHMS, projects may directly or indirectly involve migrants. Roopika Risam describes two approaches: non-collaborative studies, which use existing data sets without collaborating with those who collected data or with migrants; and collaborative studies, which are produced by and with migrants (Risam, 2019, p. 571). In both forms of work, researchers must recognize and moderate their influence and bias when presenting migrant stories. Collaborative projects, though potentially empowering for migrants, can also involve dynamics in which researchers exert authority,

often unwittingly, over what can become their "subjects" of study. In her investigation of public engagement and participation in 12 museums and galleries across the UK, Lynch (2011) found that despite their best efforts, organizations collaborating with communities often had in place invisible barriers to genuine engagement. Lynch referred to recurring themes such as policies based on "helping-out," and community collaborators being "treated as 'beneficiaries' rather than 'active agents'" in projects and content development (Lynch, 2011, p. 20). Similarly, in *Museums, Heritage and Indigenous Voice: Decolonizing Engagement*, Onciul (2015) explores the growing practice of "community engagement" in museums through four case studies of heritage projects involving First Nations Blackfoot communities in southern Alberta, Canada. Onciul warns, and demonstrates through her case studies, that if done poorly, community engagement can in fact be disempowering to communities: "Engagement creates risks and costs for participants and is not necessarily as empowering or beneficial as current discourse often purports" (Onciul, 2015, p. 2).

As by its nature DHMS is concerned with the historical, often it may not be possible to facilitate collaborative studies, especially for those projects that focus on events deep in the past. The migrants of interest may no longer be alive, and groups or institutions that gathered the data sets may not be contactable. Again, this sets up power dynamics in which researchers hold authority over the analysis and presentation of data.

Indeed, the voices and agency of migrants contend with the power dynamics of the times in which they are recorded and retrieved. The experiences of individuals—especially the "ordinary" and even more so those deemed "stateless" and "other"—can be written out and distorted by powerful discourses of state archives and authorities. Researchers in DHMS are finding ways of unearthing these past migrant voices, for example Viola and Verheul examining the discourse of Italian migrants in ethnic newspapers (Viola and Verheul, 2019). Projects are also reading against the grain of "authoritative" data traditionally used to perpetuate dominant discourses. The Real Face of White Australia, created by Bagnall and Sherratt (2010), is one such project. Using facial detection script to analyse records from the NAA, this website uncovers the faces, names, and other biographical details of thousands of individuals who were deemed not "white" and hence forced by the state to carry documents allowing them to move across borders. Exhibiting their faces en masse, the Real Face of White Australia creates a striking display that challenges dominant and ongoing discourses around race and national identity with the lives and experiences of "othered" Australians.

These examples illustrate ways that researchers can work to interrogate and contest the underlying power structures and biases through which historical materials are produced. Sets of data are always formed through a series of choices and omissions that reflect the circumstances and politics of those

---

3 At the same time, individuals struggle against dominant discourses and frameworks. Trimboli argues that digital storytelling projects can embed multicultural subjects "in relations of power that both constrain and mobilize performances according to particular notions of whiteness" (Trimboli, 2020, p. 7). She points, for example, to representations that depict protagonists as "ethnic" as defined against an implied white audience, or implicitly underscore the national aspirations of a white Australia.

times in which they are made and accessed. Olaf Berg reminds us that data are never found but always created, through many moments of complex interpretation and decision-making (Berg, 2020, p. 263–289, cited in Rass and Tames, 2020, p. 33). Rass and Tames explain that, increasingly, researchers in migration studies and historical migration studies have shifted their attention to question terms that were once accepted as objective descriptors—such as "migrant" and "displaced person"—and examine the social processes that give rise to these political categories (Rass and Tames, 2020, p. 26). However, Rass and Tames warn that with the advent of big data we can lose sight of this critical approach; they caution "that we are not seduced by it and, as a result, view it uncritically" (Rass and Tames, 2020, p. 29). This lays out a tension brought about by digitization and digitalization—between ICT advances and what is often termed the "reflexive turn" in migration studies. Again the work of Viola and Verheul is of note here. In their examination of discourses around migration in the UK's *Times Digital Archive*, the authors interrogate the very concepts on which their work is premised. They do not simply trace patterns in migration in terms of how the movement of people changes over time, but critically examine the construction of concepts like "migration," "immigration," and "emigration." By interrogating the changing semantic and historical contexts around these concepts, Viola and Verheul unearth the potential processes through which societies construct their collective meanings.

In her comparative analysis of geospatial data visualizations of migration, Risam warns of the political consequences of producing visualizations of migration that fail to interrogate their own methodologies and political categories (Risam, 2019, p. 566). She focuses on two projects in particular, "The Flow Towards Europe" and "Crossing the Mediterranean Sea by Boat." "The Flow Towards Europe," created in 2018 by Finnish startup Lucify, uses data from the United Nations Refugee Agency to show the movement of migrants into European countries from 2012 to 2018. Risam explains how this visualization uses the term "refugee" as an uninterrogated catch-all, which obfuscates different migrant motivations, experiences, and the complexities of migration (Risam, 2019, p. 572). Further, she argues that other choices around language, visual rhetorics, and spatial modes play into political narratives that dehumanize and frame the migrant as the problematic "other"—for example the repeated use of the word "crisis," and the use of dots to represent individuals moving in "unimpeded waves" across national borders (Risam, 2019, p. 572). In contrast, Risam looks at "Crossing the Mediterranean Sea by Boat," a 2017 project by the University of Warwick, the University of Malta, and the Hellenic Foundation for European and Foreign Policy. Based on over 250 in-depth qualitative interviews conducted across 2015 and 2016, this project avoids simply using the preexisting category of "refugee" and explains that it features "people who have entered the European Union, or who are contemplating making the journey, by traveling across the Mediterranean Sea without authorization" (University of Warwick, 2019). The

project website traces out the experiences of individuals using interactive story maps, detailing reasons for migration such as escaping sexual violence or conscription, and displaying individual maps plotting out the unique phases of each journey rather than a single map showing thousands of nameless dots. Risam points out that such linguistic, visual, and spatial modes counter the "migrant-as-a-problem" narrative (Risam, 2019, p. 573–574).

Researchers can also practice self-reflexivity simply through acknowledging the partiality and subjectivity of their sources. The Transnational Remembrance of Nazi Forced Labor and Migration project (TransRem), based on data from the Arolsen Archives, traces the wide variety of paths taken by those migrants now defined as displaced persons following World War II. Before exploring the interactive map, users must first read a disclaimer that details the limitations of the sources on which the map is based, such as the standardization of dates, imprecise geographical data, and the potential for error in handwritten documents filled out by those applying for the status of "displaced person." The disclaimer states,

> No guarantee can be given that these data are complete or correct. Rather, the movements displayed are a visual representation of information which was provided by historical figures and recorded at a specific point in time in the historical sources on which the map is based. The people concerned may, of course, have provided false information deliberately or may have remembered things inaccurately. This does not reveal any underlying "methodological inaccuracy" of the project, but highlights instead the fundamental challenge posed by historical sources: they can never convey an objective picture of the past (Arolsen Archives, 2022).

Navigating ethical tensions

Researchers in DHMS thus face a range of questions and considerations around the ethics of working with the data of migrants—both indirectly and directly—and negotiating the power dynamics between researchers and participants, between states and individuals. Many careful decisions must be made regarding how data is obtained, contextualized, analyzed, and presented to audiences in a way that protects the data, agency, rights, and privacy of individuals. Marie Sandberg and Luca Rossi are particularly concerned with these ethical questions when dealing with ethnographic, qualitative, and "big social data" of migrants, who they describe as "subject to precarious and insecure life circumstances" (Sandberg and Rossi, 2022, p. 4). Drawing on Annemarie Mol's notion of care, the authors argue that working with big digital data requires the development of new models that approach migrants' digital data with care before—but also during and after—the data collection and research. They highlight, for instance, the challenges of

obtaining informed consent from individuals whose data has been taken from social media and public forums, and pose the question of how the use of migrants' data might feed back to and benefit migrant communities (Sandberg and Rossi, 2022, p. 3–4).

Current research is exploring and beginning to address these challenges. Reflecting on their work on PERCEPTIONS, a project identifying public narratives on Europe and migration to Europe by analyzing social media data, Mahoney et al. (2022) outline a number of methods used to navigate the ethical tensions of their work. For example, in dealing with the issue of informed consent, the authors made the decision to use only explicitly public material where authors have no expectations of privacy (hence using only Twitter data). The authors also approached the existing power dynamics and political categories around migration with care, avoiding "profiling" individuals by identifying and labeling them as migrants, and only identifying individuals with over 10,000 followers (Mahoney et al., 2022). Concerning the protection and privacy of data, they developed principles around data sharing between partners: only those project partners deemed able to provide adequate data protection (in line with General Data Protection Regulation) were responsible for data processing, and partners such as law enforcement were not given access to raw data (Mahoney et al., 2022, p. 232–234). Similarly, in their discussion around the tensions between pushes for open data and the ethics of migration research, Bloemraad and Menjívar (2022) suggest measures including the use of oral consent, confidentiality training for those involved in data collection, and the masking of personal and identifying information. They assert that "More broadly, researchers should double-down on data security, from data collection and storage to analysis and the communication of findings" (Bloemraad and Menjívar, 2022, p25–26). Bloemraad and Menjívar also recommend that rather than following generic ethical guidelines, researchers thoroughly, and critically assess the idiosyncrasies of their particular studies and engage in a three-step process of interrogating open-science and ethical principles, evaluating the "vulnerability" of migrants involved, and then forming methods appropriate to their projects.

Conclusion

Vast and growing digital data collections, tools, and theoretical frameworks have opened up the field of DHMS. Newly digitized migrant records, analytical models, and creative modes of digital storytelling offer diverse combinations of quantitative and qualitative materials, and opportunities for data produced by and with migrants. At the same time, these examples underscore the questions and challenges around how researchers navigate the structures of power underlying data, the biases of their research, and the ethical tensions inherent in working with human data. Future surveys of this field might expand the scope to a more diverse range of examples, particularly projects from non-European and Euro-Settler-Colonial societies, to unearth further opportunities and advances in DHMS and to explore further complexities to the tensions and challenges of the work. This work might help researchers in approaching questions and methods around the protection of migrants' data, voices, agency and security, and in integrating diverse voices and subjectivities when representing these complex histories and experiences.

Data availability statement

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.

Author contributions

Both authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Arolsen Archives. (2022). Available online at: https://arolsen-archives.org/en/search-explore/; https://transrem.arolsen-archives.org/maps/ (accessed March 2, 2022).

Australian Centre for the Moving Image. (2017). Available online at: https://www.acmi.net.au/stories-and-ideas/Collections-digital-storytelling/ (accessed March 2, 2022).

Bagnall, K., and Sherratt, T. (2010). *The Real Face of White Australia*. Available online at: https://www.realfaceofwhiteaustralia.net/ (accessed 14 June, 2022).

BBC News. (2012). *Family History Firm Brightsolid Launches in US*. Available online at: https://www.bbc.com/news/uk-scotland-scotland-business-16858504 (accessed March 2, 2022).

Berg, O. (2020). Capturing displaced persons' agency by modelling their life events: a mixed method digital humanities approach. *Hist. Soc. Res.* 45:4, 263–289. doi: 10.12759/hsr.45.2020.4.263-289

Bloemraad, I., and Menjívar, C. (2022). Precarious times, professional tensions: the ethics of migration research and the drive for scientific accountability. *Int. Migrat. Rev.* 56, 4–32. doi: 10.1177/01979183211014455

Brown, A. (2021). Bonegilla Migrant Experience to put more than 300,000 migrant identification cards online. *ABC*, 1 May 2021. Available online at: https://www.abc.net.au/news/2021-05-01/bonegilla-id-cards-records-digitised-australian-migrants/100106834 (accessed March 2, 2022).

Collins, F. L. (2022). Geographies of migration II: decolonising migration studies. *Progr Hum Geogr*. 866–877. doi: 10.1177/03091325221100826

Cruz, P., Wihbey, J., Ghael, A., and Shibuya, F. (2018). *Simulated Dendrochronology of Immigration to the United States 1830–2015*. Northeastern University. Available online at: https://web.northeastern.edu/naturalizing-immigration-dataviz/ (accessed March 2, 2022).

Ellerman, A., and O'Heran, B. (2021). "Unsettling migration studies: indigeneity and immigration in settler colonial states," in *Research Handbook on the Law and Politics of Migration*, ed C. Dauvergne (Cheltenham: Edward Elgar), p. 21–34. doi: 10.4337/9781789902266.00011

Georgiou, M., and Leurs, K. (2022). Smartphones as personal digital archives? Recentring migrant authority as curating and storytelling subjects. *Journalism* 23, 668–689. doi: 10.1177/14648849211060629

Gheorghiev, O. (2021). "Labour migration in the Czech Republic during COVID-19: between agency and vulnerabilities," in *International Migration Research Network Annual Conference* (Online 7–9 July, 2021).

Gibson Group. (2019). *A Ticket to Paradise?* Available online at: https://www.gibson.co.nz/ticket-to-paradise (accessed March 2, 2022).

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., and Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: opportunities, practical considerations, and challenges. *Perspect. Psychol. Sci.* 11, 838–854. doi: 10.1177/1745691616650285

Huynh, M. (2015). *The Boat*. Available online at: https://www.matthuynh.com/theboat (accessed March 2, 2022).

Huynh, M. (2019). *Cabramatta*. Available online at: https://www.matthuynh.com/cabramatta (accessed March 2, 2022).

Joseph, P., Gregg, M., and May, S. (2013). Digitisation of the WA Welcome Wall collection: a case study. *Arch. Manus.* 41, 183–202. doi: 10.1080/01576895.2013.829752

Kaufmann, K. (2020). "Mobile methods: doing migration research with the help of smartphones," in *The SAGE Handbook of Media and Migration*, eds K. Smets, K. Leurs, M. Georgiou, S. Witteborn, and R. Gajjala (London: SAGE), 167–179. doi: 10.4135/9781526476982.n22

Kingdom of the Netherlands. (2020). *Shared Cultural Heritage Fact Sheet 2017–2020*. Available online at: https://english.cultureelerfgoed.nl/publications/publications/2019/01/01/shared-cultural-heritage-fact-sheet-2017-2020 (accessed March 2, 2022).

Lambert, J. (2013). *Digital Storytelling: Capturing Lives, Creating Community*. New York, NY: Routledge. doi: 10.4324/9780203102329

Lynch, B. (2011). *Whose Cake is it Anyway? A Collaborative Investigation Into Engagement and Participation in Twelve Museums and Galleries in the UK, London*. London: Paul Hamlyn Foundation.

Mahoney, J., Le Louvier, K., Lawson, S., Bertel, D., and Ambrosetti, E. (2022). Ethical considerations in social media analytics in the context of migration: lessons learned from a Horizon 2020 project. *Res. Ethics* 18, 226–240. doi: 10.1177/17470161221087542

National Archives of Australia (2022). *Destination: Australia: Sharing our Post-War Migrant Stories*. Available online at: https://www.destinationaustralia.gov.au/ (accessed March 2, 2022).

Onciul, B. (2015). *Museums, Heritage and Indigenous Voice: Decolonizing Engagement*. New York, NY: Routledge. doi: 10.4324/9781315770246

Otterstrom, S. M., and Bunker, B. E. (2013). Genealogy, migration, and the intertwined geographies of personal pasts. *Ann. Assoc. Amer. Geogr.* 103:3, 544–569. doi: 10.1080/00045608.2012.700607

Palaiologou, N., Sounoglou, M., and Kyriazi, E. (2021). "I have a dream': Refugee children's aspirations from a child-centred perspective," in *International Migration Research Network Annual Conference* (Online 7–9, July 2021).

Peters, N., Marinova, D., van Faassen, M, and Stasiuk, G. (2017). "Digital preservation of cultural heritage," in *Technology, Society and Sustainability: Selected Concepts, Issues and Cases*, ed L. W. Zacher (Springer eBook), 107–114. doi: 10.1007/978-3-319-47164-8_7

Rass, C., and Tames, I. (2020). Negotiating the aftermath of forced migration: a view from the intersection of war and migration studies in the digital age. *Hist. Soc. Res.* 45, 7–44. doi: 10.12759/hsr.45.2020.4.7-44

Risam, R. (2019). Beyond the migrant 'problem': visualizing global migration. *Telev. New Media* 20, 566–580. doi: 10.1177/1527476419857679

Sandberg, M., and Rossi, L. (2022). "Caring for (big) data: an introduction to research methodologies and ethical challenges in digital migration studies," in *Research Methodologies and Ethical Challenges in Digital Migration Studies. Approaches to Social Inequality and Difference*, eds M. Sandberg, L. Rossi, V. Galis, and M. Bak Jørgensen (Cham: Palgrave Macmillan), 1–21. doi: 10.1007/978-3-030-81226-3

Sander, N., Abel, G. J., and Bauer, R. (2014). *The Global Flow of People*. Available online at: http://download.gsb.bund.de/BIB/global_flow/ (accessed March 2, 2022).

Simondson, H. (2009). *Connecting Through Digital Storytelling*. Available online at: https://www.cbaa.org.au/article/connecting-through-digital-storytelling (accessed March 2, 2022).

Tatarko, A., Jurcik, T., and Hadjar, A. (2021). "How migration policy shapes the subjective well-being of the non-immigrant population in European countries," in *International Migration Research Network Annual Conference* (Online, 7–9 July 2021). doi: 10.1177/00220221211001531

Trimboli, D. (2020). *Mediating Multiculturalism: Digital Storytelling and the Everyday Ethnic*. London: Anthem Press. doi: 10.2307/j.ctvsn3pqq

University of Warwick. (2019). *Crossing the Mediterranean Sea by boat*. Available online at: https://crossing-the-med-map.warwick.ac.uk/ (accessed March 2, 2022).

Viola, L., and Verheul, J. (2019). Mining ethnicity: discourse-driven topic modelling of immigrant discourses in the USA, 1898–1920. *Digit. Scholar. Human.* 35:4, 921–943. doi: 10.1093/llc/fqz068

Viola, L., and Verheul, J. (2020). One hundred years of migration discourse in *The Times*: a discourse-historical word vector space approach to the construction of meaning. *Front. Artif. Intell.* 3:64. doi: 10.3389/frai.2020.00064

Western Australian Museum. (2022). *Welcome Walls*. Available online at: https://museum.wa.gov.au/welcomewalls/ (accessed March 2, 2022).

Arrival declaration forms. A new gateway for mapping migration to Luxembourg

Machteld Venken* and Arnaud Sauer

Centre for Contemporary and Digital History, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Since the late 19th century, foreigners moving to Luxembourg have been required to declare their residency at the local municipality upon arrival. New digital technologies have made it possible to analyze the wealth of information contained in arrival declaration forms. This article offers a first digital analysis of a selection of these sources for the municipalities of Dudelange and Differdange in the mid-1920s. This was a pivotal period during which migratory flows to the Minett region, one of Europe's most dynamic industrial regions characterized by its iron mines and steelworks, were increasing and diversifying. Using a digital hermeneutic approach, the article evaluates the choices, opportunities and difficulties involved in using these sources within the virtual environment nodegoat. It presents insights gained by visualizing migration paths and settlement patterns: differences in mobility between Italian and German (un)married migrants, the case of the owner of a café and hostel (*café-pension*) in Dudelange who hosted Italian migrants from his place of birth, and the fact that social networks among family members and friends were also active on the road, with family members waiting for each other in the French transit town of Trieux. We end our article by identifying avenues for further research.

Introduction

Ernesto Leperoni was 35 years old when he arrived in Luxembourg in March 1924 and was registered by a civil servant under the Aliens Police Act of December 30, 1893, which required all foreigners to declare their arrival. This newcomer to Differdange was born in 1889 in Genga in central Italy, which provided a large contingent of workers for the steel and mining industry in the Minett iron region. In addition to civil status data, the declaration form includes detailed information on the declarant's places of residence over the previous 10 years. Ernest Leperoni had emigrated to the United States 10 years earlier. In 1912 he arrived in Minnesota, where he remained for 6 years. He then crossed the border and lived for 8 months in Winnipeg, Canada, before returning to the European continent, probably driven out by post-war unemployment.

He returned to his native village in 1919, where he married a young woman 3 years later. His vaccination certificate indicates that he was still in Genga in 1924. We then find him in Villerupt, a French town on the border with Luxembourg, where he stayed for 3 months before moving to Mont-Saint-Martin near Longwy, a French working-class town he also left after 3 months. Arriving in Differdange without his wife and three children, he stayed with Biondi, a fellow Italian living in Rue de l'Usine, who sublet him a room. He was hired by Hadir,[1] one of the largest integrated steelworks in the world, as a wheeler, handling loads of materials for fusion. Leperoni traveled with a fellow Italian born in the same town, who was registered in Differdange on the same day at the same address.[2]

When Hein de Haas, Stephen Castles and Mark J. Miller wrote that "France was the only Western European country to experience substantial immigration in the interwar years," they forgot about Luxembourg (2005, p. 92). Historical studies on migration to Luxembourg are few in number, despite the scale of the migratory phenomenon. This article offers a first digital analysis of arrival declaration forms, an important source revealing migrants' complex trajectories to and settlement patterns in Luxembourg. These forms show varied migratory paths, mobility practices and social networks. New digital technologies have made it possible to analyze the wealth of information contained in the vast collections of declaration forms archived in Luxembourg's former industrial towns in detail. Digital humanities tools facilitate a spatiotemporal analysis of the past by using different scales. Zooming in and out "between distant and close, macro and micro (...) becomes the norm" (Burdick et al., 2016, p. 30). "Space and time, brought together," (Zephyr, 2015, p. 416) argued, "become visible as co-present elements in history through the analysis of movement (...) visualizations of historical data can capture something of the sense of movement".

This article analyzes declaration forms completed by migrants when they arrived in one of two Luxembourgish municipalities, Dudelange or Differdange, in 1924. This was a pivotal year during which migratory flows to the Minett region, one of Europe's most dynamic industrial regions characterized by its iron mines and steelworks, were increasing and diversifying. The two municipalities lie on the French-Luxembourg border, not far from the center of the mining region in Esch-sur-Alzette. They were linked to other municipalities by an extensive cross-border public transport network. Both above and below ground, the area was crisscrossed by mining concessions and industrial complexes that extended beyond national borders.

We present the new insights gained by visualizing migration paths to and settlement patterns in Luxembourg and offer a critical methodological reflection on how these findings were gathered by means of a digital hermeneutic approach. The article starts out by evaluating the choices, difficulties and opportunities related to the sources and nodegoat software. It then presents an analysis of the data through geographical visualizations. We end our article by identifying avenues for further research.

Research question

This article presents new findings obtained from creating and interpreting geographic visualizations of migration paths leading to Luxembourg and settlement patterns in the country. More specifically, it addresses the following questions: 1. At what points and in what ways was geographical visualization a precondition for the interpretation of our data?, 2. What role did consecutive geographical visualization play in our analytical thinking?, and 3. How did geographical visualization support an analysis of migration to and between the two municipalities of Dudelange and Differdange?

The findings are presented using a digital hermeneutic approach consisting of (digital) source criticism, tool criticism and visualization criticism (Reisinger, 2010). Such an approach encourages reflection on the conditions of knowledge production for digitized data (Stiegler, 2014). Digital source criticism is used to analyze the quality of the data for digital display. Tool criticism involves reflecting on the possibilities of the software program nodegoat. Visualization criticism includes an assessment of other possibilities to present and interpret the spatiotemporal characteristics of migration to Luxembourg. The digital hermeneutic approach enables us to provide an evaluation of the difficulties in mapping space and time in the humanities. Our study should be seen as a first step toward a more tailored research project addressing specific research questions about migration to and within the Minett region through the analysis of larger data sets composed of different historical sources.

Historiography of the Minett region

Luxembourg entered the industrial era at the end of the 19th century following the invention of the Thomas-Gilchrist dephosphorization process for the exploitation of iron ore. The Minett region, which became the main center for iron ore mining in Europe and the second largest iron ore deposit in the world, covers a vast area from the south of Nancy in France to southern Luxembourg and south-east Belgium. A single working

---

1   The "Hauts-Fourneaux et Aciéries de Differdange, Saint-Ingbert, Rumelange," a French-Belgian company formed after the First World War from the assets of the German company "Deutsch-Luxemburgische Bergwerks- und Hütten-A.G. of Bochum" located in Luxembourg and Lorraine (France).

2   Communal archives of Differdange (further Differdange).

environment crossing national borders with no natural obstacles arose. Within a few decades, this industrial space dominated by the mono-industry of iron formed a dense network of steel and mining companies. Given the strong demand for industrial workers as well as the low availability of local workers who were employed in agriculture, industrial companies extended their recruitment policies. Foreign workers were invited from neighboring countries first, and from the Italian peninsula and young interwar states in Central and Eastern Europe afterwards (Klessmann, 1978; Hohengarten, 1995, p. 194–199; Ponty, 2005).

Our case study contributes to research in migration history, foregrounding a regional and transnational approach (Pisarevskaya et al., 2020, p. 457–458). Let us here take the example of Dudelange, one of the two municipalities under study, which grew from a village into a thriving town in response to the frantic pace of development of industrial production and the influx of foreign workers. Dudelange saw the birth of a district referred to as "little Italy" near its steelworks (Rainhorn, 2005; Blanc-Chaléard et al., 2007). The demographic explosion required industrial plants and the state to develop an active housing policy, but these initiatives could not mask the poor housing conditions in overcrowded dwellings (Lehners, 1991, p. 9–12). Political upheavals, social and trade union movements and economic crises subsequently led to an outflux of migrants (Galloro, 2001, p. 93–126).

Although the historiography of the steel company ARBED offers a detailed picture of Luxembourg's past industrial strategy in comparison to the rest of Europe (Chomé, 1964), the history of other Luxembourgish mining and steel companies remains in the shadows (Leboutte et al., 1998, p. 161–188). On the other side of the border, in Lorraine, more detailed studies focus on the evolution of the (foreign) labor market (Walter, 1935; Prêcheur, 1959). A history of migration to Luxembourg in the interwar years still needs to be written. Figures show that whereas in 1913, 60% of the workers employed in the steel and mining industry were foreigners, i.e., 11,000 (mainly Italians and Germans) out of a total of 19,000 workers, by 1930, 55,831 foreigners were registered in Luxembourg. The workforce of the plant in Dudelange reached 1,799 foreigners in 1927 out of a total workforce of 4,105 people. These figures hide the turnover; between 1919 and 1939, there were 13,503 declarations of arrival of foreigners in Dudelange and 22,349 in Differdange (Scuto, 2010, p. 13–38).

What do we know about the profile of foreign workers in Luxembourg? The priest Gallo (1987) indicated that many of them came from various places within Italy at different speeds and at different times. The historian Scuto (2010, p. 12–38) noted that most Italian migrants arriving in Luxembourg were from northern and central Italy, a finding confirmed for Dudelange by Trinkaus (1984, p. 151). More detailed studies on the characteristics of foreign workers are available for those

who settled across the border in France. The French sociologist Serge Bonnet and the historian Gérard Noiriel conducted studies in Lorraine documenting the influence of foreigners on the development of trade unions (Bonnet, 1976; Noiriel, 2019). Moreover, the geographer Sömme (1930) showed the high mobility of different foreign workers crossing national borders within the Minett region. More recently, the sociologist Galloro (2001) analyzed the role of industrial companies and the government in the process of recruiting and dismissing foreign labor in Lorraine, as well as the strategies of foreign workers to circumvent the various restrictions associated with their status.

The Second Industrial Revolution led to a strong influx of new populations of foreign workers in Luxembourg. With the rise of nation-states and the increase in tensions between neighboring nations following the Franco-Prussian War of 1870, border control became a major concern (Noiriel, 2001, p. 125–144; Scuto, 2012, p. 155–156). To exert more control over this highly mobile foreign population, the Luxembourg government decided to legislate, like the governments in other countries. An Act adopted on December 30, 1893 required foreigners to declare their arrival in Luxembourg.[3] Control was further stepped up by a decree adopted on May 31, 1934, which introduced special identity cards for foreigners.[4]

Heavy industry slowly restarted to recruit foreign workers in the aftermath of the First World War and experienced rapid growth from 1922 onwards (Trausch, 2017, p. 185–198). The foreign population entering the municipality of Dudelange rose from 446 in 1923 to 1,115 in 1924. By 1923, various laws had been passed to restrict the influx of foreign workers. The decree of August 21, 1923 required all employers to apply for an authorization from at the Directorate General of Agriculture, Industry and Social Welfare indicating that a given vacancy could not be occupied by a Luxembourg national. In addition, the Act of October 28, 1920 aimed to protect the national population against competition from foreign workers suspected of social dumping and wage stagnation by stemming the influx of foreigners.[5]

Before discussing the geographic visualizations, the article will present and evaluate the sources and the digital tool used, offering an overview of how we used declaration forms as sources and reflecting on our initiatives to expand the source base.

---

3   This first legislative instrument was amended several times, such as in 1911, by a Decree regarding the implementation of the Act.

4   The text was modified in July of the same year and then in 1935, 1937, 1938, 1952, 1958 and 1964.

5   Legilux: https://legilux.public.lu/eli/etat/leg/loi/1920/10/28/n1/jo.

Sources and digital source criticism

Our research is based on one-page typed arrival declaration forms meticulously completed by civil servants. They contain spatiotemporal information reconstructing the life and journey of the declarant, including birth, marriage, places of residence before and upon arrival in Luxembourg and employment. A first overview of the data shows that of the 1,115 declarants in Dudelange, there were 663 Italians and 452 Germans (including 10 migrants whose regional origins from the Territory of the Saar Basin were indicated specifically, as it functioned under a League of Nations mandate) (Figure 1). Other nationalities were less well represented, including 18 French nationals and 20 individuals from Alsace and Lorraine, a regional particularity that reflected the fact that citizenship issues were still being negotiated after the territory had changed from German to French sovereignty following the First World War (Zind, 1979, p. 128–129). Other nationalities with 10 or more migrants included Polish (some of whom were referred to as "Russ Polen" given their former residence in the Russian Empire) (15), Belgian (12) and Austrian (understood as formerly belonging to the Austro-Hungarian empire, although Czech nationality was indicated separately on one occasion) (10). Individuals migrated from Spain (4), what was called Russia (despite the establishment of the Soviet Union in 1917) (3), Switzerland (2), the United States (2) and the Netherlands (1), among others. The fact that 980 declarations were made by men and 245 by women is explained by the fact that married couples who arrived together were declared through the husband. We counted 107 married and 138 unmarried female declarants. In general, these married women traveled alone (though often with their children) to join their husbands. In 1924, 21 married husbands arrived in Dudelange by themselves and were joined by their spouses within the same year and one woman arrived 2 months before her husband.[6] As most migrants needed some time to find and be able to afford suitable living conditions for a family, the spouses often remained separated for a longer period.

Although this article, like many studies in migration history, follows the categorizations imposed by policy makers (Schrover and Moloney, 2013, p. 9), it nevertheless critically assesses the quality of the data by asking how civil servants completed their task in practice. Data in the original sources were sometimes inaccurate. The places of birth and marriage of Russian emigrants in Cyrillic script were not correctly transcribed by Luxembourg officials. The former Russian prisoner of war Theodor Wolkow, for example, declared "Gribatschi" as his place of birth, which we were unable to geolocate with the help of the GeoNames portal.[7] A second difficulty was imprecise dating as a result of descriptions such as "for 2 and a half years," "3 weeks" or even more vague wordings such as "since the war." Information included in the field "Residence during

the last 10 years" could offer a succession of estimated periods, forcing the researcher to make approximate calculations or ultimately to accept a certain margin of error. Luigi Bison, for example, arrived in Dudelange in December 1924 and declared eleven successive previous places of residence, all of approximate duration. He lived for 6 months in Creola, 2 years in Moyeuvre-Grande, 4 months in Belgium and 1 year in Creola. Previously, he had spent 9 months in Moyeuvre-Grande, 6 months in Creola and 5 years in the Italian Army. Before the war, he mentioned having lived for 6 months in Duisburg in Germany, 1 year in Oberhausen and 9 months in Dortmund, and in Creola before that (Figure 2).[8] The historian Andrei Petruș stated that "Nodegoat does not currently include an option to flag fictional data" (2020, 35), but we included a separate field next to a date in our data model, where we indicated whether that date was exact or calculated by the researcher, in order to offer as accurate an interpretation as possible of our geographical visualizations. Although temporal descriptions such as "military service" could be approximately equated to the duration of the war, if there was a question mark in the source, precise dating was impossible.[9]

Border changes after the First World War caused other difficulties in data processing. Luxembourg civil servants regularly referred to an era that was no longer applicable at the time of declaration; the French in Alsace-Lorraine, for example, were frequently documented with their places of birth, marriage and residence using the German forms of the place names (e.g., Diedenhofen for Thionville). Under the Treaty of Versailles, French nationality was not granted automatically to certain categories of people from Alsace-Lorraine. This difference in status between inhabitants of the annexed territory was shown in practice by the establishment of identity cards of different types depending on the person's origin (Zind, 1979, p. 128–129).

In the same way, imprecise information on the places of origin of German or Polish emigrants led to confusion, as in the case of municipalities located in a border area which may have had different names depending on the national origin of the declarant or on shifting borders.[10] A German migrant, for example, indicated that he came from Tucheln, a town that changed its name to Tuchola when it became Polish after the First World War.[11] We used the GeoNames portal to manually trace the main geolocalizations, but for small villages that often merged into bigger administrative entities after 1924, the task remained difficult and sometimes even impossible.[12]

---

6   DA, Dudelange, 1924, 683 and 814.

7   DA, Dudelange, 1924, 786.

---

8   DA, Dudelange, 1924, 1068.

9   DA, Dudelange, 1924, 237, 909.

10   DA, Dudelange, 1924, 347, 784.

11   DA, Dudelange, 1924, 667.

12   When Emanuel Saternus arrived in Dudelange in May 1924, this Polish emigrant was registered as a native of Königshütte, a Prussian town in Silesia. When Königshütte became Polish in 1921, it was renamed Chorzów. His wife was from Neuderbor, Pless district, which we were unable to geolocate (DA, Dudelange, 1924, 347).

FIGURE 1
Places of residence of declarants in Dudelange according to citizenship.

Toward an expansion of the digital source base

As Faassen and Hoekstra note, "the challenges in exploring digitized collections with new methodologies reinforces the tendency to concentrate on a single collection. However, this contrasts with how historians usually work" (Faassen and Hoekstra, 2022). In an attempt to respond to this critique, we extended our digital source base by including data from cadastral sources and declaration forms for the town of Differdange.

Because no historical maps with street names are available for Dudelange for the period between 1890 and 1972, the

Résidence pendant les 10 dernières années	Kommt von Creola / wo er 6 Monate war / vorher 2 Jahre zu Gross-Moyeuvre / vorher 4 Monate zu Belgien / vorher 1 Jahr zu Creola / vorher 9 Monate zu Gr. Moyeuvre / vorher zu Creola / vorher 5 Jahre italienischer Soldat / vorher 6 Monate zu Duisburg / Alt Markt 2 / vorher 1 Jahr zu Oberhausen / Wiesenstr. No.2 / vorher zu Dortmund 9 Monate / vorher seit Geburt zu Creola

FIGURE 2
The transcription of a fragment of the data from the original declaration form of Luigi Bison in our database under the object "Form[a]."
[a]DA, *Dudelange*, 1924, 1068.

FIGURE 3
Migrants also tended to live in closer proximity to the steelworks.

maps we initially created on the basis of old street names and the old form of numbering mentioned in the arrival declaration forms included significant inaccuracies regarding the settlement pattern of migrants. A map of migrants who declared their residence as "Hüttenstrasse" (also known as Rue des Usines) in Dudelange, for example, displays these migrants on the equivalent street in the post-war period, when it became known as Rue de la Libération. Migrants were geolocated on the northern part of the street, although old pictures show that employees also lived on the southern part of the street which was closer to the steelworks.[13]

By consulting cadastral sources, we could include historical addresses of residence. Cadastral sources registered the sale

---

13 Postcard showing the southern part of Hüttenstrasse/Rue des Usines leading toward the steelworks, including the café run by Joseph Hirschberger, born in 1870 in Luxembourg as the son of a Jewish bartender originally from Gerolzhofen in Bavaria, Germany (Municipal Archives of Dudelange, early 20th century).

and purchase of land by means of unique land plot numbers. Although it was not common practice to document the street name and number, these were sometimes mentioned, for example in the case of the house at 225 Rue des Usines, which was sold on October 20, 1932 after the death of Boban François, a former steel worker in Dudelange. The buyer was Catherine Fritsch, the divorced wife of Charles Reisen. The address 225 Rue des Usines/Hüttenstrasse corresponds to plot number C-2992/2755.[14]

By consulting a contemporary Luxembourg government database on geospatial data, we were able to link the land plot numbers to their current streets and street numbers.[15] The address 225 Rue des Usines/Hüttenstrasse from before the Second World War corresponds to the contemporary address 205 Rue de la Libération. After adapting our data model and integrating the cadastral data into the database, we were able to map streets and street numbers more accurately. The corrected map shows that migrants also tended to live in closer proximity to the steelworks. It shows that we had not taken into consideration the fact that Rue de la Libération merged the two prewar streets Hüttenstrasse (in the south) and Deichstrasse (in the north) after the Second World War; (Figure 3).[16]

We also compared our data with the data included in declaration forms issued by the nearby town of Differdange. In 1924, the number of foreigners arriving in Differdange was twice that of Dudelange (2,442 compared with 1,115). In the first 3 months, there were 412 declarations, including a high proportion of Germans (204) but also Italians (139), French (16) and Belgians (15). The steelworks was managed until 1920 by a German company, the *Deutsch-Luxemburgische Bergwerks- und Hütten-A.G.* The end of the war and the change of ownership of the company—in 1920, the plant was taken over by a French-Luxembourgish industrial consortium (Hadir)—does not seem to have disrupted the structure of the foreign workforce. Despite anti-German campaigns published in the French-speaking press, there continued to be a high number of German workers.[17] The majority of German declarants came from neighboring regions and had close professional links with metallurgical and mining regions in the Ruhr and Saarland.[18]

---

14 Source: Cadastre de Dudelange, Relevé des Actes de Mutations contenant tous les actifs de Propriétés foncières entre le 1 janvier 1932 et le 31 décembre 1940.

15 www.geoportail.lu, consulted on April 25, 2022.

16 See also: Stadt Düdelingen. Strassenbenennungen, 1956 for imprecise descriptions of changes in street locations.

17 See: "Un scandale Boche," *L'Indépendance Luxembourgeoise*, 8 octobre 1919, no. 233, p. 1.

18 DA, Dudelange, 1924, 319.

Our nodegoat data model and tool criticism

The nodegoat platform is a digital environment for the creation of data models and databases, network analysis and the online visualization of dynamic historical data using cartographic support. The software uses iterative data modeling, allowing researchers to adapt and further develop their data model while inserting and interpreting data (van Bree and Kessel, 2017, p. 2). We used nodegoat as a tool to detect faintly perceptible phenomena and to develop new hypotheses that are difficult to explore by other means of analysis. Our intention was to enrich our perception of migration to and within Luxembourg by interpreting these data through exploration. As our data model developed, the possibility of cross-referencing data multiplied, as did the options for experimenting with the data.

We developed an input mask that was sufficiently detailed and nuanced to respond to the characteristics and subtleties of our sources and allow for multiple interpretative activities (Burrows, 2017, p. 45). Our data model introduced three new features that have yet to be mentioned in publications by scholars using nodegoat: scalable reading, a collaborative learning environment for university students, and geolocalization enriched with cadastral source data (Burrows; Palomba et al., 2020; Petruş, 2020). In order to document our interpretation of spatiotemporal data, we developed a data model to facilitate scalable reading, allowing a combination of distant reading (through overview visualizations) and close reading (the interpretation of single items of data) by linking data to scanned source materials for all the interpretation phases of data insertion (Weitin, 2017). Our data model included two levels of data processing. The first step created an object entitled "Form," allowing the information to be transcribed as it had been in the source, respecting the structure of the form and adding information such as the status of the source (e.g., a torn document). Since the declaration forms in different municipalities documented information differently, for reasons of coherence we included fields indicating whether information was asked for or not. The second step created a new object called "Person," where information about declarants and each of their family members was prepared for mapping by interpreting time (including estimations) and geolocalization (including self-created geometries estimating space) (Figure 4). Particular attention was paid to making wives more visible in the database than they were in the original source material. Wives who arrived with their husbands in Dudelange were mentioned on the declaration forms of their husbands. In our database, however, they are included as individuals who are linked through their marriage with their husbands. We also made distinctions between missing, incomplete, approximate or illegible data. Through the "Person" object, it was possible to
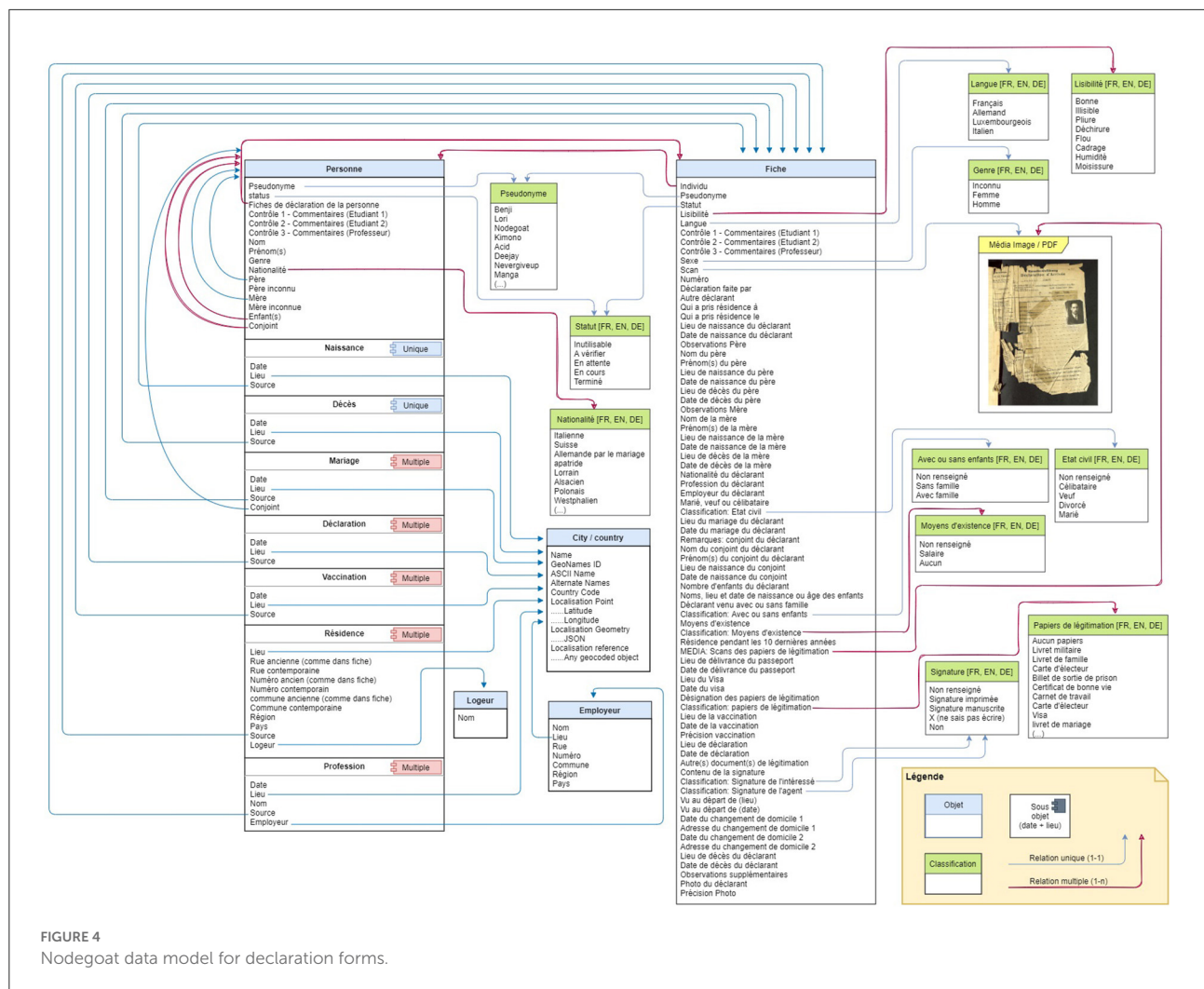
FIGURE 4
Nodegoat data model for declaration forms.

FIGURE 5
The color code used in our database facilitates data insertion for students.

link different forms with information about the same individual, for example a foreigner moving between Dudelange and Differdange,[19] or to link several individuals, such as brothers and sisters.[20]

───────

19 DA, Differdange, 1924, 362; DA, Dudelange, 1924, 667.

20 DA, Dudelange, 1924, 563, 564.

We designed a database in which students could insert data manually and the entire class could interpret data collaboratively interpretation. Two different groups of students received individual pseudonyms based on their initials in order to facilitate recognizability within each of the groups, while at the same time guaranteeing individual students' pseudonymity outside of that group. The pseudonyms were

Adresse géolocalisée: Rue Saint Martin Rue Saint Martin Rue Martin Martinusstrasse Dudelange (LU) 16 28
(ngZO1r17hYunhYuQqY3h3oEr)

Pseudonyme	🔍 Silverware	ver
Statut	🔍 A vérifier	ver
Commentaire	Silverware: according to Excellsheet Martinusstrasse 15 is now 29. I therefore picked the opposite side of the road, as 16 is not in the list. STOP	ver
rue ancienne (comme dans fiche)	Martinusstrasse	ver
LIEN: Rue contemporaine	🔍 Rue Saint Martin Rue Saint Martin Rue Martin Martinusstras	ver
Numéro (donnée renseignée)	🔍 Donnée renseignée + réponse approximative	ver
Numéro ancien (comme dans fiche)	16	ver
Numéro contemporain	28	ver
Lien internet geoportal.lu [1]	<iframe src='https://map.geoportail.lu/theme/main?version=3&z	ver
Lien internet geoportal.lu [2]	<iframe src='https://map.geoportail.lu/theme/main?version=3&z	ver

Sub-Objects: Editor   [Located]

FIGURE 6
Student interpretation of a geolocalization.

FIGURE 7
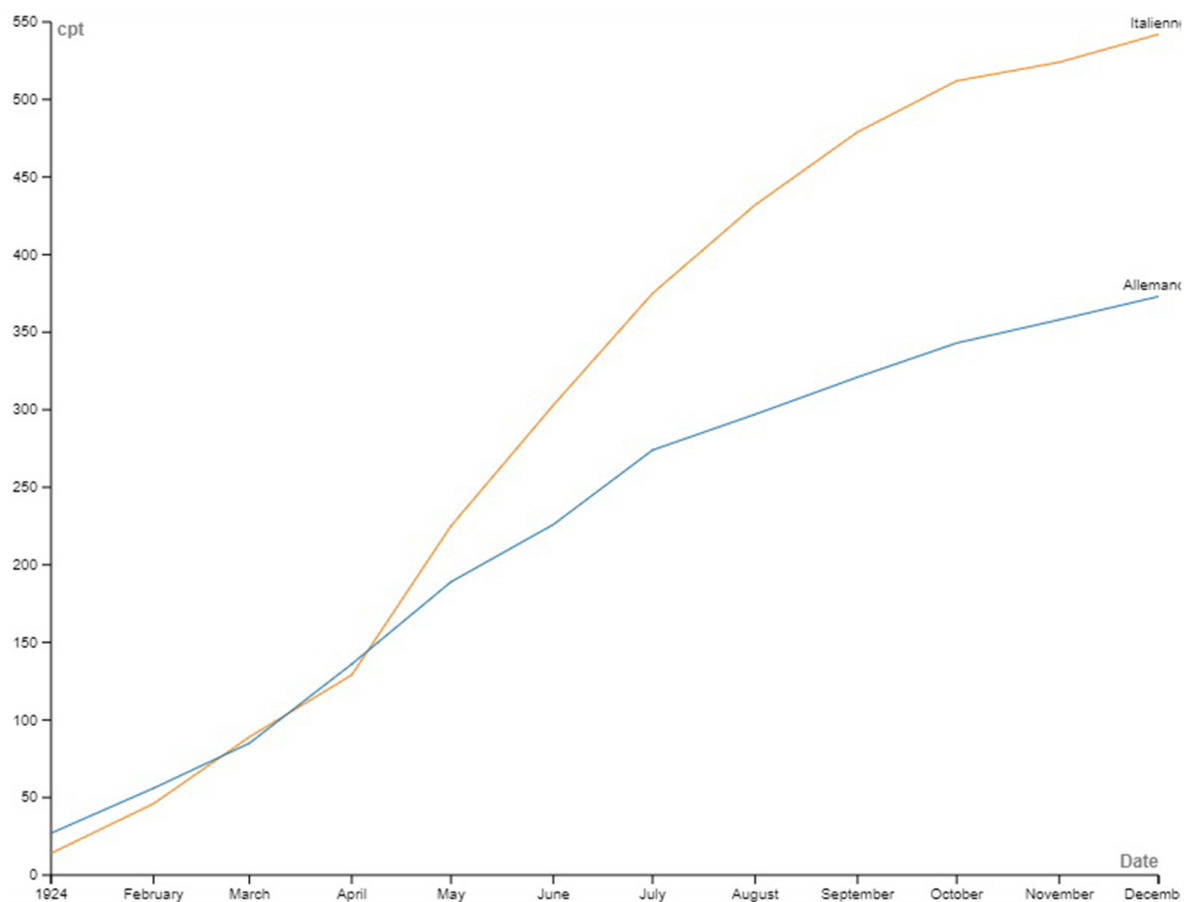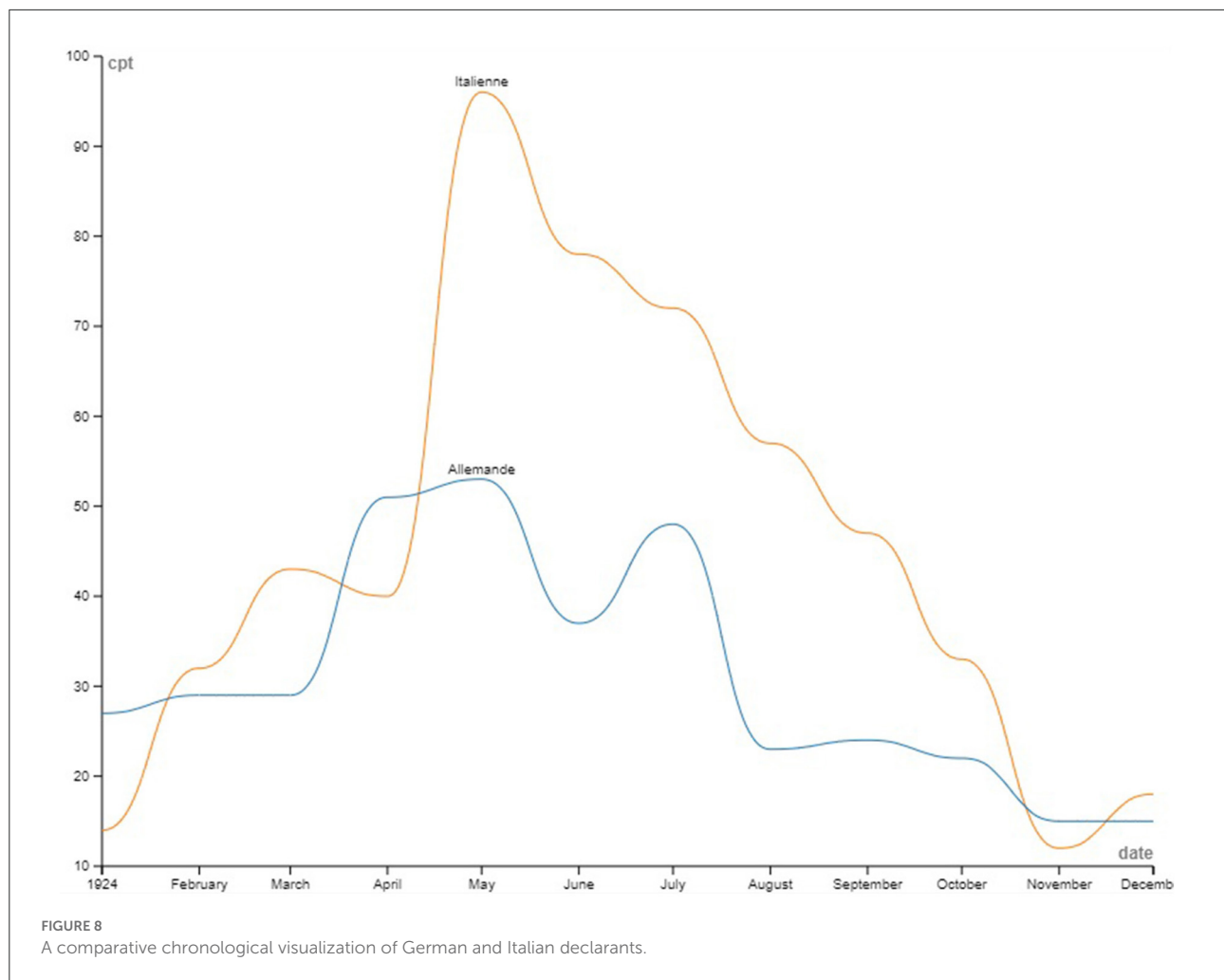A cumulative chronological visualization of German and Italian declarants.

FIGURE 8
A comparative chronological visualization of German and Italian declarants.

decided upon at the beginning of the project when we did not yet know that our work would eventually lead to a research publication. We recognize that a pseudonym such as Silverware objectifies a participant in the project and may not enable readers of this article to identify with the human being who inserted and interpreted the data. An explanatory manual was produced for students to understand and comment on the data entry process. The database included fields for comments and questions. We also created a color code for forms with missing or incorrectly entered data, indicating in red the records requiring verification (Figure 5). After peer control among students, the teacher checked the students' work and approved the data insertion process.

A last special feature of our data model was a more developed system for historic geolocalization by linking the location data in the forms with the Dudelange land registry (*cadastre*). We created a concordance table in Excel to link old and contemporary addresses of the same cadastral entity within nodegoat (Figure 6).

When we sought to compare the rising numbers of arriving German and Italian migrants in 1924, we discovered that nodegoat's chronological visualizations were informative for indicating how many declarants holding one citizenship arrived in Dudelange on a specific day, but could not offer cumulative or comparative chronological visualizations. When we imported our data in a CSV format in the statistical program RawGraphs 2.0 and composed a cumulative distribution plot, we discovered that German declarants outweighed Italian declarants until May, after which point there were more Italian declarants (Figure 7). This may be explained by the implementation of the American Immigration Act aimed at preventing immigration from various countries, including Italy, which was signed into law on May 26, 1924. Whereas in the years 1900–1910, approximately 200,000 Italians had migrated to the United States, in 1924 their number decreased to 4,000 (Shaw-Taylor and McCall, 2020, p. 63). A comparative chronological visualization, however, indicates that the Italian influx was more seasonal than German migration and that the higher number of Italian declarants in May could also be interpreted as the beginning of the high season (Figure 8).

Geographic visualization and criticism

We will now present the findings we obtained by creating and interpreting geographic visualizations in nodegoat. An accepted definition of visualization is "a method of deforming, compressing, or otherwise manipulating data in order to see them in new and enlightening ways" (Graham et al., 2016, p. 159). Mapping is understood as visualizing geographically: "Mapping constructs while it objectifies the world it represents, selectively, therefore shaping thought and guiding action" (McCarty, 2005, p. 33). Scholars have argued that explorative geographic visualizations can help shed light on interconnections between series of data: "even before a dataset is complete, visualizations can be used to recognize errors in the data collection process" (Graham et al., 2016, p. 163). In this way, exploratory visualizations become an integral element of the research workflow, sitting "particularly well as an additional layer in the hermeneutic process of hypothesis formation" (Graham et al., 2016, p. 163; see also: Burrows, 2017, p. 51). Mapping is used to "test hypotheses, discover patterns, and investigate historical processes and relationships" (Presner and Shepard, 2015, p. 209). Maps are no longer illustrations added to an analytical text, or "displays of the results of an analysis," but are appreciated as "primary modes of knowledge production" (Drucker, 2020, p. 17, 120), since they have the potential to "generate questions that might otherwise go unasked, reveal historical relations that might otherwise go unnoticed, and undermine, or substantiate, stories upon which we build our own versions of the past" (Presner and Shepard, 2015, p. 209). Cross-referenced data on mobility enabled us to identify phenomena and trends that would otherwise be very difficult to perceive.

A precondition for analysis

Geographic visualization could function as a precondition for analysis, as a comparison of the mobility paths of unmarried and married Italian and German declarants in Dudelange in 1924 makes clear. Although the link between high mobility among unmarried migrants and a less varied migration path for married migrants has already been documented for migrants settling outside Luxembourg, it has not been documented in historiographical literature on migration to Luxembourg (Williams, 2013). Detailing the specific nature of that relationship required a geographical visualization; given the multitude of different geographical names and time periods related to the mobility and settlement of the many migrants arriving in Dudelange in 1924, it would have been impossible to come to any conclusions without mapping. The four figures map the migration patterns of Italian and German migrants. More

FIGURE 9
Married German and Italian migrants arriving in Dudelange in May 1924.

FIGURE 10
Unmarried German and Italian migrants arriving in Dudelange in May 1924.
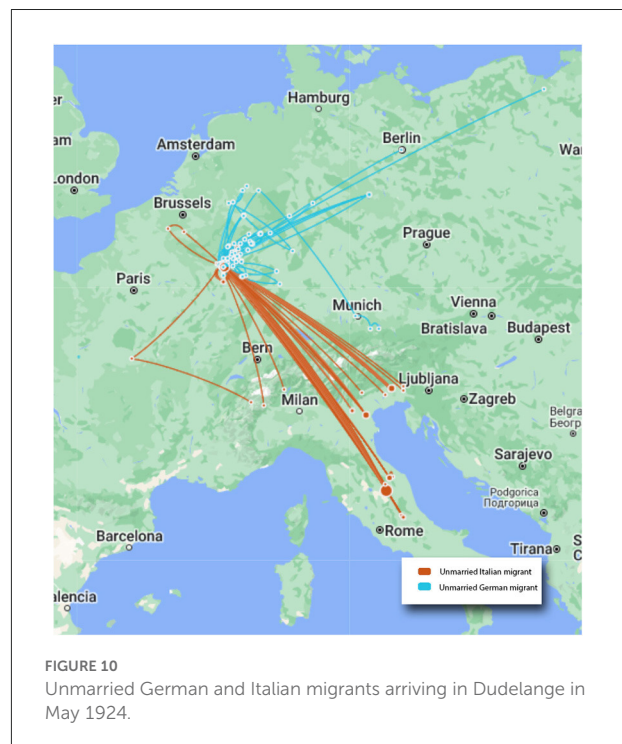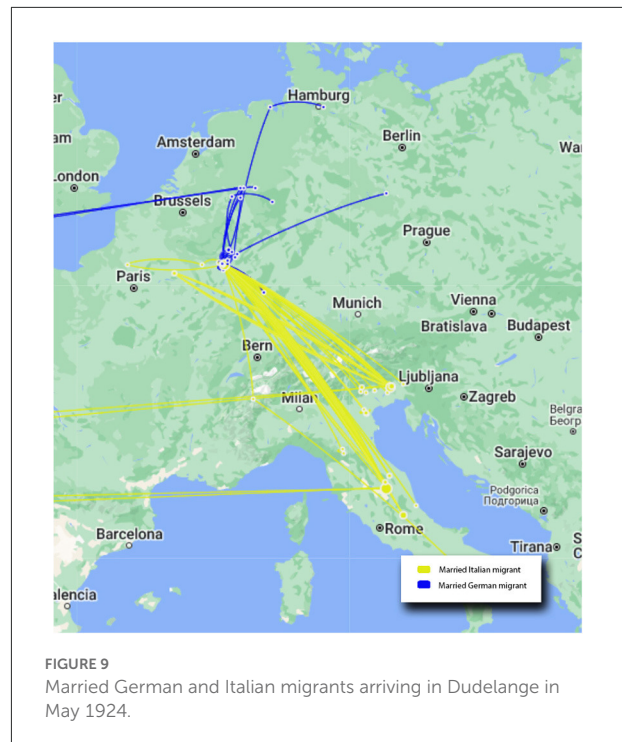
specifically, we selected unmarried and married migrants who arrived in Dudelange either in May, the month for which there were the most declaration forms, or in November, the month in 1924 during which the fewest migrants arrived. In May, 47 married and 50 unmarried Italian migrants and 10 married
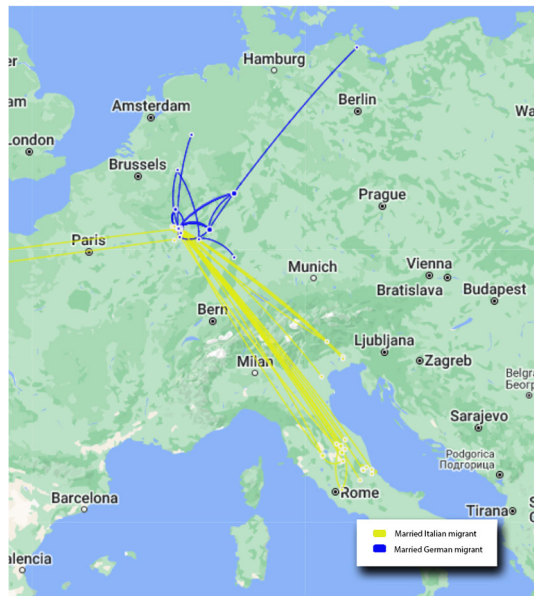
FIGURE 11
Married German and Italian migrants arriving in Dudelange in
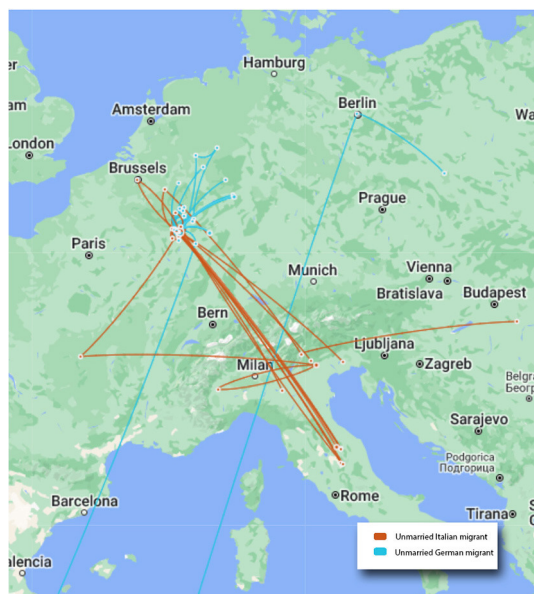November 1924.

FIGURE 12
Unmarried German and Italian migrants arriving in Dudelange in
November 1924.

and 44 unmarried German migrants arrived in Dudelange
(Figures 9, 10). November saw the arrival of 14 married and
13 unmarried Italian migrants, as well as 11 married and 19
unmarried German migrants (Figures 11, 12). The visualizations
show the migration trajectory of migrants up to 10 years before
1924, as documented in their declaration forms.

A comparison of the visualizations for the month of
November reveals the mostly direct migration trajectories
between the places of residence in Germany or Italy and
Dudelange for married migrants, compared to trajectories
involving more places of residence within the Greater Region
for unmarried German and Italian migrants. The maps show
how migrants gave meaning to the Minett region as a borderland
by means of their mobility practices. The borderland indeed
became "a region jointly shared by two nations that houses
people with common social characteristics in spite of the
political boundary between them" (McKinsey and Konrad, 1989,
p. 10).

A comparison between married German migrants and
unmarried German migrants arriving in Dudelange in May and
November indicates that there were different recruitment areas
for the two groups. Whereas married German migrants were
almost exclusively recruited from areas north of Luxembourg,
unmarried German migrants had a greater chance of being
recruited from across Luxembourg's eastern border. Comparing
the migration patterns of unmarried Italian migrants between
May and November shows that unmarried Italian migrants
arriving in May had less intense migration trajectories within
a perimeter of 200 kilometers than both their German
counterparts in May and unmarried Italian migrants arriving
in November. This demonstrates that another type of Italian
migrant was recruited for seasonal work in the summer, when
more migrants were recruited directly from Italy, than for work
in the colder months of the year, when recruitments were
made among those who were already in the neighborhood or
had already worked there. As the maps only show arriving
migrants, we do not know how their situation evolved. There are
indications that mixed marriages were a common phenomenon
among Italian, German, French and Luxembourgish citizens in
the Minett region, but the relationship between marriage and
mobility remains under-researched for Luxembourg (Reuter,
2018, p. 36).

Consecutive mapping supporting the thinking process

We often treated our self-created geographic visualizations
as summaries of selected data inviting us to a deeper analysis by
means of consecutive mapping (Hayles, 2012, p. 47). A first map
created in nodegoat, for example, displayed declarants' places of
residence in café hostels (*café-pensions*) (Figure 13).

We detected a relationship between certain Italian migrants'
places of birth and their places of residence in Dudelange.
Among the most popular destinations were the Ruffini *café-
pension*, which hosted 57 declarants, and the Baldelli *café-
pension*, which hosted 11 declarants. These two *café-pensions*
respectively housed 19 and 9 Italian migrants born in Nocera
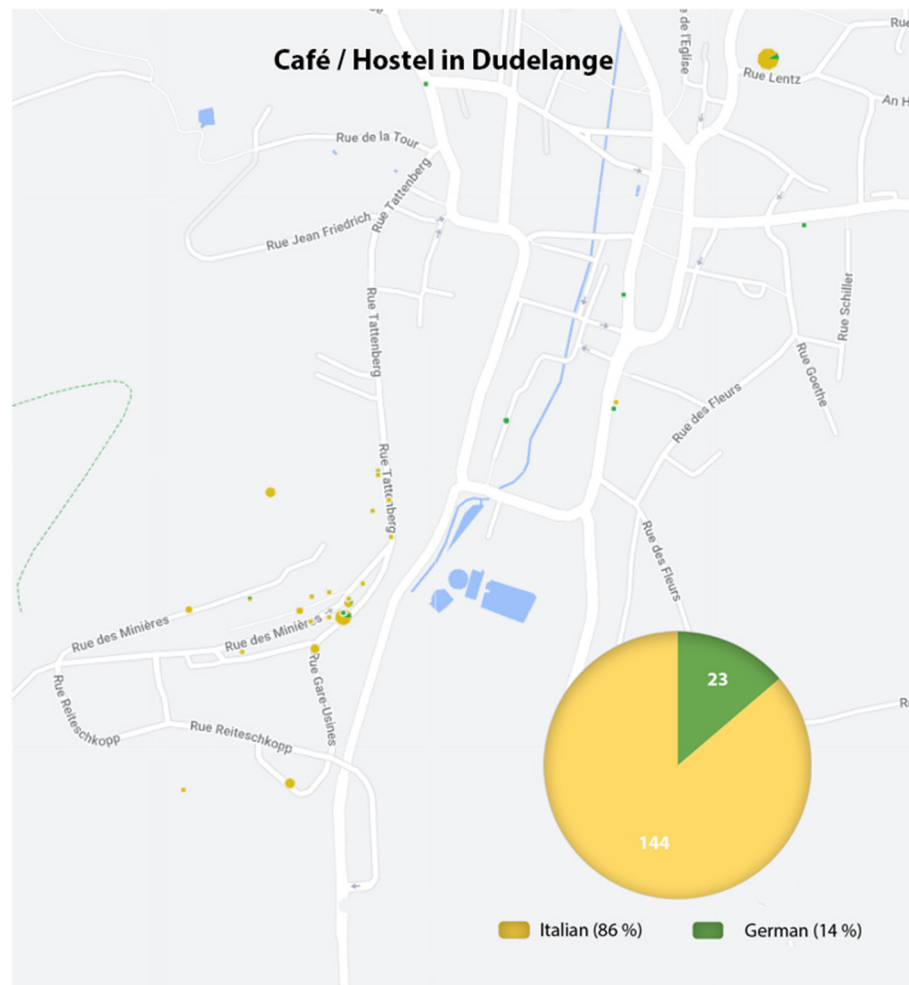
FIGURE 13
Declarants taking up residence in a *café-pension*.

Umbra. A search in an old repository of marriages revealed that the *café-pension* owner Francesco Giovanni Giuseppe Baldelli, born in 1877 in Gualdo Tadino, a village in the vicinity of Nocera Umbra, married on November 27, 1902 in Dudelange.[21] The *café-pension* owner and the nine Italian immigrants arriving at his *café-pension* came from the same village. All these declarants traveled directly from Nocera Umbra in central Italy to Dudelange, and most traveled in groups. Three declarants had left Nocera Umbra earlier for a stay in the United States, but had returned to Nocera Umbra before departing for Dudelange.[22] The nodegoat map shows not only that Italian migrants could come from one municipality or a few neighboring municipalities, but also where they arrived in Dudelange and the ties between the two places (Figure 14).

The direct migration path from Nocera Umbra to Dudelange indicates that family members and acquaintances followed earlier cohorts of migrants or that locals were recruited in Italy to come and work in Luxembourg (Gallo, 1987, p. 73–75). Official recruiters like Giulio Moschen, an Italian worker-entrepreneur in Dudelange, originally from Levico Terme in the province of Trento, traveled to Italy to recruit workers on behalf of the ARBED company (Reitz, 1984, p. 83–84; Wiegandt-Sakoun, 1986, p. 471–480). Other migrant workers were approached directly at customs posts or in major stations. In the 1920s, Italian migration slowed down because the Italian government tightened its control over emigration, and Luxembourg industrialists turned to other places: Poland, the Kingdom of Serbs, Croats and

---

21   ANLux, CT-03-02-0293.

22   DA, Dudelange, 1924, 449, 576, 944.

FIGURE 14
Visualization of the direct migration path of 28 Italian migrants from Nocera Umbra to Dudelange in 1924.

Slovenes, and the Ruhr region (Pairault, 1928, p. 77–110). For the small dataset used in this study, this consecutive mapping could also have been achieved using a more conventional approach, such as consecutive selection options in an Excel file. For the bigger dataset that we aim to compose in the future, however, this would no longer be possible.

Mapping spatiotemporal data from two towns

Geographic visualization offered us the possibility to map the spatiotemporal data from Dudelange and Differdange together and to analyze the mobility of migrants between these two towns. Only 21 declaration forms from Dudelange

and Differdange mention a place of residence in these two municipalities. The migrants were skilled German workers, such as metal workers. There was also a small proportion of apprentices in foods trades (e.g., butchers or bakers). Let us look at the case of the Engelberty brothers, which represents a fairly typical form of family migration.[23] Like many German migrants, Nicolaus and Jakob originated from the border area that had been part of Luxembourg until 1815. The two brothers were born respectively in 1896 and 1900 in Gashof, a few kilometers from the German-Luxembourgish border. Their forms indicate residence in Dudelange on April 16, 1924 and a job as "Hüttenarbeiter" at the main steel and mining company.

---

23   DA, Differdange, 1924, 178, 179; DA, Dudelange, 1924, 245, 246.
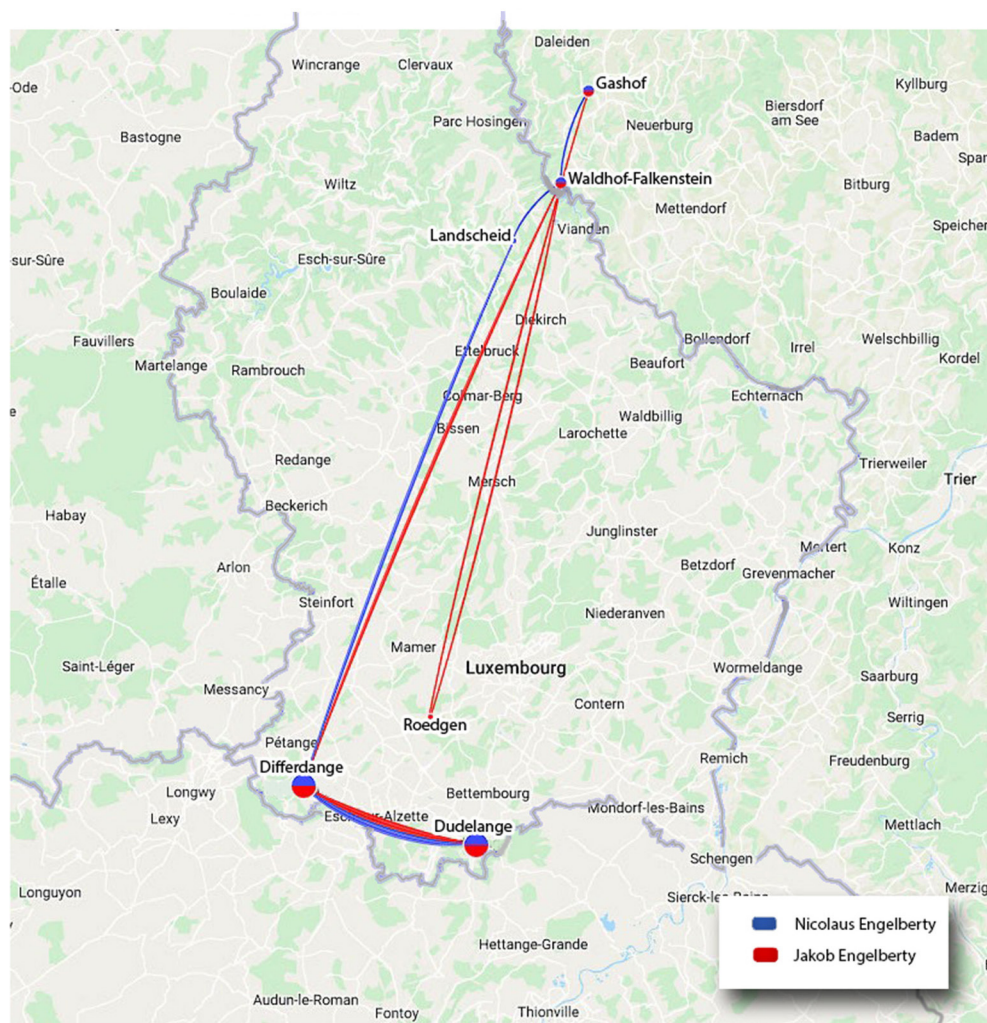
FIGURE 15
Migration paths of the brothers Jakob and Nicolas Engelberty between Germany and Luxembourg.

Aged 24 and 28 and both single, nothing seems to have been able to retain them in their rural home region, a source of emigration for decades, in particular to the industrial south of Luxembourg.

The two brothers did not come directly to Dudelange. Both were recorded on the same day, but 2 months earlier, in Differdange, where they were hired as "Hüttenarbeiter" at the Hadir steelworks. We observed a relative nomadism for workers, who moved quite easily from one company to another during an economic boom. The brothers' declaration forms in Differdange tell us that they were not on their first stay in Luxembourg. Jakob was registered for the first time in June 1922 in Roedgen, while Nicolaus lived in Landscheid from July 1923 onwards. In these rural towns, the brothers were probably hired as agricultural workers and continued what they had learned on the family farm. Once their contracts had come to

an end, the brothers left Luxembourg and found themselves in early 1924 in Waldhof, a German border town only separated from the Luxembourg town of Vianden by the Our river. It was from there that they both went to Differdange with their passports obtained on January 18, 1923 at the German Embassy in Luxembourg and their foreigner's booklet authorizing them to work legally. In both Differdange and Dudelange, the brothers lived together at the same addresses not far from the steelworks where they worked.

Although Jakob and Nicolaus should have mentioned their first entry into Luxembourg when arriving in Dudelange, they did not, and this information can only be retrieved by consulting their declaration forms in Differdange. Successive declarations for the same individual can be used to provide more comprehensive data on the migratory trajectory of declarants. The potential unreliability of residence data in declaration
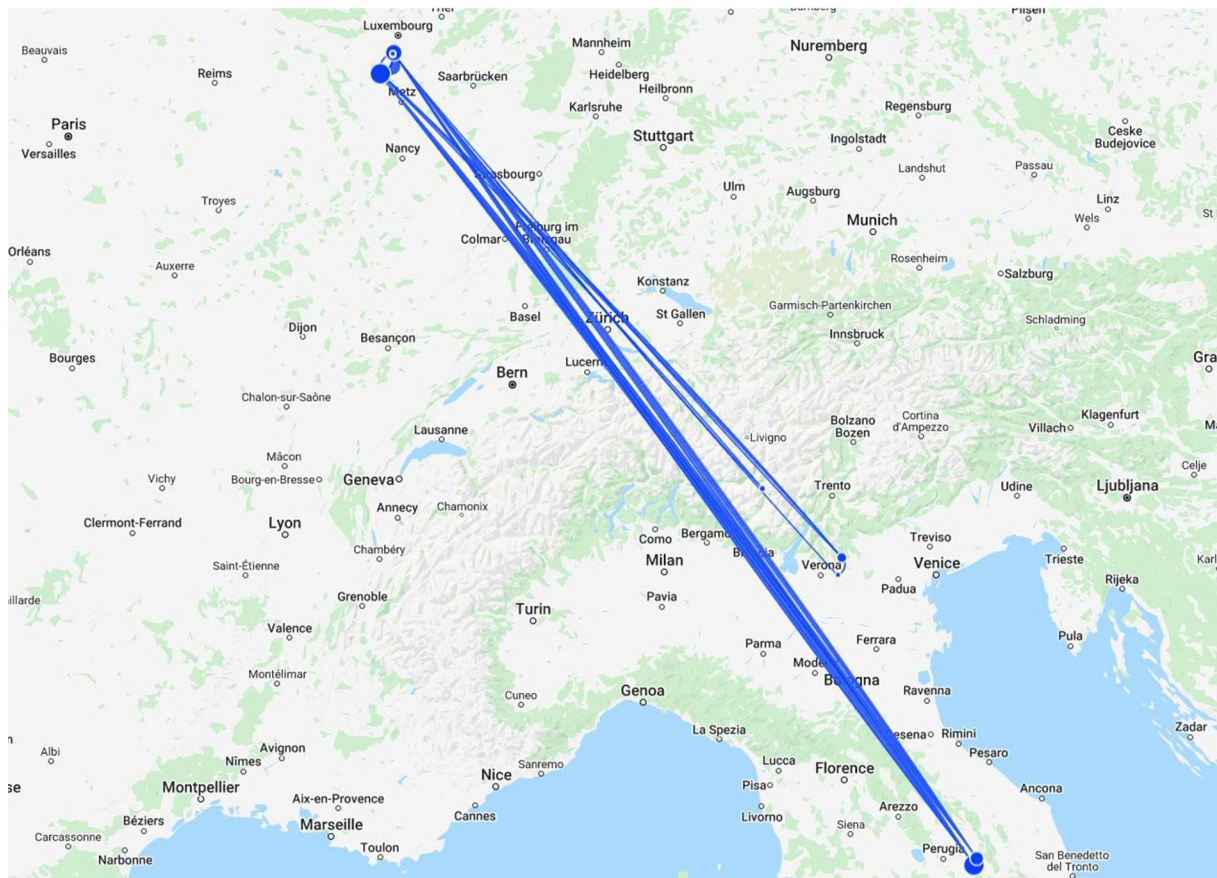
FIGURE 16
The migration trajectories of Italian migrants who declared residency in the French town of Trieux before arriving in Dudelange.

forms caused by the fact that declarants did not always present the same papers to civil servants tends to disappear when more data of an individual can be taken into account. This leads us to identify more detailed trajectories within and outside Luxembourg (Figure 15).

A social network analysis of waiting

More than 20 years ago, Dirk Hoerder encouraged researchers to decipher the importance of networks of individuals in order to understand migrations in the past (Hoerder, 1997, p. 85). Taking the example of Trieux, which was a popular short-term destination in Lorraine for migrants on the way from Italy to Luxembourg, we will indicate how social networking functioned for migrants on the road, and discuss the limits of nodegoat in visualizing this phenomenon accurately.

In Lorraine, major German and Italian communities emerged at the end of the 19th century following the establishment of new mines and factories. Algrange was referred

to as "little Berlin," and Italian migrants gathered in industrial towns in the Orne and Fensch valleys. A consulate was opened in Nancy in October 1911 and a consular agency followed in the district of Briey, to which Trieux belonged, in 1912. In 1913, that district numbered 46,700 Italians out of 75,000 foreigners (Bonnet et al., 1962, p. 92). The Sancy mine in Trieux was opened in 1903 and the town grew from 258 inhabitants in 1901 to 1,924 in 1911, the majority of whom were Italian workers. Trieux functioned as a stopover that was frequently cited by Italians arriving in Dudelange. One Italian who received financial compensation for a work accident in Luxembourg moved to Trieux, where he built a wooden hut that he rented out to newcomers from Italy (Jeandin, 1977, p. 29).

Mapping migration from Italy through Trieux to Dudelange in 1924 showed the migration paths of 14 Italians (Figure 16). However, the dots and lines on the map do not show how migrants used space and time. Figure 17 details the migrants' trajectories differently with a temporal social network visualization composed in the software program Tableau, demonstrating how social networks operated during migrants'

FIGURE 17
A temporal social network visualization of migration from Italy through Trieux to Dudelange.
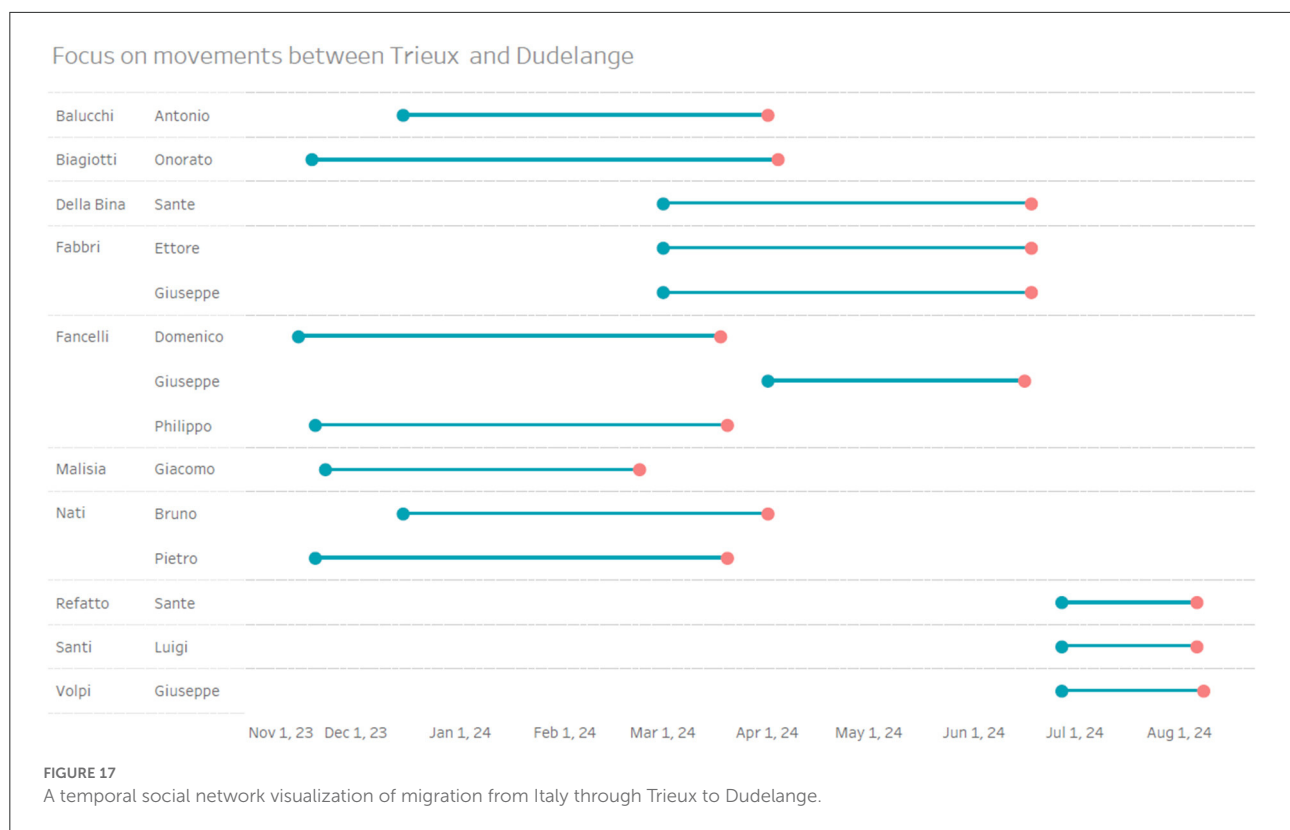
trajectories. It shows us who traveled together during the first part of the journey (e.g., Onorato Biagotti and Domenico Fancelli), who waited for whom in Trieux and for how long before traveling alone or together to Dudelange (e.g., Sante Della Bina and the Fabbri siblings waited for Giuseppe Fancelli), as well as how siblings followed each other (e.g., Nati). Based on the available data about Trieux in the database, we could not trace whether spouses traveled from Italy to Trieux alone, waited there for each other, and then traveled together to Dudelange, because upon arrival in Luxembourg the couple was declared on one form and only the migratory path of the male declarant was indicated in detail.

Conclusion

Mapping in nodegoat gave us new insights into migration trajectories and settlement patterns of migrants arriving in Dudelange in 1924. The ability to filter according to a vast number of criteria enabled us to decipher trends that would otherwise have been buried in the mass of data. For example, mapping unraveled a detailed relationship between migration and marriage. It also facilitated our work on precise samples through consecutive mapping, which provided knowledge about the personal ties between villages in central

Italy and some *café-pensions* in Dudelange. In addition, our data model allowed us to visualize data from declaration forms of different municipalities together, which made it possible to display more complex migration trajectories. Our digital hermeneutic approach ultimately allowed us to venture beyond nodegoat so that we could create and interpret comparative and cumulative chronological visualizations of the arrival of German and Italian migrants in Dudelange, as well as a temporal social network visualization of Italian migrants traveling to Dudelange through the French town of Trieux.

Visualizing migration data also revealed new avenues for further research. Our analysis of mobility was hampered by the absence of departure data and could be enriched by the data found in tens of thousands of handwritten resident cards of individuals or families produced in the interwar years, which kept track of arrivals and departures.[24] Another interesting avenue for further research would be to see when (and how) the mobility of unmarried declarants changed once they married. Preliminary research by Daniel Richter on living patterns in the town of Esch-sur-Alzette indicates that married couples

_____

24 Example: The resident card of Anton Vukovic indicating the dates of successive entries to (April 26, 1928 and September 9, 1928) and exits (July 9, 1928 and July 7, 1931) from Dudelange (Communal Archives of Dudelange).

very rarely lived in *café-pensions* and that unmarried women seemingly only lived in such homes in emergency situations, for example following a dispute with their husbands or in the case of (accused) prostitution (Richter, 2022). This research could be conducted on the basis of data included in the individual files compiled by the Luxembourg Alien Police for each migrant and conserved in the National Archives of Luxembourg. Qualitative sources such as letters from migrants and newspaper articles about migration could inform a more human-oriented approach to space and time modeling, expressing how spaces were made, observed and experienced by migrants themselves (Löw, 2016; Musekamp, 2019; Drucker, 2020).

Another possibility would be to research the often neglected aspect of transnational return migration (Oberbichler and Pfanzelter, 2021). Although sources created by the Luxembourg government were not intended to be used to document return migration, the declaration forms created in 1924 do enable us to document return migration during the First World War. Nodegoat visualizations show many Italian declarants in 1924 had already been residents of Dudelange before the First World War, and that they moved back to Italy within a year of war breaking out (Gallo, 1987, p. 289; Buzzi, 2015).

A last possible avenue for future research is the combined processing of declaration forms over more years and several municipalities within the Minett region, which would make it possible to complete migration data for individual declarants and to display more complex migration trajectories within the borderlands over a longer time period, potentially from a comparative perspective. Including more data will require a different methodological approach, such as the automated linking of historical data (Abramitzky et al., 2019).

Data availability statement

The original sources presented in the study can be consulted in the communal archives of Dudelange and Differdange in the Grand-Duchy of Luxembourg. Further inquiries can be directed to the corresponding author.

Author contributions

MV and AS wrote the article. MV conducted the case study on Dudelange. AS conducted the case study on Differdange.

Both authors contributed to the article and approved the submitted version.

Websites

www.geonames.org, accessed on April 25, 2022. www.geoportail.lu, accessed on April 25, 2022. https://legilux.public.lu, accessed on April 25, 2022.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abramitzky, R., Mill, R., and Pérez, S. (2019). *Linking Individuals Across Historical Sources: A Fully Automated Approach*. Cambridge: National Bureau of Economic Research, Working Paper 24324. doi: 10.3386/w24324

Blanc-Chaléard, M.-Cl., Bechelloni, A., Deschamps, B., Dreyfus, M., and Vial, É. (eds.) (2007). *Les Petites Italies dans le monde*. Rennes: Presses universitaires de Rennes. doi: 10.4000/books.pur.6543

Bonnet, S. (1976). *L'Homme du fer, Mineurs de fer et ouvriers sidérurgistes lorrains, vol. 1*. Nancy: Centre Lorrain d'Études Sociologiques.

Bonnet, S., Santini, Ch., and Barthélemy, H. (1962). *Les Italiens dans l'arrondissement de Briey avant 1914*. Nancy: Berger-Levrault. p. 92.

Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., and Schnapp, J. (2016). *Digital Humanities*. Cambridge, Massachusetts: MIT Press.

Burrows, T. (2017). *The History and Provenance of Manuscripts in the Collection of Sir Thomas Phillipps: New Approaches to Digital Representation*. Speculum 92/1, 39–64. doi: 10.1086/693438

Buzzi, P.-L. (2015). *Les immigrés italiens en Lorraine française pendant la Première Guerre mondiale (1914-1918)* (Master's dissertation supervised by Jean-Noël Grandhomme), Université de Strasbourg.

Chomé, F. (1964). *ARBED, Un demi-siècle d'histoire industrielle, 1911-1964*. Luxembourg: ARBED/Aciéries Réunies de Burbach-Eich-Dudelange.

Drucker, J. (2020). *Visualization and Interpretation: Humanistic Approaches to Display*. Cambridge, Massachusetts; London, England: The MIT Press. doi: 10.7551/mitpress/12523.001.0001

Faassen, M. V., and Hoekstra, R. (2022). Migrant visibility: Digitization and heritage policies. *Front. Hum. Dyn.* 4, 908456. doi: 10.3389/fhumd.2022.908456

Gallo, B. (1987). *Les Italiens au Grand-Duché de Luxembourg, Un siècle d'histoire et de chroniques sur l'immigration*. Luxembourg: Imprimerie Saint-Paul.

Galloro, P.-D. (2001). *Ouvriers du fer, princes du vent, Histoire des flux de main-d'œuvre dans la sidérurgie lorraine (1880-1939)*. Metz: Éditions Serpenoise.

Graham, S., Milligan, I., and Weingart, S. (2016). *Exploring Big Historical Data: The Historian's Macroscope*. London: Imperial College Press. doi: 10.1142/p981

Hayles, N. K. (2012). *How We Think: Digital Media and Contemporary Technogenesis*. Chicago and London: Chicago University Press. doi: 10.7208/chicago/9780226321370.001.0001

Hoerder, D. (1997). "Segmented macrosystems, networking individuals, cultural change: balancing processes and interactive change in migration," in *Citizenship and Exclusion*, ed V. Bader (Houndsmill: Palgrave Macmillan), 81–95. doi: 10.1057/9780230374591_5

Hohengarten, A. (1995). "'Polonia' oder die polnische Immigration in Luxemburg im 20. Jahrhundert," in *Itinéraires croisés, Luxembourgeois à l'étranger, étrangers au Luxembourg/Menschen in Bewegung, Luxemburger im Ausland, Fremde in Luxemburg*, eds A. Reuter and D. Scuto (Luxembourg: Éditions Le Phare), 194–199.

Jeandin, J. (1977). *Trieux, Soixante-dix-neuf jours au fond pour la Lorraine*. Paris: Social Éditions.

Klessmann, Ch. (1978). *Polnische Bergarbeiter im Ruhrgebiet: 1870-1945*. Göttingen: Brill GmbH. doi: 10.13109/9783666359828

Leboutte, R., Puissant, J., and Scuto, D. (1998). *Un siècle d'histoire industrielle (1873-1973), Belgique, Luxembourg, Pays-Bas, Industrialisation et sociétés*. Paris: Éditions SEDES. doi: 10.3917/sedes.lebou.1998.01

Lehners, J.-P. (1991). "Wohnen in Düdelingen zu Beginn des 20. Jahrhunderts," in *Tageblatt 222* (Esch-sur-Alzette: Tageblatt), 9–12.

Löw, M. (2016). *The Sociology of Space – Materiality, Social Structures and Action. Cultural Sociology*. Houndsmill: Palgrave Macmillan. doi: 10.1057/978-1-349-69568-3

McCarty, W. (2005). *Humanities Computing*. Houndsmill: Palgrave Macmillan. doi: 10.1057/9780230504219

McKinsey, L., and Konrad, V. (1989). *Borderlands Reflections: The United States and Canada*. Orono, Me: University of Maine.

Musekamp, J. (2019). "Big history and local experiences: migration and identity in a European Borderland," in *Mapping Migration, Identity, and Space*, eds. T. Linhard and T. Parsons (Houndsmill: Palgrave Macmillan), 55–83. doi: 10.1007/978-3-319-77956-0_3

Noiriel, G. (2001). *État, nation et immigration, Vers une histoire du pouvoir*, Paris: Belin.

Noiriel, G. (2019). *Immigrés et prolétaires, Longwy, 1880-1980*, Marseille: Agone.

Oberbichler, S., and Pfanzelter, E. (2021). Topic-specific corpus building: a step towards a representative newspaper corpus on the topic of return migration using text mining methods. *J. Dig. Hist.* 1, 74–98. doi: 10.1515/jdh-2021-1003

Pairault, A. (1928). *L'Immigration organisée et l'emploi de la main-d'œuvre étrangère en France*. Paris: PUN.

Palomba, P., Montanari, R., and Garzia, E. (2020). "FORMAL. Mapping Fountains over Time and Place. Mappare il movimento delle fontane monumentali nel tempo e nello spazio attraverso la geovisualizzazione," in *Digitized and Digitalized Humanities: Words and Identity. Atti del IX Convegno Annuale AIUCD. La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica, AIUCD*, ed C. Clivaz (Milano: AIUCD), 67–73.

Petruş, A. (2020). *Network Analyses of Foreign Travelers Through Wallachia, Moldavia and Transylvania Between 1831-1840*. Studia Universitatis Babe?-Bolyai – Digitalia 65/2, 31–47. doi: 10.24193/subbdigitalia.2020.2.03

Pisarevskaya, A., Levy, N., Scholten, P., and Jansen, J. (2020). *Mapping Migration Studies: An Empirical Analysis of the Coming of Age of a Research Field*. Migration Studies 8/3, 455–481. doi: 10.1093/migration/mnz031

Ponty, J. (2005). *Polonais méconnus, Histoire des travailleurs immigrés en France dans l'entre-deux-guerres*. Paris: Publications de la Sorbonne.

Prêcheur, Cl. (1959). *La Lorraine sidérurgique*, Paris: S.A.B.R.I.

Presner, T., and Shepard, D. (2015). "Mapping the geospatial turn," in *A New Companion to Digital Humanities*, eds S. Schreibman, R. Siemens, and J. Unsworth (New York: John Wiley and Sons), 199–212. doi: 10.1002/9781118680605.ch14

Rainhorn, J. (ed.) (2005). *Petites Italies dans l'Europe du Nord-Ouest*, Valenciennes: Presses universitaires de Valenciennes.

Reisinger, G. (2010). "Digital source criticism: net art as a methodological case study," in *Netpioneers 1.0: Contextualizing Early Net-Based Art*, eds D. Daniels and G. Riesinger (Berlin: Sternberg Press), 123–142.

Reitz, J. (1984). *L'immigration étrangère à Differdange au début du XXe siècle (1898-1914)*, Mémoire de maîtrise sous la direction de François Roth, Université de Nancy II.

Reuter, A. (2018). "La ville de Dudelange à la veille de la Première Guerre mondiale, un étonnant laboratoire du vivre en commun," in *Être d'ailleurs en temps de guerre (1914-1918), Étrangers à Dudelange, Dudelangeois à l'étranger, Mutations, Mémoires et perspectives du bassin minier, no. 10* (Luxembourg: Fondation Bassin Minier), 31–39.

Richter, D. (2022). "Making home in the neighborhoods of Esch-sur-Alzette (LU) during the first half of the 20th century," in *Paper Presented at the International Migration Conference in Dudelange on June 16, 2022* (Dudelange).

Schrover, M., and Moloney, D. M. (2013). *Gender, Migration and Categorisation: Making Distinctions Between Migrants in Western Countries, 1945-2010*. Amsterdam: Amsterdam University Press.

Scuto, D. (2010). "Histoire des immigrations au Luxembourg (XIXe-XXIe siècles)," in *25 ans d'action pour l'immigration, 1985-2010, OGBL, 25ème anniversaire du Département des Immigrés (*Luxembourg: OGBL/Département des Immigrés), 13–38.

Scuto, D. (2012). *La nationalité luxembourgeoise (XIXe-XXIe siècles)*. Brussels: Éditions de l'Université de Bruxelles.

Shaw-Taylor, Y., and McCall, L. (2020). *Immigration, Assimilation, and Border Security*. Lanham: Bernan Press.

Sömme, A. (1930). *La Lorraine métallurgique*. Paris: Éditions Berger-Levrault.

Stiegler, B. (2014). *Digital Studies: organologie des savoirs et technologies de la connaissance*. Limoges: Fyp éditions.

Trausch, G. (2017). *Histoire économique du Grand-Duché de Luxembourg, 1815-2015*. Luxembourg: Imprimerie Centrale S.A.

Trinkaus, F. (2014). *Arbeiterexistenzen und Arbeiterbewegung in den Hüttenstädten Neunkirchen, Saar und Düdelingen, Luxemburg: 1880-1935/40: ein historischer Vergleich*. Saarbrücken: Kommission für Saarländische Landesgeschichte.

van Bree, P., and Kessel, G. (2017). *Iterative Data Modelling*. From Teaching Practice to Research Method. Abstract retrieved from Digital Humanities Montreal database.

Walter, G. (1935). *L'évolution du problème de la main-d'œuvre dans la métallurgie de la Lorraine*. Mâcon: J. Buguet-Comptour.

Weitin, T. (2017). Scalable reading. *Zeitschrift für Literaturwissenschaft und Linguistik* 47, 1–6. doi: 10.1007/s41244-017-00-4

Wiegandt-Sakoun, C. (1986). "Les missions catholiques italiennes dans l'entre-deux-guerres: l'exemple français," in *Les Italiens en France de 1914 à 1919*, ed P. Milza (Rome: École française de Rome), 471-480.

Williams, L. (2013). "Transnational marriage migration and marriage migration," in *Transnational Marriage. New Perspectives from Europe and Beyond*, ed K. Charley (New York, NY: Routledge), 32–46. doi: 10.4324/9780203111772-9

Zephyr, F. (2015). "Spatial history as scholarly practice," in *Between Humanities and the Digital*, eds P. Svensson and D. T. Goldberg (London and Cambridge: MIT Press), 411–428.

Zind, P. (1979). *Elsass-Lothringen Alsace-Lorraine, une nation interdite, 1870-1940*. Paris: Éditions Copernic.

Collecting migrants' Facebook posts: Accounting for ethical measures in a text-as-data approach

Helena Dedecek Gertz*

Department of Intercultural Education Research, Faculty of Education, Hamburg University, Hamburg, Germany

Based on the heuristics proposed by Helen Nissenbaum to assess ethical issues surrounding research using new technologies, this paper discusses the ethics of the collection and analysis of migrants' digital traces for academic research purposes. Concretely, this paper is grounded on an empirical research that applies a topic modeling approach to a large dataset of migrants' posts written on Facebook groups. After discussing the nine aspects proposed by Nissenbaum, the paper contends that while researchers strive to comply with ethical measures by, for instance, asking adequate questions and protecting the collected data, the lack of transparency of social networking sites is harmful to critical social sciences and can hamper findings that contribute to understanding migratory patterns and decisions.

Introduction

The analysis of migrants' media use can produce valuable knowledge about decision-making and networks in migratory context. Methodologically, the interest in researching migrants' media use has added to the complexity of carefully handling migrants' data, and obtaining informed consent for research purposes. Accordingly, the ethics of collecting migrants' digital traces has been gaining attention, particularly among qualitative researchers (Leurs, 2017; Siapera and Creta, 2020; Sandberg et al., 2022b). Most of these ethical reflections agree that, because of the vulnerability of certain migrant populations, researchers need to go beyond procedural ethics and care for the safety and well-being of researched subjects.

Quantitative studies based on migrants' digital traces generate different problems relating to "profiling, informed consent, data sharing processes and ethical approval and data management procedures" (Mahoney et al., 2022b, p. 230). As there are fewer studies about migration applying topic modeling to social media data created by migrants, there are correspondingly fewer analyses on the ethics of collecting and analyzing such quantitative data. Mahoney et al. (2022b, p. 232) analyzed large textual datasets from migrants on

Twitter, collecting only "explicitly public social media data". They contend that ethical issues of such data collection become more intricate the more social media develops and the identification of public and private spaces becomes more complex (Idem, p. 235). Elsewhere, they carefully commend that datasets coming from migrant Facebook groups require consent, while collecting migrants' Twitter data would be closer to observing public behavior and therefore less problematic (Mahoney et al., 2022a; p. 339–340). A similar recommendation comes from Sandberg et al. (2022a).

As detailed on the section "Comparative evaluation based on studies using Facebook posts and a topic modeling approach," studies using large amounts of Facebook texts tend to acknowledge that their methodological procedure can be liable to ethical critique, but do not analyze that ethical critique systematically. The most common solution for this dilemma tends to be to collect data that is interpreted as public or with fewer privacy constraints, such as posts from profiles with less privacy settings. By discussing the collection of large textual datasets posted by migrants on a social media platform, or Facebook more specifically, this paper systematically analyses the ethical decisions of an empirical research situation and argues in favor of research in digital humanities and social sciences. Against that background, this paper asks "how to justify the collection and analysis of migrants' digital traces for academic research purposes?" The discussion emerges from the procedures of collection and analysis of quantitative textual data from Facebook groups of migrants and aspiring migrants. Here, the outcomes of that empirical analysis are put in the background, giving way to a detailed reflection on the choices and consequences of the methodological decisions.

Following, first, the context and research design of the empirical base study is outlined. After that, the ethical issues and corrective measures are discussed guided by the heuristics of "contextual integrity" of information flow within new technologies proposed by Nissenbaum (2010) and as applied by Zimmer (2018). Although Nissenbaum presented her heuristics over a decade ago, they are based on broader concepts which make them comprehensive and abstract enough to be applied to different analytical situations. So much so that the heuristics are appropriate to discuss different empirical topics, such as data breaches from a dating app (Zimmer, 2018) and migrants' posts on Facebook groups. The paper summarizes the ethical boundaries of automated data collection, as encountered in the empirical base study conducted by Helena Dedecek Gertz and Florian Süßer (henceforth "we", "our" or "the authors") and presents our suggested measures to comply with migrants' data protection, adding up to the arguments for a reflexive and critical data collection based on ethics of care (Leurs, 2017). The central argument is that, although acknowledging that collecting textual data from social media users without their explicit consent is rightfully

prone to critique, researchers, as a community, can care for migrants' anonymity throughout the process of research by making careful decisions to this end and by asking adequate research questions.

Context of the data collection

The data that motivates the discussion here derives from a research project that aimed at identifying the roles of media in migratory pathways relating to education. Empirically, the project focuses on media uses of Brazilians who live in Germany or who aspire to do so. The project was based on a mixed-methods approach, consisting of a qualitative content analysis of interactions in migrant Facebook groups, qualitative interviews with participants of these groups, and a topic modeling of posts made in the groups. The ethical discussion in this paper derives from the empirical paper that applied the topic modeling approach to establish the prevalence of topics relating to education in debates among Brazilian migrants in Facebook groups. The outcome reveals that vocational education and training (VET) and language learning for certification purposes are the most relevant education-related topics debated among these migrants.

The background of that project is based on research that shows that, in the context of migration, formal education can represent a means to secure residence status, access the job market, and acquire certificates that contribute to building migrants' cultural capital (Waters, 2015). People who migrate to pursue educational pathways contribute to the transnationalization of educational institutions in the country of destination. Transnational education (TNE) is more often approached in research about higher education; nevertheless, migrants from families with low income and low educational attainment are also actors in TNE. Fürstenau (2019) and Carnicer (2019) have described how Brazilian women from such backgrounds migrate to Germany first as Au Pairs, then complete VET (which is usually remunerated in Germany), and thereby secure both employment and stable residence status.

Based on that background, the topic modeling (the empirical analysis that motivates this paper) had two assumptions. Based on the findings presented in the previous paragraph, one is that access to education can be a motivator for migration across socioeconomic classes, i.e., not only among migrants who can afford the pursuit of a university degree or educational exchanges abroad. The other assumption is that information exchanges through social media platforms are important for migrants' decision-making (Dekker et al., 2018; Richter et al., 2018). Although studies in this direction are mostly conducted among migrants who fled war and conflict using their smartphones' to evaluate the safest options to reach their countries of destination, information and communication technologies (ICTs) and exchanges with latent

ties (Haythornthwaite, 2002) are relevant in other migratory contexts, such as those associated with educational aspirations (Jayadeva, 2020). Based on these two assumptions we contended that people cross borders, regardless of their socioeconomic background following educational projects, and that digitally mediated communication, particularly through social media, plays a role in decision-making for these projects.

Specifically for the Brazilian case, similar findings confirm the relevance of social media information exchange in contexts of migration. Brazilian migrants have been exchanging information on social media for at least a decade when the most used platform among them was Orkut (Schrooten, 2012; Oosterbaan, 2013). Nowadays, Brazilian migrant groups on Facebook groups have taken on that role in these online interactions (Foletto, 2018). Most of these studies on Brazilian migrants' on social media describe its uses for solving bureaucratic issues, job-seeking, and also for organizing social gatherings. Educational aspects remain under-commented, although education is a means to fulfilling migratory pathways and it can become part of migrants' life once they are established in the country and their children start attending school. While it is known that migrants who wish to pursue university degrees abroad use social media to facilitate that process (Jayadeva, 2020), the role of media use for achieving other educational levels, such as VET or schools for migrants' children, remains understudied.

Against that background, the quantitative textual analysis that motivated this ethical reflection reveals that topics relating to education, VET, and language-learning in particular, are among the most prevalent ones in information exchanges on Facebook groups of Brazilians in Germany. That conclusion was only possible through a topic modeling approach, which demanded the collection of quantitative textual data produced by migrants in the context of a social media platform. The following sections reflect on the ethics of collecting and analysing this data produced by individuals that are potentially vulnerable due to their legal status in Europe. Following, the methodological decisions that were ethically critical for this analysis are detailed; after that, we analyze our decisions based on the heuristics to guide ethical decision making in projects involving ICTs proposed by Nissenbaum (2010) and commented by Zimmer (2018). We conclude by arguing that, while researchers strive to comply with anonymization and data security, the lack of transparency from social media platforms can be harmful for critical, independent, and public-interest-oriented research, which in turn can impair the development of knowledge about social phenomena.

Methodological decisions

In this section, we first discuss the choice of Facebook as a data source followed by an overview of ethical discourses in research about migrants' social media use and digital data collection. After that, we present our rationale for choosing Facebook groups adequate to answer our research question, explain our procedure of textual data collection, and argue in favor of a topic modeling approach to analyze the data.

Creating a Facebook account for research

In migration research, Facebook has been mostly used as an empirical data source in qualitative approaches. Some accounts based on interviews about Facebook use among migrants (Leurs, 2014; Dekker et al., 2018) are exempt from a discussion such as the one we propose here, as informed consent can be acquired. As Leurs (2017) observes, however, researchers must still be careful with publishing digital traces of migrants, such as print screens or detailed information about certain media use patterns, even though interviewees themselves might have agreed to provide such data. That position is aligned with a way of arguing for a careful collection, management, and analysis also of quantitative textual data from Facebook, as "informed consent does little to protect participants" (Brown et al., 2016, p. 855) and researchers share the responsibility of caring for research participants' privacy and anonymity at all situations. Following such principles of care and transparency towards research participants, one of the authors created an account on Facebook.

The Facebook profile used for research was created using the researcher's real name, and with information identifying her as a researcher. Some friends and acquaintances added her and she joined five groups of her private interest (university and academic research related). She "liked" 76 public pages, most of them from organizations of Brazilians in Germany, but also some university profiles and a few of private interests. Finally, she joined 43 groups of Brazilians in Germany. Although she created this profile for research purposes, it is not a dummy account used simply to collect data, because she is clearly identified, with her name and picture, and with information signaling her as a researcher at Hamburg University. Her university e-mail and the website leading to the university's website, where she figures as a researcher, are also available on the page, as to make public other forms of contact with her (e-mail and telephone number on the faculty website) and to have some proof of her identity (the link to the website). Indeed, one person with whom the researcher got in touch through Facebook to ask for an interview for the qualitative part of the study, replied to her *via* email— and not through the Facebook chat where she sent out the message requesting the interview—to "be sure about the identity of the person who contacted me", as the potential interviewee explained. Also, Facebook allows users to add free

text to their profiles. In this space, she wrote her position at the university, the name of the project she works for, and informed that the profile was created for research purposes. This information was written in Portuguese, German, and English. Apart from this research-related information, there are traces of her personal interests (university groups and "likes" on pages) and information about her background (the town where she was born, where she currently lives, and her educational pathway).

Contextualizing migrant Facebook groups selected for research

This section contextualizes the space of our data collection, namely Facebook groups gathering Brazilian migrants in Germany. First, we define these groups. Afterward, we describe the rationale behind the choice to analyze groups of Brazilians in Germany and the data collection procedure. Finally, we argue in favor of our decision to work with quantitative textual data in this context accounting for the research quality and ethics of our decisions. This section is already part of the contextual integrity analysis. Nissenbaum (2010) proposes nine points for the decision heuristics (see section "Discussing ethical decisions of research with migrants' textual data"). The second point of the heuristics is to identify the prevailing context of the information flow. As this section does exactly that, namely giving background information about the source of the data, this section substitutes the section "Prevailing context".

Facebook groups of Brazilians in Germany gather registered users with similar interests, locations, jobs or professions, and aims. Some of these groups are public, meaning that their content can be seen by any other user logged on to Facebook. Other groups are private and might request users to fill up a form upon entry in order to be accepted by the administrators. Posts and comments on these groups are visible to all participants. These two types of groups can be found using Facebook's search tool and were included in this collection. There are "secret groups" for which one has to be invited to participate—there are none of these types in the dataset.

Although it has been argued (e.g., Naughton, 2022) that the use of Facebook has been declining, Brazilian migrant groups are still active and diverse, ranging from the general "Brazilians in Germany", to the location-based groups, like "Brazilians in [German city]", to specific groups like "Brazilian women in Germany", to work-related groups, such as "Brazilian IT professionals in Germany", "Brazilian Au Pairs in Germany", aim-related groups "Ausbildung in Germany from Brazil", and other interest groups "Gardening for Brazilians in Germany". Finally, Brazilian migrants' fondness for social media groups is not new, as it has been reported already over a decade ago (Oosterbaan, 2010; Schrooten, 2012), nevertheless its uses for educational projects have not been studied. The paid VET

programmes in Germany have the potential to attract migrants who are excluded from tertiary education in their countries of origin, as is the case of some Brazilians (Carnicer, 2019; Fürstenau, 2019)—hence the choice to focus our research on Germany as a country of destination. As to the decision to focus on Brazilians, the best-described case of migrants from low-income backgrounds pursuing tertiary education in a European country seems to be that of Brazilians. Similarly, the well-described case of Brazilians using social media to establish migrant networks in Europe (Oosterbaan, 2010; Schrooten, 2012; Foletto, 2018) lays the grounds for the choice of that particular nationality.

To select relevant groups for the empirical research, first Facebook was searched for the terms "Brazilians" and "Germany" (in Portuguese). After that, groups relating to migration were selected and these were once again filtered according to their level of interaction: using the information provided by Facebook itself, the author joined 43 "active groups", i.e., groups with at least a thousand participants and three posts made in 1 week. From that total, 30 groups required participants to fill up a form upon entry to inform group managers about their interest in joining the group. We used the forms to inform the managers about our research interests, data collection, and anonymization measures. To perform the analysis, we used the structural topic modeling (STM) approach (Roberts et al., 2019), which allowed us to correlate the posts with the groups they came from. In this context, a topic is "a mixture over words where each word has a probability of belonging to a topic" (Roberts et al., 2019, p. 2). The outcome of the procedure reveals that among the seven most relevant topics, two are related to education, particularly to language learning and accessing vocational education and training in Germany. Across all groups, there is some mention of one of these topics, meaning that in a group gathering Brazilians in a specific city (but not explicitly related to education) or in a group gathering Brazilians who wish to pursue a degree in Germany, there is some mention of both education-related topics.

Automated textual data collection using a web-scraper

The empirical data for this analysis is textual and comes from posts and comments made by group participants. We solved the issue caused by the "APIcalypse" (Bruns, 2019) by automating our data collection using the WebDriver API *Selenium*, which allows us to automatically control a web browser. The scraper logs into the researcher's Facebook account and systematically goes through the groups that we could join. All posts, comments, and sub-comments were copied to a local file system. *Selenium* controls the web browser as if a human is sitting in front of it: all data obtained is exactly the same data available to the human Facebook user. No clicking behavior or friends list is collected, for instance. A human could do the same procedure,

however with a much bigger investment of time and effort. Although we could have set up a Facebook Developer account that would allow us to use Facebook's API, the process using *Selenium* is not subject to Facebook's Graph API which includes rate limits. Summing up, automation simply sped up the process of data collection.

Reliability, reproducibility, and ethics in migrants' textual data collection

For the automated data collection, we considered whether and to what extent we would be violating terms of use from a giant social media company and if that would make us liable to a legal process. In that regard, we argue with those who stand for critical research (Hargittai and Sandvig, 2016; Bruns, 2019) and we support that independent data collection for social science and digital humanities research is rules-based and can comply with user privacy. Still, that does not solve the issue of the impossibility of gaining consent from all users when conducting such large data collection.

There are central differences between big techs' data extractivism and our procedures for data collection, storage, and analysis. These differences are due to the scientific interest, access, and use of the data. Regarding our data gathering, we as researchers using a Facebook user's interface with *Selenium* only had access to what other individual Facebook users also have. In that sense, we could control what information was collected. As for the storage procedure, the textual data was saved on a file shared only between the two authors of the empirical paper and that could only be accessed through a closed network. The applied "text as data" approach (Grimmer and Stewart, 2013) follows a standardized and, theoretically, reproducible methodology while complying with measures for data protection and having no financial profit. Nevertheless, because data from Facebook groups can be erased, entire groups or Facebook itself can cease to exist, and the platform can change its access rights, a reproducibility test is unlikely to result in the same *corpus*, hence such a test is not feasible.

It could be argued that there are other ways to research media use for educational projects on migration that would not demand a large textual data collection, therefore sparing this paper's discussion. That critique can also be directed to the nature of such quantitative textual data from social media: these Facebook posts are not connected to traits that characterize social positionality (e.g., socioeconomic background, educational attainment), which does not allow for an analysis that accounts for inequality and discrimination. In that sense, we agree with Leurs' critique (Leurs, 2017) that such procedures assume a "detachment from a discrete, knowable world" and tend to "naturalize the politics of knowledge production" (p. 134). However, because our empirical research question could only be answered with quantitative textual data focused on the prevalence of interactions involving

education in migrant Facebook groups, the lack of such background information about group participants is not critical. Our topic modeling study is part of a larger project that includes a qualitative content analysis of the posts made in these Facebook groups and a digital ethnography, for which 30 group participants were interviewed. Following the logic that "all quantitative models of language are wrong—but some are useful" and that topic models need validation (Grimmer and Stewart, 2013, p. 269–270), apart from providing new research outputs by themselves, these qualitative approaches were used to validate findings from the topic modeling.

Finally, we do not treat data as "public" (Zimmer, 2010): we did not reproduce posts word-by-word, both as an ethical measure and because that was not useful for answering our empirical question. As a final compliance measure to counterbalance the impossibility of getting consent from group users and to reassure that anonymity is preserved, the empirical paper was presented to group managers and opened to their critique.

Discussing ethical decisions of research with migrants' textual data

After contextualizing the study that generated this debate in the first section and describing methodological decisions in the previous section, we move on to discuss ethical decisions of our research design based on Nissenbaum's (2010) nine points for a contextual integrity analysis when using emerging media technologies for research. We also rely on ethical guidelines for internet research elaborated by the Association of Internet Researchers (AoIR) (franzke et al., 2020, p. 9-23).

Nissenbaum's nine points for contextual integrity are the following:

1. Describe the new practice in terms of its information flows (see our section Information flows).
2. Identify the prevailing context in which the practice takes place at a familiar level of generality, which should be suitably broad such that the impacts of any nested contexts might also be considered (section Prevailing context referring to section Contextualizing migrant Facebook groups selected for research).
3. Identify the information subjects, senders, and recipients (section Information subjects, senders, and recipients).
4. Identify the transmission principles: the conditions under which information ought (or ought not) to be shared between parties. These might be social or regulatory constraints, such as the expectation of reciprocity when friends share news, or the obligation for someone with a duty to report illegal activity (section Transmission principles and its subsections).

5. Detail the applicable entrenched informational norms within the context, and identify any points of departure the new practice introduces (section Detail the Entrenched Information Norms and its subsections).

6. Make a prima facie assessment: there may be a violation of contextual integrity if there are discrepancies in the above norms or practices, or if there are incomplete normative structures in the context to support the new practice (section Prima facie assessment).

7. Evaluation I: Consider the moral and political factors affected by the new practice. How might there be harms or threats to personal freedom or autonomy? Are there impacts on power structures, fairness, justice, or democracy? In some cases, the results might overwhelmingly favor accepting or rejecting the new practice, while in more controversial or difficult cases, further evaluation might be necessary (section Evaluation I).

8. Evaluation II: How does the new practice directly impinge on values, goals, and ends of the particular context? If there are harms or threats to freedom or autonomy, or fairness, justice, or democracy, what do these threats mean in relation to this context? (section Evaluation II).

9. Finally, on the basis of this evaluation, a determination can be made as to whether the new process violates contextual integrity in consideration of these wider factors (section Final determination). (Nissenbaum, 2010; p. 182–183).

In what follows, we address these items proposed by Nissenbaum thereby analyzing the ethical issues of collecting migrants' digital traces for academic research.

Information flows

There are at least nine information flows on Facebook:

1. The first one is from the users to the company "Facebook" (clicking patterns, location, cookies).

2. The second one is the flow of information about the author who created the profile from the Facebook service to search engines and other non-users (if the profile is indexed on Google, for instance).

3. The third one is non-textual information from the profile owner to other users registered on Facebook (such as "likes" on profile pages and participation in groups).

4. The fourth one is from the private posts and friends list on the personal profile of the author who created the profile to their friends.

5. The fifth one is composed of posts written on private groups (groups, for which the administrator has to grant access to the requester), which can only be read by other group participants.

6. The sixth one is composed of posts written on public groups (groups, whose content can be seen by people who are not participating in it), which can be read by anyone who opens the group link.

7. The seventh one are replies to questionnaires elaborated by group administrators, as a requirement to be accepted in certain private Facebook groups.

8. The eight one are direct messages exchanged through the Facebook chat (which was used to contact potential interview partners for the qualitative study), i.e., a two-way flow between the profile owner and another person.

9. The ninth one are multiple-way direct messages exchanged among a closed group through the Facebook chat.

From this list, only the flows described on numbers 4 and 9 were not part of the interactions of the author who owns the profile, as she did not post anything on her private profile and did not send direct messages to multiple people. We did not create a dummy profile. Although the profile was created for research purposes, the owner was clearly identified on it and she did not try to conceal her intentions to group administrators when filling in questionnaires requesting to join the groups. By running the web-scraper, we did introduce a tenth information flow from the groups to our closed database, however our data did not include users' personal information. Finally, as our database is not public, the raw information flow is kept within the circuit of Facebook users (including the author who has a Facebook profile) and will be destroyed as soon as the research is concluded. One could argue that, once the analysis based on this data is published, there would be the eleventh flow of information toward the general public, nevertheless, that information is anonymized, filtered, and analyzed based on a specific research question. That flow is not of raw data; thus, it is a new circuit of information flow (from the publisher to its readers, etc.).

Prevailing context

The prevailing context relates to the social context in which data is gathered (Nissenbaum, 2010, p. 149). For this paper, the prevailing context is that of Facebook groups of Brazilian migrants in Germany, as described particularly on section Contextualizing migrant Facebook Groups selected for research. This includes, among other aspects, the high level of use of Facebook among Brazilian migrants for networking and the opportunity of migrating to Germany to pursue tertiary education.

Information subjects, senders, and recipients

In the context of our data collection, "information subjects" are the Facebook users who interact by writing posts with a question or a piece of information or by commenting on those posts in groups of Brazilians in Germany. People who participate in the groups but never had any interaction on them are not our information subjects. Within these information subjects, the senders are those who pose questions or share other kinds of information on the groups and those who reply to such questions. The recipients are all group members who read the interactions (either group participants or not, in the case of public groups). Because our dataset is not publicly available, we did not expand the role of recipients to the general public.

Transmission principles

Transmission principles are rules that constrain the information flows (Nissenbaum, 2010, p. 145). In our case, there are three such sets of rules:

1. Because most of the information flows happen within a context controlled by a private company, Facebook's Terms of Service (ToS) are one of the regulators.
2. Because the data was collected in Germany, the General Data Protection Regulation (GDPR) from the European Union is a second regulator.
3. Because the collected data is part of an academic research, academic research ethics guidelines (e.g., franzke et al., 2020) apply.

Following, we comment on the central guidelines from these three sets of rules.

Facebook ToS

Facebook prohibits scrapping, but not manual data collection. We could have done the same procedure manually and acquired the same data however securing anonymity here would have been even worse because the person manually collecting the data would have seen what each group participant has written. The company's decision to prohibit scrapping after the Cambridge Analytica scandal is probably useful in constraining other companies to harvest and sell personal data that could be used for skewing public opinion on matters such as migration. However, that decision is harmful to social research (Sandvig, 2017; Bruns, 2019; Mancosu and Vegetti, 2020).

GDPR

GDPR recognizes that "by coupling information from registries, researchers can obtain new knowledge of great value

… within social science, research on the basis of registries enables researchers to obtain essential knowledge about the long-term correlation of a number of social conditions such as unemployment and education with other life conditions". As GDPR defines personal data as "any information relating to an identified or identifiable natural person" and as Facebook posts always appear associated with a profile, these texts could be interpreted as personal data. However, GDPR highlights that "information that identifies an individual … may be personal data if you are processing it to learn something about that individual or if your processing of this information will have an impact on that individual". In this sense, the data we collected is in a gray zone: it is being used neither to learn something about an individual in particular nor to undertake actions that would have any foreseeable impact on an individual. In fact, we use these data to describe the social world, more specifically, digital information exchange relating to transnational education and migration.

Kotsios et al. (2019, p. 6–10) provide further instructions to assess the consequences of GDPR's seven principles relating to the processing of personal data in social media research. We present these seven principles and associate them with our case based on the comments by Kotsios et al. (2019) and on our experience with the empirical data described in the previous paragraphs:

1. Lawfulness, fairness, and transparency: Processing must be lawful, fair, and transparent to the data subject.

    Because Facebook itself does not provide "transparent data access to critical, independent, public-interest research" (Bruns, 2019, p. 1561), we cannot fully comply with this point. Facebook managers are likely aware that, even after closing their API, private companies still use web scrapers as well as researchers. The issue is that now the procedure is made opaque both to researchers and, as a consequence, to research subjects as well. We have taken the measures in our power to secure fairness and transparency as to our research purposes and data management standards.

2. Purpose limitation: Data must be processed for the legitimate purposes specified explicitly to the data subject when collected.

    Our data collection was conducted for public interest purposes, not for profit (Kotsios et al., 2019; p. 9-10), as is the case of private companies that also use web scrapers. The data we collected is going to be used solely for this research purpose and with our previously determined research question.

3. Data minimization: Only as much data as absolutely necessary for the purposes specified must be collected and processed.

We had a defined timeframe for data collection (from December 2020 to January 2021) and we collected strictly data that was needed to answer our previously determined research question.

4. Accuracy: Personal data must be kept accurate and up to date.

Once the collection timeframe was closed, the texts of collected posts were not edited content-wise. For the topic modeling analysis, we deleted stop-words (e.g., pronouns and conjunctions), diacritics (e.g., the letter "ç" or "ã"), and converted typical internet shortcut words into their traditional format (e.g., in Portuguese "as well" means "também" and is often written in online interactions as "tb" or "tbm"). This manipulation does not change the accuracy of posts' content. Instead, it serves to raise the accuracy of our topic modeling results.

5. Storage limitation: Personally identifying data can only be stored for as long as necessary for the specified purpose.

The data will be destroyed as soon as the research project is finished.

6. Integrity and confidentiality: Processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (e.g., by using encryption).

We have complied with this as described in section Methodological decisions.

7. Accountability: The data controller is responsible for being able to demonstrate GDPR compliance with all of these principles.

As data controllers, we can comply with this measure.

As demonstrated, research based on social media texts can strive to comply with GDPR measures. However, the fact that social media companies like Facebook do not provide transparent information about their algorithm functionality and no longer facilitate data collection for academic research purposes puts researchers in a gray zone in regards to GDPR.

AoIR guidelines

AoIR guidelines (franzke et al., 2020) are based on similar concerns as GDPR's, such as securing data privacy. However, AoIR guidelines are not laws, but rather stances for decision-making recommendations for scientific research.

Instead of providing a panacea, AoIR guidelines emphasize researchers' ability to make sound judgments, which most importantly protect research subjects and researchers themselves (p. 23). These were the main guidelines we followed in our decision-making described at the beginning of the paper.

There is a clash: we are complying with AoIR, and we are in a gray zone of the ToS and GDPR. Our compliance rationale for these three constraints in information flow (ToS, GDPR, and AoIR) is aligned with the conclusions reached by Mancosu and Vegetti (2020), who claim that collecting textual data from Facebook pages can be "ethically and legally (GDPR) acceptable" (p. 9) but it might be in conflict with Facebook ToS.

Detail the entrenched information norms

Such norms "describe the existing practices that prevail in a given context, encompassing the flows of information, transmission principles, and expectations of the actors involved" (Zimmer, 2018, p. 8). In our context, there are three groups of actors involved: migrants and aspiring migrants who participate in the groups, the group administrator(s), and the researchers. Because the interests of a company are divergent from those of these actors, Facebook is not accounted for here. Its entrenched information norms-related expectations can be interpreted according to the ToS described in the section Facebook ToS.

Migrants and aspiring migrants' expectations

As highlighted in section Methodological decisions, migrants have different reasons to join social media platforms and exchange information on these platforms. The entrenched information norms they have, however, are likely to be similar, namely that other humans will read what they have posted on the groups. There is an expectation that these other humans probably share similarities with them: be Brazilian migrants or aspiring migrants in Germany, have some relationship with Brazil and/or with Germany. There is also an expectation that the questions and other shared information on these groups will be replied to by these other humans who are likely to hold valuable information that can help solve the issue being asked about. As these groups are formed by over a thousand participants, it is not expected that all participants see the messages and reply to them, as well as there is an understanding that there are participants who are lurking in these groups (i.e., group participants who read the interactions but do not write). As these groups are highly populated and administrators cannot guarantee the identity of those who access groups, participants are likely to be careful with sharing personal information, and it is not uncommon that migrants anonymize themselves by not using

their real names on social media. Finally, some public groups even accept the presence of company profiles that promote their services.

Group administrator(s)

Group administrators of public groups are likely to hold fewer entrenched information norms than administrators of private groups. The former is probably interested in having fewer moderation duties and possibly being recognized as the administrator of a large group highly relevant for the information exchange of a specific population (e.g., for the case of migrants in a specific town or migrants looking for education and employment in specific areas). For these administrators, which users participate in the group and for what purposes is probably irrelevant, as long as participants comply with their rules. The latter type of administrator is probably interested in having more control over who can access the group. Based on the questions from entry forms, their expectation is to filter participants who are likely to fuel disrupting discourses and those interested in using the groups to sell products and services. These administrators were informed about our interests as researchers.

Researchers

Our expectation was to observe migrants and aspiring migrants' textual interactions in a non-controlled situation. In doing so, we wanted to analyze the role of information exchanges in migratory projects related to education—or, how education projects relate to migration. More concretely, we expected to understand what migrants and aspiring migrants debate about education in these groups and to determine the relevance of education-related topics in these information exchanges. From the perspective of other group participants, as the researcher who owns the profile did not interact in the groups, she could be interpreted as a lurker.

As the author who created the profile is clearly identified on Facebook and as we did not harvest information such as location or other sensitive information that users may have made available on their personal Facebook profiles themselves, we did not disrupt users' expectation of being in a group with other people they do not know. In the eyes of these participants, we as researchers could be seen as any other lurker. As we did not expect to breach anonymity or to promote services or products, nor to do harm to participants, we complied with entrenched information norms of group administrators—and with ours as researchers. Furthermore, administrators of closed groups were explicitly informed about our expectations. Our results will also be shared with them in order to reassure them that we have secured that no group participant can be de-anonymized. Finally, as we are not making our data publicly

available, we are not disrupting the informational norms of any actor.

Prima facie assessment

Nissenbaum (2010, p. 182) contends that "a breach of informational norms yields a prima facie judgment that contextual integrity has been violated because presumption favors the entrenched practice". Here, we land in a gray area. Considering that these groups are highly populated and therefore participants are careful with the information they publish in the group, that private groups' administrators were informed and allowed us to participate, and that we did not go against the expectations of public groups' administrators, we could argue that no informational norm was breached. Nevertheless, if we consider that the expectations of group participants were to exchange information, not to participate in a research project, then an informational norm was breached, particularly in relation to group participants and public groups' administrators. Still, our academic publication opens another information flow, because we have processed and analyzed the data, and therefore we are not sharing data that is part of the information flow described in part 4.1. In that sense, because there are two information circuits (the one among Facebook users only and the one derived from the publication of a paper based on the Facebook posts), and the data of one is not shared with the other, the situation is more complex and the entrenched information norms from the second circuit should also be assessed. Shortly, in that second circuit, peer-reviewers and the academic readership would probably like to have access to the data from the first circuit in order to assess the reliability of our analysis. However, if we do that, we would merge the two circuits of information and then doubtlessly breach contextual integrity by making our data public.

Evaluation I

Considering that gray zone in relation to migrants' privacy and the breach of Facebook's ToS, we assume there is potential for a violation of contextual integrity and therefore, proceed to the first evaluation step to assess the gravity of the potential violation (Nissenbaum, 2010, p. 182).

Based on previous studies investigating migrants' use of social media (e.g., Dekker et al., 2018; Jayadeva, 2020), there is no evidence that such a topic of investigation might have caused harm to migrants. Researchers in this field have followed ethical procedures of anonymization and their research questions do not put the groups researched by them under any particular doubt or surveillance from authorities or other actors of migration. In that sense, there is no evidence that academic research about migration and

digital media use has ever caused migrants to loose control over their information. Similarly, our proposal accounts for such security measures. In this sense, we are "doing no harm", a primary ethical imperative (Fuchs and Unterberger, 2021).

Evaluation II

The second evaluative step asks to assess how the new practice directly impinges on the values, goals, and ends of the particular context (Nissenbaum, 2010, p. 182). If we consider that, although one of us was participating in the groups not for the interest of exchanging information, but rather in analyzing it, then one could argue that we are not aligned with the goals of the context. However, as the person participating in the groups was completely identified, we treated the data carefully, we did not interfere in any discussions in the group, we did not collect private information, and users are not naive about participating in a group with over thousand unknown people, hence not sharing sensitive information and sometimes anonymizing themselves with aliases, one cannot say that we bluntly disrupted the values of the groups or of participants we have researched.

A model to assess factors affecting consent suggested by McKee and Porter (quoted in Elgesem, 2015, p. 15–16) and adapted to research with social media data by Elgesem (2015) helps to think about consent and anonymity within this gray area. What is helpful in that model is that it is based on scales, not on absolute statements. Assessing these scales can inform whether there is a requirement to obtain consent or if it is important to have consideration for consent (Elgesem, 2015; p. 18–19) and for the impossibility of obtaining it. The scales account for whether the data is rather private or rather public, and whether there are rather high or rather low issues involving topic sensitivity, degree of interaction of the researcher with subjects, and degree of vulnerability of subjects. For our case, the data collected are not personal communications between a small group of people, yet it was posted in a specific group of Brazilian migrants to which we had access, hence on this scale, we would still be in a gray area. Regarding topic sensitivity, we have a clear research question focused on transnational educational projects, which is not a topic of particular concern especially because we are not closely interacting with the researched subjects. Although remembrance of the experience of shattered educational aspirations and projects, for instance, can cause distress, our empirical research question focuses on what general topics are discussed in these groups. Furthermore, as we are not requesting group participants to access memories or share plans with us, i.e., we have a very low level of interaction with subjects, both our topic and our degree of interaction imply a comparatively low requirement of consent. Regarding the last factor, the subjects' vulnerability, we have to consider that we are

researching migrants and aspiring migrants whose legal status is unknown to us. Again, we go back to our research question to judge whether there is a rather high or low requirement for consent. Differently from investigating migrants who use social media to inform their pathway to claim asylum and might have to resort to irregular practices for border crossing (Dekker et al., 2018; Fischer and Jørgensen, 2021), our research question relates to an issue that requires migrants to have a regularized status in the country, as without a residence permit, they cannot enroll in tertiary educational institutions. However, the situation is different for children and teenagers, who can access schools even though their parents might not have a regular migratory status in Germany. The possibility of inflicting direct psychological harm through our research topic is also low, as we did not interact with group participants. The possibility of inflicting indirect harm based on the outcomes of our research is also low, due to the focus of our research question in migratory projects involving education.

Comparative evaluation based on studies using Facebook posts and a topic modeling approach

This section comments upon other empirical research based on a topic modeling approach that also used data collected on Facebook. At the time of writing, to the best of our knowledge, there is no study using migrants' posts based on such an approach. The focus of this section is on other researchers' solutions and ethical justifications for data collection on Facebook regardless of the empirical topic of the studies. This overview reveals researchers' concern about the ethics of collecting such data but also an apparent avoidance to discuss these concerns in depth, perhaps either due to the earlier facilitated access to collecting Facebook posts (before the API's closure) or due to implicit perceptions of what public data is.

The discussion presented in this paper could have been spared if we had followed a less troublesome approach to data collection. An option could have been recruiting migrant Facebook users to participate in the research and requesting them to sign a consent form, as Verheijen and Stoop did for their linguistic analysis among Dutch speakers (Verheijen and Stoop, 2016, p. 249–258). They analyzed posts made only by these subjects who explicitly consented to have their posts collected. That would hardly be an option to research migrant Facebook groups. The reason for that is twofold: first, we could not force participants to post in the migrant Facebook groups, hence, if a participant did not post at all, we would have no data; and second, had we recruited participants with a high rate of posts, we would be cherry-picking the data since there is no evidence that most group participants have a high posting rate. Furthermore, even if we maximized or minimized demographic differences of such hypothetical participants, we would still have a non-representative sample because we do not know exactly what are the socio-economic characteristics of

regular participants of these migrant Facebook groups. Hence, Verheijen and Stoop's (2016) solution would not suit our research aim.

Although also not related to migrants' use of Facebook, other methodological solutions closer to the one described in this paper reveal similar ethical concerns and contend that they cannot guarantee anonymization despite measures taken by the researchers (Merrill and Åkerlund, 2018, p. 340), while others do not focus on discussing the data collection and storage procedures (Puschmann et al., 2020; Amara et al., 2021; Heft et al., 2022). Most of these analyses are based on "public posts" or "publicly available profiles", i.e., comments on Facebook pages of political parties and private organizations made by users who did not restrict who could view their posts. That decision seems to be implicitly presented as an ethical justification for collecting those posts. Furthermore, unlike ours, these other studies were conducted before Facebook closed its API, hence they do not mention the harms of that restriction to research.

The restriction posed by Facebook to social researchers is an obstacle in the analysis of social interactions, their causes, and consequences. Still, researchers keep using that platform and other platforms owned by the same company for their data collection, due to the social relevance it has reached. If earlier research, as described above, mentioned ethical concerns in a few sentences or left these concerns implicit, the closing of access to collect Facebook data has given impulse to reflect on ethics in practices of collecting digital traces (e.g., Bruns, 2019; Puschmann, 2019). That does not mean that the trade of "closing the access to relevant empirical data" for "elaborating on the ethics of collecting that data" was worth it: ethics of research using digital data had been already brought up before the closing of Facebook's API (e.g., Zimmer, 2010) and, as digitalization increases, there is no evidence that the ethical discussion in this field would have stopped. Nevertheless, this situation promotes advancements in the ethics' discussion at the same time that it sheds light on the power that a big-tech company has over academic research and researchers, as researchers might have to consider whether they make themselves liable for prosecution or decide to investigate topics through other methodological approaches even though using Facebook would be relevant.

Final determination

The last point proposed by Nissenbaum (2010, p. 182–183) is the final determination as to whether there was a violation of contextual integrity and, if so, how grave is this violation toward whom, whether and how these violations are defensible. This final point is similar to our research question about how to justify the collection and analysis of migrants' digital traces for academic research.

We could have hired someone to copy and paste all posts and comments from Facebook, thereby complying with the ToS of not using an automated web scraper. The person doing this, however, would have had much more insight into who wrote what than an automated procedure. Facebook's decision of prohibiting web scraper might be well-thought to avoid companies profiling users and tackling the criticism toward the company after the Cambridge Analytica scandal, nevertheless, these policies are harmful to researchers who care for ethics and anonymity—as they can make themselves liable even though they have the best interest of not exposing vulnerable populations.

In this paper, we made transparent our data collection procedure and analyzed it in the light of ethical and legal frameworks. Along with Bruns (2019) and other critical researchers of digital media, we have added up the argument that such social media platforms occupy nowadays an important role in social phenomena and thus must "provide transparent data access to critical, independent, public-interest research" (p. 1561). For researchers studying migration and social media use, the lack of transparency of social media platforms implies a forced lack of control over the collected data. In turn, that impacts also migrants who could profit from critical views about digital media: research in this area can provide insights into reasons to migrate and decision-making processes supported by information exchanges on social media which can inform policies and support arguments in favor of migrants and diversity in media educational institutions, public discourses, and political spheres.

Based on the heuristic described in the previous eight points, the measures we took for the data collection through a topic modeling approach and its subsequent analysis do care for the anonymity of potentially vulnerable group participants. On the one hand, our decision not to make the data collected from the groups freely available further secures anonymity. On another hand, that puts us in a criticisable situation regarding the reliability of our data, as it cannot be shared. However, as securing data protection and anonymity of migrants who participate in these Facebook groups is more important in order to avoid harm, we decided to put more weight on that aspect than on the quality assessment of the academic community. In that sense, to some extent, securing the anonymity and data protection of vulnerable populations in academic research is a group commitment.

The collection of large textual datasets of migrants' digital traces for academic research purposes can be justified when researchers are invested in securing the collected data from anonymity breaches—by not collecting certain profiling data and by not creating another information flow by making their dataset available. The fact that academic research is guided by methodological and ethical guidelines and tends to be detached from financial profit also speaks in favor of the possibility of securing such datasets collected from migrant or vulnerable populations. In the unlikely case researchers had interest in selling their dataset for target advertising or

political action against migration, for instance, the contextual integrity analysis described here would no longer be applicable and multiple contextual integrity violations would have been committed. Finally, although the procedures described here could be interpreted as in a legally gray zone, no involved parts were harmed in this data collection and analysis procedure. Therefore, such research is defensible when an appropriate research question is addressed and standards are followed, as researchers have already been doing (Mahoney et al., 2022a; Sandberg et al., 2022a).

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Acknowledgments

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Amara, A., Hadj Taieb, M. A., and Ben Aouicha, M. (2021). Multilingual topic modeling for tracking COVID-19 trends based on Facebook data analysis. *Appl. Intell.* 51, 3052–3073. doi: 10.1007/s10489-020-02033-3

Brown, B., Weilenmann, A.,MacMillan, D., and Lampinen, A. (2016). "Five provocations for ethical HCI research," in *Proceedings of ACM CHI, 7–12 May, San Jose, CA, USA.*

Bruns, A. (2019). After the 'APIcalypse': social media platforms and their fight against critical scholarly research. *Inf. Commun. Soc.* 22, 1544–1566. doi: 10.1080/1369118X.2019.1637447

Carnicer, J. (2019). Transnational Migration and Educational Opportunities: A Case Study of Migration from Brazil to Germany. *London Rev. Educ.* 17, 14–25. doi: 10.18546/LRE.17.1.02

Dekker, R., Engbersen, G., Klaver, J., and Vonk, H. (2018). *Smart* refugees: how syrian asylum migrants use social media information in migration decision-making. *Soc. Media + Soc.* 4, 2056305118764439. doi: 10.1177/2056305118764439

Elgesem, D. (2015). "Consent and information—ethical considerations when conducting research on social media," in *Internet Research Ethics, eds D. Elgesem, C. Ess, A. O. Larsson, M. Lüders, R. Prabhu, K. U. Segadal, et al. Cappelen Damm Akademisk/NOASP,* 14–34.

Fischer, L., and Jørgensen, M. B. (2021). "Impossible research? ethical challenges in the (digital) study of deportable populations within the European Border Regime," in *Research Methodologies and Ethical Challenges in Digital Migration Studies: Caring For (Big) Data?,* eds M. Sandberg, L. Rossi, V. Galis, and M. Bak Jørgensen, 141–166. London, UK: Palgrave Macmillan.

Foletto, L. (2018). Usos sociais do facebook por migrantes brasileiros na Suécia: os grupos na construção identitária. *Mediação* 20, 91–112.

franzke, A. S., Bechmann, A., Zimmer, M., Ess, C., and Association of Internet Researchers. (2020). Internet Research: Ethical Guidelines 3.0. Available online at: https://aoir.org/reports/ethics3.pdf

Fuchs, C., and Unterberger, K. (2021). *The Public Service Media and Public Service Internet Manifesto.* London: UK: University of Westminster Press. doi: 10.16997/book60

Fürstenau, S. (2019). *Educational Aspirations of Underprivileged Female Migrants. An Ethnographic Case Study of a Transnational Network of Care Workers between Brazil and Germany.* Germany: Zeitschrift Für Erziehungswissenschaft.

Grimmer, J., and Stewart, B. M. (2013). Text as data: the promise and pitfalls of automatic content analysis methods for political texts. *Pol Anal.* 21, 267–297. doi: 10.1093/pan/mps028

Hargittai, E., and Sandvig, C. (2016). "How to think about digital research," in *Digital research confidential: The secrets of studying behavior online*, eds E. Hargittai and C. Sandvig. Cambridge, MA: The MIT Press, 1–28

Haythornthwaite, C. (2002). Strong, weak, and latent ties and the impact of new media. *Inf. Soc.* 18, 385–401. doi: 10.1080/01972240290108195

Heft, A., Pfetsch, B., Voskresenskii, V., and Benert, V. (2022). Transnational issue agendas of the radical right? Parties' Facebook campaign communication in six countries during the 2019 European Parliament election. *Eur. J. Commun.* 026732312211001. doi: 10.1177/02673231221100146

Jayadeva, S. (2020). Keep calm and apply to Germany: how online communities mediate transnational student mobility from India to Germany. *J. Ethnic Migr. Stud.* 46, 2240–2257. doi: 10.1080/1369183X.2019.1643230

Kotsios, A., Magnani, M., Vega, D., Rossi, L., and Shklovski, I. (2019). An analysis of the consequences of the general data protection regulation on social network research. *ACM Transac. Soc. Comput.* 2, 1–22. doi: 10.1145/3365524

Leurs, K. (2014). Digital throwntogetherness: young londoners negotiating urban politics of difference and encounter on facebook. *Popular Commun.* 12, 251–265. doi: 10.1080/15405702.2014.960569

Leurs, K. (2017). Feminist data studies: using digital methods for ethical, reflexive and situated socio-cultural research. *Feminist Rev.* 115, 130–154. doi: 10.1057/s41305-017-0043-1

Mahoney, J., Le Louvier, K., and Lawson, S. (2022a). "The ethics of social media analytics in migration studies," in: *Information and Communications Technology in Support of Migration. Security Informatics and Law Enforcement,* eds B. Akhgar, K. L. Hough, Y. Abdel Samad, P. Saskia Bayerl, A. Karakostas. Berlin, Germany: Springer, Cham.

Mahoney, J., Le Louvier, K., Lawson, S., Bertel, D., and Ambrosetti, E. (2022b). Ethical considerations in social media analytics in the context of migration: lessons learned from a Horizon 2020 project. *Res. Ethics* 18, 226–240. doi: 10.1177/17470161221087542

Mancosu, M., and Vegetti, F. (2020). What you can scrape and what is right to scrape: a proposal for a tool to collect public facebook data. *Social Media + Soc.* 6, 205630512094070. doi: 10.1177/2056305120940703

Merrill, S., and Åkerlund, M. (2018). Standing up for Sweden? The racist discourses, architectures and affordances of an anti-immigration facebook group. *J. Comput. Med. Commun.* 23, 332–353. doi: 10.1093/jcmc/zmy018

Naughton, J. (2022). For the first time in its history, Facebook is in decline. Has the tech giant begun to crumble? *The Guardian*. Available online at: https://www.theguardian.com/commentisfree/2022/feb/06/first-time-history-facebook-decline-has-tech-giant-begun-crumble

Nissenbaum, H. F. (2010). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford Law Books.

Oosterbaan, M. (2010). 'Virtual Migration. Brazilian Diasporic Media and the Reconfigurations of Place and Space'. *Revue Européenne Des Migrations Internationales* 26, 81–102. doi: 10.4000/remi.5037

Oosterbaan, M. (2013). "Chapter two. orkontros: brazilian migrants, social network sites and the european city," in *Contemporary Culture: New Directions in Art and Humanities Research*, eds J. Thissen, R. Zwijnenberg and K. Zijlmans. Amsterdam: Amsterdam University Press, 37–50

Puschmann, C. (2019). An end to the wild west of social media research: a response to Axel Bruns. *Inf. Commun. Soc.* 22, 1582–1589. doi: 10.1080/1369118X.2019.1646300

Puschmann, C., Ausserhofer, J., and Šlerka, J. (2020). Converging on a nativist core? Comparing issues on the Facebook pages of the Pegida movement and the Alternative for Germany. *Eur. J. Commun.* 35, 230–248. doi: 10.1177/0267323120922068

Richter, C., Emmer, M., and Kunst, M. (2018). "Flucht 2.0. Was Geflüchtete wirklich mit ihren Smartphones machen," in *Social Transformations*, *Vol. 2*. Available online at: https://www.socialtrans.de/index.php/st/article/view/20

Roberts, M. E., Stewart, B. M., and Tingley, D. (2019). stm : An R Package for Structural Topic Models. *J. Stat. Soft.* 91. doi: 10.18637/jss.v091.i02

Sandberg, M., Mollerup, N. G., and Rossi, L. (2022a). "Contrapuntal connectedness: analysing relations between social media data and ethnography in digital migration studies," in *Research Methodologies and Ethical Challenges in Digital Migration Studies*, eds M. Sandberg, L. Rossi, V. Galis, M. Bak Jørgensen. London, UK: Palgrave Macmillan, Cham.

Sandberg, M., Rossi, L., Galis, V., and Bak Jørgensen, M. (2022b). *Research Methodologies and Ethical Challenges in Digital Migration Studies*. Berlin, Germany: Springer International Publishing.

Sandvig, C. (2017). *Heading to the Courthouse for Sandvig v. Sessions [web log]*. Available online from: https://socialmediacollective.org/2017/10/19/heading-to-the-courthouse-for-sandvig-v-sessions/ (retrieved April 27, 2022).

Schrooten, M. (2012). Moving ethnography online: researching brazilian migrants' online togetherness. *Ethnic Racial. Stud.* 35, 1794–1809. doi: 10.1080/01419870.2012.659271

Siapera, E., and Creta, S. (2020). "The ethics of media research with refugees," in S. Jeppesen, P. Sartoretto. *Media Activist Research Ethics. Global Transformations in Media and Communication Research—A Palgrave and IAMCR Series*. London, UK: Palgrave Macmillan, Cham.

Verheijen, L., and Stoop, W. (2016). "Collecting facebook posts and whatsapp chats," in *Text, Speech, and Dialogue. TSD 2016. Lecture Notes in Computer Science, vol. 9924*, eds P. Sojka, P, A. Horák, I. Kopeček, K. Pala. Berlin, Germany: Springer, Cham, 249–258.

Waters, J. (2015). Educational imperatives and the compulsion for credentials: family migration and children's education in East Asia. *Child Geogr.* 13, 280–293, doi: 10.1080/14733285.2015.972646

Zimmer, M. (2010). "But the data is already public": on the ethics of research in Facebook. *Ethics Inf. Technol.* 12, 313–325. doi: 10.1007/s10676-010-9227-5

Zimmer, M. (2018). Addressing conceptual gaps in big data research ethics: an application of contextual integrity. *Social Media + Soc.* 4, 205630511876830. doi: 10.1177/2056305118768300

Transnational death and technological haunting

Natalia Sánchez-Querubín*

University of Amsterdam, Media Studies, Amsterdam, Netherlands

The paper proposes "technological haunting" as a concept in migration and transnational death studies. Existing theory and empirical work in media studies explore connections between new media and grieving practices and how affordances of co-presence and portability shape how people maintain bonds with the deceased. The unique considerations that "technological haunting" brings to the study of transnational families and death still need to be addressed by both media scholars and researchers within digital migration studies.

KEYWORDS

transnational families, transnational death, online media, mourning (bereavement), online funerals

Introduction

Transnational families maintain bonds and practice care while living separated by significant distances during prolonged periods (Baldassar, 2014). The "mediated co-presence" (Madianou and Miller, 2012) and portability afforded by new media are essential for how transnational families maintain relationships, considering that face-to-face interactions are limited (Leurs, 2014; Francisco, 2015; Alinejad, 2019; Wilding et al., 2020; Abel et al., 2021; Hillyer, 2021; Tariq et al., 2022). Digital media are also important for practicing intergenerational care and death, which are sub-areas in transnational family research (Blouin et al., 2022). They explore, for example, caregiving between aging parents in the home country and children who live abroad and the need for more compassionate policies and perspectives that can account for aging migrants (Ciobanu and Hunter, 2017), such as the role of distant support networks and media literacy (Wilding and Baldassar, 2018).

Online rituals and communication during times of crisis are also especially significant when members of transnational families have limited resources or work leave (Giralt, 2019), unstable migration statuses (Bravo, 2017), and are subjected to border regimes, including border closure during the Covid-19 pandemic, all of which shape how individuals may care for their loved ones and be present at the end of their lives (Alexis-Martin, 2020; Hinkson et al., 2022). Research also finds that media affordances like video streaming do not alleviate the need for physical presence but can help people cope with the distance during a crisis (Baldassar, 2014; Bravo, 2017; Le Gall and Rachédi, 2019; Brandhorst et al., 2020). While this body of literature is rich and extensive, there is still a need for more research on grief experiences as "less is known about processes of transnational grieving and their impacts on migrants and their family relationships after the bereavement has occurred" (Giralt, 2019; p. 578). This paper invites considering how the concepts of "digital remains" (Lingel, 2013; Wright, 2014; Morse and Birnhack, 2022) and "haunting" can contribute to this aim by potentially informing studies of grief in transnational families.

Research on grieving, amongst other aspects, investigates how people continue bonds with the deceased (Gibson, 2015; Walter, 2015; Arnold et al., 2017; Wagner, 2018; De Vries, 2019; van der Beek et al., 2019; Eriksson Krutrök, 2021). The theory of continuing bonds

emphasizes that the bereaved sustain a relationship with deceased loved ones by talking to them, writing letters, praying, sharing memories, conserving objects (Root and Exline, 2014), visiting online cemeteries (De Vries and Rutherford, 2004), and interacting with digital remains, which are a deceased person's personal data in social media accounts, text messages in phones, digital photos in cloud services, and voice messages in WhatsApp (Cumiskey and Hjorth, 2017). Media studies scholarship investigates people's use of online media for continuing bonds with the dead while also considering how the media affordances and infrastructures of apps and commercial platforms shape these practices (Gibbs et al., 2014, 2015; Lagerkvist and Andersson, 2017; Thimm and Nehls, 2017; Leaver and Highfield, 2018). Indeed, digital remains are framed by debates about their place in contemporary understandings of grief, potential misuse by media companies, and memorialization in participatory media spaces, often suggesting "that this tension between this anticipated decoupling between the body and data is indeed a source of anxiety in our lives" (Graham et al., 2013; p. 134). The concept and metaphor of *haunting* emerges in cultural and media studies for capturing the sense of nostalgia and connection, as well as ambivalence and anxiety associated with digital remains. As media researchers and philosophers, Lagerkvist and Andersson (2017) argue, "studying death online both enables and requires a re-conceptualization of our culture of connectivity as an existential and ambivalent terrain" (2017; p.551). After all, the media we use to maintain mediated co-presence and "lifelines" (Lagerkvist and Andersson, 2017) with people we love, become also archives of traces left behind after death, and in this way also, a "media of absence" in need of managing (Lagerkvist, 2019a; p. 190).

The subject of digital remains and thus questions about their status, ritualization, and management are absent in studies about how transnational families experience grief. It is true that families and friends living in proximity also deal with digital remains; however, bringing the subject to transnational death may illuminate grief behaviors and anxieties – *forms of haunting*— informed by migrant experiences. After all, migration studies teach us that distance and mobility are aspects of kin work and that "transnational bereavement and grieving relate to the ever-changing emotional geographies of migration and transnational families" (Giralt, 2019; p. 578). Living apart together in a networked and data-intensive world also means dying in it.

To summarize, this paper focuses on a potential area of inquiry: how may "digital remains" and "haunting" concepts inform research in transnational families and grief? The paper is a conceptual piece, meaning it does not include original research based on data. Instead, it approaches the question by discussing literature and presenting ideas for future research. The structure is the following: first, I review transnational care and death literature. Afterward, I introduce digital remains and haunting as concepts developed in media studies. The last section brings both areas together to suggest future research directions about the grief experiences of transnational families.

Transnational care and media

Transnational families are "families that live some or most of the time separated from each other yet hold together and create something that can be seen as a feeling of collective welfare and unity, namely familyhood, even across national borders" (Bryceson and Vuorela, 2020; p. 18). Research into transnational family approaches care as a complex phenomenon that connects affect with material and media practices (Tronto, 1993) and with the intention of "at least removing the assumption that distance is implicitly a barrier to care exchange" (Baldassar, 2016; p. 161). Transnational family studies, in fact, tend to contest "very powerful normative notions of care and intimacy as inherently proximate forms of relatedness and of our taken for granted assumptions about emotional closeness and 'being there' requiring physical co-presence" (Baldassar, 2016; p. 146).

Whereas in the past, "transnational family members would stay in touch only by long-awaited letters that traveled by sea, today, people can be virtually constantly present in each other's lives" (Baldassar et al., 2016; p. 477). For example, people use videoconferencing (e.g., Skype and Zoom) to have dinner together and celebrate birthdays, enabling "the ability of family members to be co-present (emotionally 'there' for each other) across distance" (Baldassar, 2016; p. 145). People also leave video calls open for several hours so that relatives may have a window into their everyday lives (Neustaedter et al., 2015) and mothers play "hide-and-seek with their children (with the help of an adult who moves the laptop around the house to find the children in their hiding places)" (Madianou and Miller, 2012; p. 71). Likewise, exchanging text, photos, videos, and memes in family group chats is "used by long-distance families to affirm their relationships" (Abel et al., 2021; p. 643). These chats create "ambient" and "connected" presence and phatic communication, namely, it is "the repetition, rather than the content of these short messages that maintain the relationship tie and form a sense of connectedness" (Vetere et al., 2009; p. 179). The "portability of care" (Huang et al., 2012; p. 131) and these forms of anytime/anywhere connection "have become woven into the rhythms of family life" (Clark, 2012; p. 202).

Aging and end of life care in transnational families

Managing aging kin, illness, and the end of life represent challenges for people separated geographically. For example, Saramo (2019) frames death as a moment of family rupture, when "we joined countless other families, today and in centuries past, in the processes and emotions of transnational death. Such intimate negotiations, hinged on individual deaths, collectively shape, and reshape identities, traditions, symbols, and cultural borders" (p. 8). Transnational family studies also investigate "death work" performed at a distance: "transnational families "do" kinship on an everyday basis through acts of support and negotiations that defy distance. Can they also "do" death"? (Matyska, 2019; p. 49).

Parents staying in their home countries while adult children relocate to pursue education and long-term employment and residency abroad is now a common scenario. The serious illness of a parent can then set into motion a distant "crisis" of care. It calls for attending medical appointments, coordinating exams and treatments remotely, arranging for prolonged stays, accompanying the person using various media, hiring caregivers, and even having parents emigrate to live with their children. The responsibility

might be assumed alone or distributed between siblings, relatives, and the local community. Illness and aging reveal aspects of transnational caregiving that "remain hidden during those periods when 'routine' forms of distant care are adequate" (Baldassar, 2014; p. 391). After all, "it is during these 'crisis events' of the family life-course when physical co-presence is most acutely required to deliver hands-on personal care and intimate emotional support to the sick family member. It is also a time when close kin, including those who are living faraway, feel they need 'to be there' for their own sense of well-being" (Baldassar, 2014; p. 394).

Many factors will determine how people can be there for each other during times of crisis. Family members with low digital literacy will likely struggle to use apps and videoconferencing platforms to connect with relatives (Baldassar, 2014). Those with dual citizenship, stable migration status, and generous work leaves will be able to move and engage in caregiving more freely. The accounts from undocumented Latino migrants living in the United States collected by Bravo (2017) describe the opposite, namely, scenarios where people "had no other choice than to have a 'virtual' presence at home, rather than a physical one, when a person in the family gets terminally ill or dies at home" (Bravo, 2017; p. 36). It is not uncommon for members of transnational families to express feelings of guilt and anxiety toward their obligations, which are exacerbated by border regimes—"access to mobility is one of the sharpest stratifying dimensions of our age" (Brandhorst et al., 2020; p. 262). For example, one of the participants in Bravo's study, who maintained constant communication with his father during the period leading to his death, concludes: "I helped in my own way, sending money and making sure our relatives were with him all along the way. I also sent money for his funeral. Still, I never felt so defeated in my whole life, but I had no option" (Bravo, 2017; p. 39). Digital technologies, particularly webcams, "facilitate visual interaction with the dying person anywhere in the world" (Le Gall and Rachédi, 2019; p. 66). However, hearing, touching, and physically interacting with the dying person "allows for a form of intimacy that is not attainable through long-distance communication" (Le Gall and Rachédi, 2019; p. 67).

Transnational death: online funerary rituals

A person's passing is followed by a period when both professionals (e.g., funeral directors) and kin engage in death work. The term "death work" helps to draw "attention to the role of death in the making of transnational families and to stress that death, similarly to kinship, is work rather than a biologically determined phenomenon, enacted through mutual agency and effort by the dying and the survivors, who as transnational family members simultaneously do transnational kinship by doing death" (Matyska, 2019; p. 49). Transnational death challenges the performance of such work and "raises questions about identity, belonging, and customs, but also about the logistical care of bodies, rituals, and commemoration" (Saramo, 2019; p. 8). As with distant intergenerational care, "people fulfill different necessary roles according to their abilities and where they are located" (Saramo, 2019; p. 14).

Funerals help people grieve and process a loss. Having a funeral can also be part of the deceased person's vision of a good death. Migrant communities have historically set up mutual aid funds to secure proper burials aligned with their traditions so that they do not become "some foreigner hastily buried" in a shallow, unmarked grave (Saramo, 2019; p. 11). Funerals are also arranged in the host country through specialist religious organizations, while repatriation is a priority for some communities. For example, Nunez and Wheeler (2012) write about the importance of dying and resting on one's native land for migrants from African countries residing in South Africa. Johannesburg-based organizations and repatriating funeral parlors serve these groups by navigating the social, governmental, "and spiritual channels necessary for determining the course of a deceased migrant body" (Nunez and Wheeler, 2012; p. 212). As repatriation is lengthy, funerary parlors' refrigeration technologies are a selling point.

Family and friends living abroad may return to their home country for a funeral. In addition to the *in-situ* funeral, online memorial pages, digital guest books, and other online rituals complement the grieving process. A person "may visit and interact with an online memorial site for a loved one over many years, but this does not mean she will not also physically tend to the body of the deceased and its final resting place, and the material belongings and meaningful memorabilia left behind" (van Ryn et al., 2017; p. 114). The statement mentioned above is, however, not applicable to all cases. Studies in transnational death include many instances when mediated co-presence is the only option. The (im)mobility regimes related to immigration status, visas, expensive airfares, and latest the covid pandemic prevent people from gathering. Covid-19, one may argue, made distant grieving into a mainstream issue.

Le Gall and Rachédi (2019) explore how missing a relative's funeral affects people deeply. Between 2013 to 2015, they interviewed people who participated in funerary rituals through Skype. A Mexican woman recounts: "I was in contact during Mass, the burial, and the prayers. My sister and a cousin helped me with Skype. I would also call, and I heard everything. There were a lot of people. I wouldn't talk. I would just accompany them via the internet" (Le Gall and Rachédi, 2019; p. 69). Similarly, an Indian woman living abroad could not travel to her father's funeral. Videoconferencing enabled her to be present: "We were on Skype and whatever was going on—I was there. The whole night, sitting online, praying, and seeing my daddy until the last moment when they took him away. So, I felt that I was there with him all the time" (Nesteruk, 2018; p. 1021). These subjects wanted to be present but could not.

Hundreds of people die crossing borderlands, such as the ones between Mexico and the United States and between the Mediterranean Ocean and Europe (Cuttitta and Last, 2020). The post-mortem management of border deaths concerns "counting, mourning, and engaging dead bodies" by governmental institutions and security agencies (Cuttitta and Last, 2020; p. 12). Regarding the Mediterranean route, "most of the bodies disappear into the sea or are buried in anonymous graves at cemeteries on either side of the European Union (EU) border" (Horsti, 2019; p. 672). For every corpse washed ashore, "there is a family living with ambiguity, not knowing if their loved one is dead or alive" (Kovras and Robins, 2016; p. 41).

Researchers, activists, and humanitarian organizations have critiqued the EU's "inability to respond ethically to migrant death as a social loss worthy of common grief" (Horsti, 2019; p. 672). In turn, M'charek et al., 2020 argue for treating migrant bodies as "matters of care," through proper documentation and memorialization. A more caring treatment of border deaths may include more accessible online databases to help with body identification and offer narrative context to statistics. Organizations also engage in grief activism (M'charek et al., 2020) and grassroots forensics (Schwartz-Marin et al., 2016). An example is the Centre for Political Beauty (CPB), a Berlin-based art collective that staged a symbolic burial for refugees in front of the German Parliament and, with the families' permission, exhumed deceased migrants buried on the shores of Italy and transported them to Berlin to provide a dignified funeral. Photographs and social media can also support grieving. For example, Horsti (2019) writes about how digital photographs of burials in Greece taken by relatives (for example, people already living in Europe or who survived the crossing) help those back in the country of origin to grieve. These "online memorials are accessible across borders and in some ways stand in for the victims' unknown or inaccessible graves" (Horsti, 2019; p. 672). The photographs of the rituals "documented the act of mourning for those who were not present at the time" (Horsti, 2019; p. 672).

The sections above only offer a glimpse into the complex subject of intergenerational care and death within transnational families. While I have used elderly parents and migrant children as examples, the situation extends to illness in other family members and different configurations. Additional issues are the "reduced mobility that often accompanies bodily aging" (Brandhorst et al., 2020; p. 265; Askola, 2016) and the restrictive migration policy that sees older people as an economic burden. Furthermore, retirees establish themselves in countries where their pensions afford them a better lifestyle. It means also "new challenges such as the distance from family and an increasing insecurity about legal status, access to public health and aged care services" (Brandhorst et al., 2020; p. 266). Aging in exile requires society's understanding of what it means to age in one's adoptive country (Nesteruk, 2018; p. 1015). Ultimately, distant mourning can be the product of (in)immobility regimes and thus a concern for mobility justice and grief justice. Mobility justice "recognizes that while mobility is a fundamental right for everyone, it is experienced unequally along the lines of gender, class, ethnicity, race, religion, age, and able-bodiedness" (Brandhorst et al., 2020; p. 262). The capacity for caring and grieving is framed by devices such as work leave and remote work policy. For example, in her study of Australian transnational families, Nesteruk (2018) remarks how "the infrastructure of employment and bereavement leave policy is designed for limited periods of grief recovery, and a fast return to productive activities both at work and at home is expected" (Nesteruk, 2018; p. 1021). Family reunification visas for the elderly are complex, airline fares are costly, and visa procedures are lengthy and harrowing. There is a need for more compassionate and structural undertakings of the needs of transnational families.

Access to technology and digital literacy in the face of death and care become issues of justice. Indeed, there are "inequalities in people's ability to afford travel and access ICTs, which raises the issue of whether these should be considered as new kinds of 'human rights and civic policy issues: the right to have contact with faraway kin and the right to the technologies which facilitate distant care" (Baldassar, 2014; p. 395). Access to co-presence, with good quality image and audio and digital death work, becomes a site of intervention, especially outside commercial funerary enterprises.

Digital remains and technological haunting

In the sections above, I reviewed literature attending mainly to issues of mediated distant care, crisis management, and funerary rituals in transnational family settings. This body of work can be further enriched by focusing on an additional aspect of people's grief practices, namely, digital remains and issues emerging around them (Maciel and Pereira, 2013; Stokes, 2015). Digital remains are defined as "orphaned data whose creator and owner is now dead" (Morse and Birnhack, 2022; p. 1344), "online content on dead users" (Lingel, 2013; p. 191), and "digital traces that will remain even after we die" (Wright, 2014;). They include people's social media accounts, devices like laptops and mobile phones, photos and texts stored in them, and digital footprints in channels like WhatsApp. Questions about digital remains involved their status in terms of value, privacy, and ownership (Maciel and Pereira, 2013; Stokes, 2020; Morse and Birnhack, 2022) and about deciding how to sort, delete, store, share, memorialize the personal data of someone who has passed away (Gach and Brubaker, 2021). For example, Facebook and Instagram have protocols for dealing with the public accounts of people who have died. Family can contact the company, and after submitting proof of death, they can memorialize the account, thus saving it for posterity, or shut it down. Studies describe memorialized accounts in terms of how they function as "biographical objects" (Ebert, 2014; p. 35) suited for the preservation of bonds and as sites for "performative displays of mourning" that allow wider audiences to pay respect to remains and maintain bonds with the dead (Marwick and Ellison, 2012. p.378). In their study of digital remains, Gray and Coulton (2013) remark that "as an immaterial and immanent form, the dead can effectively, but not formally, exist" (p.37), thus we find them in acts of recall and in a "broader range of connections between the senses, agencies, memory and history that is enmeshed through our emotional and aesthetic experiences" (p.37). Social media is also part of the landscape where these emotional and esthetic experiences occur (Walter et al., 2012; Brubaker et al., 2013; Gotved, 2014; Refslund Christensen and Sandvik, 2015).

The dead are "a continuing co-presence on social media platforms" (Eriksson Krutrök, 2021; p. 1; Leaver, 2019). Terms like "ghost in the machine" (Stokes, 2012; Wortham, 2015; Kasket, 2019; Pasquali et al., 2022) and "digital specters" (Cumiskey and Hjorth, 2017), frequently used in news and academic media, capture the haunting nature of these encounters. Ideas about "haunting" also help examine the ambivalence of public digital remains like social media accounts. Haunting evokes attraction and nostalgia, albeit accompanied by uncertainty and anxiety (Sconce, 2000; Blanco and Peeren, 2013). Moreover, "haunting" speaks of presence and activity that lingers beyond a clear source, which may bring comfort and pain. Digital remains are intertwined

with interactivity, automation, and popularity metrics of the social media platforms that host them, which also means a lack of control over someone's digital afterlife. On the one hand, social media companies ultimately dictate digital legacy management terms and own the data, which continues to be stored and used in servers. Conservation is not entirely up to users. On the other hand, interaction with the memorialized account, namely, new forms of maintaining bonds with the deceased, can be unsettling. People post public content speaking directly at the deceased person rather than about them, mimicking ongoing conversations. For example, a woman describes being uncomfortable when content from her deceased relative shows up on her social media feed; to avoid it, she tweaked the settings but found de-friending the account to be "an act she could not bring herself to do, though she found the page hard to visit" (Pennington, 2013; p. 625). Quantification is also a potentially painful aspect of grief online, when for instance, a person sees that others no longer interact with the memorial account they manage—"how do we understand the expressed hurt some bereaved feel when not receiving enough likes or visitors at the site of their commemoration?" (Lagerkvist, 2019b; p. 15). Unsettling activity also includes someone logging in as the deceased and impersonating them and Facebook recommending users to interact with a diseased person's profile (Zaveri, 2019). Questions about post-death activity will continue growing as the memorialization industries innovate by offering bots, holograms, and avatars trained on a diseased person's online data to recreate their communication styles (Jiménez-Alonso et al., 2022). In these products it "is not simply the presence of the deceased that causes anxiety, but the supposed fullness of that presence, formed by near-totalized recording, networked and beyond the control of the user" (Bollmer, 2013; p. 145).

In addition to public digital remains like social media accounts, people need to deal with digital remains found in private channels and stored in people's devices. Anthropologists like Gibson (2008) have examined how families manage the objects left behind by a loved one. Deciding what to keep, distribute amongst family and friends, sell, or donate is underlined by how grief changes value: "When a loved one dies suddenly their personal belongings and defining possessions come to the foreground of consciousness— they are truly noticed. This noticing is complex and often poignant. Death reconstructs our experience of personal and household objects in particular ways; there is the strangeness of realizing that things have outlived persons" (Gibson, 2008; p. 8). A similar logic can help think about how grief changes people's perception of digital objects, albeit with consideration for their (im)materialities. While it is true that "digital objects of the dead lack the integrity of a physical form and boundary and the kind of relation and intimacy that can be had when this is the case" (Gibson, 2014; p. 234) people store and interact with digital remains and the devices that host them, often in ways that evoke interactivity. For example, Cumiskey and Hjorth (2017) authors of Haunted Hands, explore how people's relationship with their mobile phones changes after losing a loved one. One of the subjects in their ethnographic study keeps a mobile phone with text messages from her parents lost in Typhoon Haiyan, constantly checking old messages they had sent. "She holds the phone as if it contains her parents' spirits" (Cumiskey and Hjorth, 2017; p. 1). Another subject carries her old phone, "refusing to get it upgraded because it has a text messaging from her father who has

passed away" (Cumiskey and Hjorth, 2017; p. 1). Another person continued to pay their brother's phone bill for six months after his death—"as if shutting down the phone would close the last avenue to her brother" (Cumiskey and Hjorth, 2017; p. 2). The phone represented a digital extension of him, "a digital ghost left behind to keep her company, as well as a memorial used to reminisce about him and to revisit shared memories" (Cumiskey and Hjorth, 2017; p. 2). Likewise, a father uses his phone as a tool for ritual by texting his son, who passed away. Moreover, there is uncertainty about how to put digital remains to rest in ways that feel meaningful. For example, this anxiety underlines the artistic research project "Requiem for my mothers' data" (Petrozzi, 2023), which started the death of author's mothers and being left with their devices and data to look after. They describe their project as an inquiry "into techno-capitalism as a new realm in need of ceremony, rituality and care practices evolving around and in it" (Petrozzi, 2023).

A consideration for transnational migration and family studies

In this paper, I have, on the one hand, reviewed literature from death and transnational family studies and, on the other, about the topic of digital remains and memorialization, as developed in media studies. Now, I conclude by combining insights from these two areas to propose directions for future reflection and inquiry; namely, researchers could incorporate considerations about how death transforms people's relationships with media and data into ethnographic and theoretical work on grief in transnational family settings. While these considerations are relevant also for families living in geographical proximity, it is important to include them in transnational family research and explore if specific dynamics or issues emerge due to distance and their communication practices. I propose that research could, more concretely, integrate the following three points.

Memorialization unbound to location

In principle, memorialized online spaces (e.g., social media profiles turned into memorials) complement and expand rather than replace visits to cemeteries and other meaningful locations and contact with other people affected by the loss. However, do limited opportunities to visit graves and other sites of commemoration faced by transnational families affect their perception and use of online memorial spaces? Existing literature describes the importance of online funerals, but do other practices that unbound mourning from location emerge as significant for migrants?

Sorting through objects, devices, and private digital remains

The expectation is that after a person dies, relatives sort through their belongings, including clothes, furniture, and digital remains like mobile phones, digital photographs stored in cloud services, and emails. Accessing and deciding which digital remains to save

is challenging due to the volume of content (e.g., hundreds of digital photos, emails). Researchers could integrate these concerns into investigations about grief by exploring how members of transnational families sort through personal and household items. Do difficulties in accessing locations and transporting physical objects (for example, taking furniture to another country) shape the sorting process? Do digitization practices such as scanning photos and photographing a person's house emerge as a response? Likewise, there are no clear rituals or scripts for disposing of objects such as a deceased person's mobile phones and laptops meaningfully. How do people manage a deceased person's devices and data? Also, are practicalities such as making sure passwords are accessible to others part of how the family prepares for death?

Ritualistic behavior with and through media

A third area of investigation pertains to ritualistic behavior with and through media. Do transnational families develop shared online rituals that help collectively maintain bonds with the disease and grief? An example could be organizing online masses to celebrate anniversaries, sharing photos from the cemetery with those who are not in the home country, and video chatting. Also, texting, sharing images, voice messages, and emailing are lifelines between people living apart and activities that generate large quantities of content and data. Literature on grief and media describes people engaging in ritualized behaviors with digital communication channels, such as texting someone who has died. This behavior is framed as a way for continuing bonds with the deceased. A general inquiry pertains to how participants in family groups deal with these digital remains stored in cloud services and phones and with the absence of one member. Pertaining to communication between two people, what type of emotions do

these communication channels, still visible in a person's device yet inactive, generate? Do people engage in ritualistic behavior like texting a dead person? How is the intensity of engagement with these devices—as most interactions between family members occurred through them rather than in person shaping-people's relationship with them post-mortem?

The paper's goal has not been answering these questions but instead creating a context that invites asking them and, hopefully, has made a case for looking into how people's relationship with media changes after the death of a loved one and if the management of digital remains might have a specific quality and issues for migrants.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abel, S., Machin, T., and Brownlow, C. (2021). Social media, rituals, and long-distance family relationship maintenance: a mixed-methods systematic review. *New Media Soc.* 23, 632–654. doi: 10.1177/1461444820958717

Alexis-Martin, B. (2020). Sensing the deathscape: digital media and death during COVID-19. *J. Environ. Media.* 1, 11–1. doi: 10.1386/jem_00032_1

Alinejad, D. (2019). Careful co-presence: the transnational mediation of emotional intimacy. *Soc. Med. Soc.* 5:2, 2056305119854222. doi: 10.1177/2056305119854222

Arnold, M., Gibbs, M., Kohn, T., Meese, J., and Nansen, B. (2017). *Death and Digital Media*. London: Routledge. doi: 10.4324/9781315688749

Askola, H. (2016). Who will care for grandma? Older women, parent visas, and Australia's migration program. *Au. Feminist Law J.* 42, 297–319. doi: 10.1080/13200968.2016.1258750

Baldassar, L. (2014). 'Too sick to move': distant 'crisis' care in transnational families. *Int. Rev. Sociol.* 24:3, 391–405. doi: 10.1080/03906701.2014.954328

Baldassar, L. (2016). De-demonizing distance in mobile family lives: co-presence, care circulation and polymedia as vibrant matter. *Global Networks*.16, 2.145–163. doi: 10.1111/glob.12109

Baldassar, L., Kilkey, M., Merla, L., and Wilding, R. (2016). "Transnational families, care and wellbeing," in *Handbook of Migration and Health*, ed. Felicity Thomas (Northampton, Massachusetts: Edward Elgar Publishing), p. 477–497. doi: 10.4337/9781784714789.00039

Blanco, M., and Peeren, E. (2013). *The Spectralities Reader: Ghosts and Haunting in Contemporary Cultural Theory*. New York: Bloomsbury Publishing.

Blouin, S., Gerson, S. M., and Cavalli, S. (2022). Assistance in dying across borders: How the transnational circulations of persons, terms and themes influence the construction of a public problem. *Death Stud.* 46, 1557–1568. doi: 10.1080/07481187.2021.1926632

Bollmer, G. D. (2013). 'Millions now living will never die': cultural anxieties about the afterlife of information. *Inform. Soc.* 29, 3.142–151. doi: 10.1080/01972243.2013.777297

Brandhorst, R., Baldassar, L., and Wilding, R. (2020). Introduction to the special issue: 'transnational family care 'on hold'? Intergenerational relationships and obligations in the context of immobility regimes. *J. Intergener. Relatsh.* 18:3, 261–280. doi: 10.1080/15350770.2020.1787035

Bravo, V. (2017). Coping with dying and deaths at home: how undocumented migrants in the United States experience the process of transnational grieving. *Mortality* 22, 33–44. doi: 10.1080/13576275.2016.1192590

Brubaker, J. R., Hayes, G. R., and Dourish, P. (2013). Beyond the grave: Facebook as a site for the expansion of death and mourning. *Inform. Soc.* 29, 152–163. doi: 10.1080/01972243.2013.777300

Bryceson, D., and Vuorela, U. (2020). *The Transnational Family: New European Frontiers and Global Networks*. New York: Routledge. doi: 10.4324/9781003087205

Ciobanu, R. O., and Hunter, A. (2017). Older migrants and (im) mobilities of ageing: an introduction. *Popul. Space Place.* 23, e2075. doi: 10.1002/psp.2075

Clark, L. S. (2012). *The Parent App: Understanding Families in the Digital Age*. New York: Oxford University Press. Available online at: https://academic.oup.com/book/6470

Cumiskey, K. M., and Hjorth, L. (2017). *Haunting Hands: Mobile Media Practices and Loss*. Oxford: Oxford University Press.

Cuttitta, P., and Last, T. (2020). *Border Deaths: Causes, Dynamics, and Consequences of Migration-related Mortality*. Amsterdam: Amsterdam University Press.

De Vries, B. (2019). *Kinship Bereavement in Later Life: A Special Issue of "Omega-Journal of Death and Dying"*. Routledge.

De Vries, B., and Rutherford, J. (2004). Memorializing loved ones on the World Wide Web. *Omega J. Death Dying* 49, 5–26. doi: 10.2190/DR46-RU57-UY6P-NEWM

Ebert, H. (2014). "Profiles of the dead: mourning and memorial on Facebook," in *Digital Death: Mortality and beyond in the Online Age*, ed. C.M. Moreman, and A.D Lewis, (ABC-CLIO), 23-42.

Eriksson Krutrök, M. (2021). Algorithmic closeness in mourning: vernaculars of the hashtag #grief on TikTok. *Soc. Med. Soci.* 7, 396. doi: 10.1177/20563051211042396

Francisco, V. (2015). 'The internet is magic': technology, intimacy and transnational families. *Crit. Sociol.* 41, 173–190. doi: 10.1177/0896920513484602

Gach, K., and Brubaker, J. (2021). Getting your Facebook affairs in order: User expectations in post-mortem profile management. Proc. ACM Hum.-Comput. Interact. 5, CSCW1, Article 174 (April 2021), 29 pages. doi: 10.1145/3449248 (accessed June 14, 2023).

Gibbs, M., Meese, J., Arnold, M., Nansen, B., and Carter, M. (2015). Funeral and Instagram: death, social media, and platform vernacular. *Inform. Commun. Soc.* 18:3, 255–268. doi: 10.1080/1369118X.2014.987152

Gibbs, M., Nansen, B., Carter, M., and Kohn, T. (2014). *Selfies at* Funerals: Remediating Rituals of Mourning. AoIR Selected Papers of Internet Research. Daegu.

Gibson, M. (2008). *Objects of the Dead: Mourning and Memory in Everyday Life*. Melbourne: Melbourne Univ. Publishing.

Gibson, M. (2014). "Digital objects of the dead: negotiating electronic remains," in *The Social Construction of Death: Interdisciplinary Perspectives*, eds L. Van Brussel and N. Carpentier (Palgrave: Houndsmills), 221–238.

Gibson, M. (2015) Automatic and automated mourning: messengers of death and messages from the dead. *Continuum.* 29, 339–353. doi: 10.1080/10304312.2015.1025369

Giralt, R. (2019). Bereavement from afar: transnational grieving and the emotional geographies of migration. *Child. Geograph.* 17, 578–590. doi: 10.1080/14733285.2018.1447087

Gotved, S. (2014). Research review: death online—alive and kicking! *Thanatos.* 3, 1.

Graham, C., Gibbs, M., and Aceti, L. (2013). Introduction to the special issue on the death, afterlife, and immortality of bodies and data. *Inform. Soc.* 29, 133–141. doi: 10.1080/01972243.2013.777296

Gray, S. E., and Coulton, P. (2013). "Living with the Dead: Emergent Post-mortem Digital Curation and Creation Practices," in *Digital Legacy and Interaction. Human–Computer Interaction Series*, ed. C. Maciel and V. Pereira (Springer, Cham).

Hillyer, R. S. (2021). Staying connected: effects of online platforms on transnational family relations and social capital. *Contemp. Japan.* 33, 3–23. doi: 10.1080/18692729.2020.1847389

Hinkson, G. M., Huggins, C. L., and Doyle, M. (2022). Transnational caregiving and grief: an autobiographical case study of loss and love during the Covid-19 pandemic. *OMEGA-J. Death Dying.* 4, 5689. doi: 10.1177/00302228221095689

Horsti, K. (2019). Digital materialities in the diasporic mourning of migrant death. *Eur. J. Commun.* 34, 671–681. doi: 10.1177/0267323119886169

Huang, S., Yeoh, B. S. A., and Toyota, M. (2012). Caring for the elderly: the embodied labor of migrant care workers in Singapore. *Global Networks* 12, 195–215. doi: 10.1111/j.1471-0374.2012.00347.x

Jiménez-Alonso, B., Brescó, de., and Luna, I. (2022). Griefbots. A new way of communicating with the dead? *Integrative Psychol. Behav. Sci.* 3, 1–16. doi: 10.1007/s12124-022-09687-3

Kasket, E. (2019). *All the Ghosts in the Machine: The Digital Afterlife of your Personal Data*. London: Hachette UK.

Kovras, I., and Robins, S. (2016). Death as the border: managing missing migrants and unidentified bodies at the EU's Mediterranean frontier. *Polit. Geography* 55, 40–49. doi: 10.1016/j.polgeo.2016.05.003

Lagerkvist, A. (2019a). *Digital Existence*. Milton: Routledge. doi: 10.4324/9781315107479

Lagerkvist, A. (2019b). "Numerical Being and Non-Being," in *A Networked Self and Birth, Life, Death*, ed. Z. Papacharissi (New York, N.Y: Routledge), 11–34. doi: 10.4324/9781315202129-2

Lagerkvist, A., and Andersson, Y. (2017). 'The grand interruption:' death online and mediated lifelines of shared vulnerability. *Fem. Media Stud.* 17, 550–564. doi: 10.1080/14680777.2017.1326554

Le Gall, J., and Rachédi, L. (2019). "The emotional costs of being unable to attend the funeral of a relative in one's country of origin" in *Transnational Death*, ed. S. Saramo, E. Koskinen-Koivisto, and H. Snellman, (Helsinki: Finnish Literature Society/SKS), 65-84.

Leaver, T. (2019). "Co-Creating Birth and Death on Social Media," in *A Networked Self and Birth, Life, Death*, ed. Z. Papacharissi (New York, N.Y: Routledge), p. 35–49. doi: 10.4324/9781315202129-3

Leaver, T., and Highfield, T. (2018). Visualising the ends of identity: pre-birth and post-death on Instagram. *Inform. Commun. Soci.* 21, 30–45. doi: 10.1080/1369118X.2016.1259343

Leurs, K. (2014). The politics of transnational affective capital: digital connectivity among young Somalis stranded in Ethiopia. *Crossings J. Mig. Cult.* 5, 87–104. doi: 10.1386/cjmc.5.1.87_1

Lingel, J. (2013). The digital remains: Social media and practices of online grief. *Inform. Soc.* 29, 190–195. doi: 10.1080/01972243.2013.777311

Maciel, C., and Pereira, V. C. (2013). *Digital Legacy and Interaction: Post-Mortem Issues*. Cham: Springer. doi: 10.1007/978-3-319-01631-3

Madianou, M., and Miller, D. (2012). *Migration and New Media*. Florence: Routledge. doi: 10.4324/9780203154236

Marwick, A., and Ellison, N.B. (2012). There isn't WIFI in heaven! Negotiating visibility on Facebook memorial pages. *J. Broadcast. Elect. Media.* 56, 378–400. doi: 10.1080/08838151.2012.705197

Matyska, A. (2019). "Doing death kin work in Polish transnational families," in *Transnational Death*, ed. S. Saramo, E. Koskinen-Koivisto, and H. Snellman (Helsinki, Finland: Finnish Literature Society/SKS), 49-64.

M'charek, A., Black, J., Cuttitta, P., and Last, T. (2020). "Engaging Bodies as Matter of Care: Counting and Accounting for Death during Migration," in *Border Deaths* (Amsterdam: Amsterdam University Press), 85-101. doi: 10.2307/j.ctvt 1sgz6.9

Morse, T., and Birnhack, M. (2022). The posthumous privacy paradox: privacy preferences and behavior regarding digital remains. *New Media Soc.* 24:6, 1343–1362. doi: 10.1177/1461444820974955

Nesteruk, O. (2018). Immigrants coping with transnational deaths and bereavement: the influence of migratory loss and anticipatory grief. *Family Process* 57, 1012–1028. doi: 10.1111/famp.12336

Neustaedter, C., Pang, C., Forghani, A., Oduor, E., Hillman, S., Judge, T. K., et al. (2015). Sharing domestic life through long-term video connections. *ACM Transact. Comp. Human Interact.* 22, 1–29. doi: 10.1145/2696869

Nunez, L., and Wheeler, B. (2012). Chronicles of death out of place: management of migrant death in Johannesburg. *African Stud.* 71:2, 212–233. doi: 10.1080/00020184.2012.702966

Pasquali, F., Bartoletti, R., and Giannini, L. (2022). 'You're just playing the victim:' online grieving and the non-use of social media in Italy. *Soc. Med. Soc.* 8, 4. doi: 10.1177/20563051221138757

Pennington, N. (2013). 'You don't de-friend the dead: an analysis of grief communication by college students through Facebook profiles. *Death Stud.* 37:7, 617–635. doi: 10.1080/07481187.2012.673536

Petrozzi, G. (2023). *Requiem for my Mother's Data Project Proposal: Ritual for Performing Grief in the Digital Age*. Mediamatic. Available online at: https://www.mediamatic.net/en/page/386358/requiem-for-my-mother-s-data (accessed June 14, 2023).

Refslund Christensen, D., and Sandvik, K. (2015). Death ends a life not a relationship: timework and ritualizations at Mindet. *Dk. New Rev. Hyperm. Multimed.* 21, 57–71. doi: 10.1080/13614568.2014.983561

Root, B. L., and Exline, J. J. (2014). The role of continuing bonds in coping with grief: overview and future directions. *Death Stud.* 38, 1–8. doi: 10.1080/07481187.2012.712608

Saramo, S. (2019). "Introductory Essay. Making transnational death familiar," in *Transnational Death*, eds. S. Saramo, E. Koskinen-Koivisto, and H. Snellman (Helsinki: Finnish Literature Society/SKS),1-8. doi: 10.21435/sfe.17

Schwartz-Marin, E., Cruz-Santiago, A., and Cruz-Santiago, A. R. L. Y. (2016). Pure corpses, dangerous citizens: transgressing the boundaries between experts and mourners in the search for the disappeared in Mexico. *Social Res.* 83, 483–510. doi: 10.1353/sor.2016.0038

Sconce, J. (2000). *Haunted Media: Electronic Presence from Telegraphy to Television*. London: Duke University Press.

Stokes, P. (2012). Ghosts in the machine: Do the dead live on in facebook? *Philos. Technol.* 25, 363–379. Available online at: https://philpapers.org/rec/STOGIT

Stokes, P. (2015). Deletion as second death: the moral status of digital remains. *Ethics Inf Technol.* 17, 237–248. doi: 10.1007/s10676-015-9379-4

Stokes, P. (2020). Dealing with digital remains. *ADI Policy Briefing Papers, Australia.* 1, 3. [Epub ahead of print].

Tariq, A., Muñoz Sáez, D., and Khan, S. R. (2022). Social media use and family connectedness: a systematic review of quantitative literature. *New Med. Soc.* 24, 815–832. doi: 10.1177/146144482110 16885

Thimm, C., and Nehls, P. (2017). Sharing grief and mourning on Instagram: digital patterns of family memories. *Communications* 42, 327–349. doi: 10.1515/commun-2017-0035

Tronto, J. C. (1993). *Moral Boundaries: A Political Argument for An Ethic of Care*. New York: Routledge.

van der Beek, S., Hoondert, M., Bruin-Mollenhorst, J., Matthee, N., Kreek, d. e., Wenz, M. K., et al. (2019). *Ritual in a Digital Society*. Amsterdam: Institute for Ritual and Liturgical Studies, Protestant Theological University and Institute for Christian Cultural Heritage, University of Groningen.

van Ryn, L., Kohn, T., Nansen, B., Arnold, M., and Gibbs, M. (2017). "Researching Death Online," in *The Routledge Companion to Digital Ethnography*, eds. L. Hjorth, H. Horst, A. Galloway, and G. Bell, (New York, N.Y: Routledge), 112-120.

Vetere, F., Smith, J., and Gibbs, M. (2009). "Phatic interactions: being aware and feeling connected," in Awareness Systems (London, U.K: Springer), p. 173-186. doi: 10.1007/978-1-84882-477-5_7

Wagner, A. J. (2018). Do not click "like" when somebody has died: the role of norms for mourning practices in social media. *Social Med. Soc*. 4, 1. doi: 10.1177/2056305117744392

Walter, T. (2015). Communication media and the dead: from the stone age to Facebook. *Mortality*. 20, 215–232. doi: 10.1080/13576275.2014.993598

Walter, T., Hourizi, R., Moncur, W., and Pitsillides, S. (2012). Does the Internet change how we die and mourn? Overview and analysis. *Omega: J. Death Dying* 64, 275–302. doi: 10.2190/OM.64.4.a

Wilding, R., and Baldassar, L. (2018). Ageing, migration and new media: The significance of transnational care. *J. Sociol*. 54, 226–235. doi: 10.1177/1440783318766168

Wilding, R., Baldassar, L., Gamage, S., Worrell, S., and Mohamud, S. (2020). Digital media and the affective economies of transnational families. *Int. J. Cult. Stud*. 23, 639–655. doi: 10.1177/1367877920920278

Wortham, J. (2015). Ghosts in the machine. Social media has changed the way we mourn, for the Better, *The New York Times*, 23 December. Available online at: https://www.nytimes.com/2015/12/27/magazine/ghosts-in-the-machine.html (accessed April 20, 2023).

Wright, N. (2014). Death and the internet: the implications of the digital afterlife. *First Monday* 19, 6. doi: 10.5210/fm.v19i6.4998

Zaveri, M. (2019). 'R.I.P. to a startling Facebook feature: reminders of a dead friend's birthdays. *The New York Times*, 10 April. Available online at: https://www.nytimes.com/2019/04/10/technology/facebook-dead-users-happy-birthday.html (accessed April 20, 2023).

Narratives of Italian Transatlantic (re)migration, 1897–1936

Lorella Viola*

Luxembourg Centre for Contemporary and Digital History, University of Luxembourg, Esch-sur-Alzette, Luxembourg

Remigration is typically envisioned as the final stage of the migration experience, a one-way movement from the host country to the country of origin. This article offers a novel, intimate view of historical return migration as a more complex and discursive process. The case study is Italian American migrants at the turn of the twentieth century, one of the groups which – according to historical statistics – was most actively engaged in Transatlantic remigration; more recent readings, however, show that many of these returnees eventually re-emigrated to the US. Using for the first time immigrant newspapers against the baseline of the Italian public discourse, the article analyzes Italian migrants' own accounts of remigration as a way to access the more subjective dimension of migration. The integration of text mining and Critical Discourse Analysis will show that migrants were experiencing migration as a sense of identity crisis manifested through feelings of being misunderstood, rejected and unappreciated. These results indicate a less material reading of (re)migration, that is beyond economic reasons, and that for many individuals remigration was a bi-directional movement, only fully concluded when they were no longer experiencing a sense of identity crisis, be it in their homeland or the host society. The article will argue that this was the visible outward sign of a much more profound issue: the Italian Government's view of (r)emigration −mainly through the lens of domestic economic advantage − deeply underestimated the complexity of migration as a social phenomenon and as a profoundly changing psychological experience. In the long run, this error of judgment deeply damaged Italy as many of those *ritornati* felt misunderstood and disillusioned and crossed the Atlantic again, this time never to return.

KEYWORDS

remigration, Italian American migration, ethnic press, discoursehistorical analysis, digital humanities, identity

Introduction

Since the early mass migration movements of the past centuries, return migration has always represented a rather complex and critical aspect. Although generally less researched than other topics, studies on the subject has on the one hand primarily revolved around the *impact* of return migration, that is around the one question of whether return migration is beneficial or detrimental, especially economically but also socially, particularly for the sending countries (Saloutos, 1956; Boyd Betty, 1973; Bohning, 1975; Böhning, 1975; Rosoli, 1977; Musillo, 1981; Gentileschi and Simoncelli, 1983; Wyman, 1993; Bonifazi and Heins, 1996). On the other, research has investigated the reasons for returning, positioning the discussion along the conceptual binary of integration/failure to integrate in the receiving country (Cerase, 1967, 1974; Rogers, 1983; Kubát and Center for Migration Studies, 1984). Within the first orientation, a pioneering book is *They remembered America* Saloutos (1956) which investigates the impact of repatriated Greek Americans onto Greek society. From the analysis of two-hundred interviews, the author explores both the reasons behind the repatriation and the potential larger significance

for Greek society. Although it may be difficult to generalize conclusions due to the small sample and the highly qualitative focus, the results suggest a minimal overall impact of the returnees on the sending country. This would be explained by returnees' ambivalent sentiments towards their return experience as well as no ability or desire to Americanize Greece.

Another study which assesses the socio-economic impact of the migratory flows particularly for the sending countries is that by Bohning (1975). Geared more towards exploring issues of development and employment, Bohning's analysis of return migration focuses on emigration statistics from the Mediterranean mostly to western European countries. The book questions the supposed benefits of return migration brought by emigration specifically, and argues against a "laissez-faire" policy. Focussing on the Italian case, Rosoli's analysis (Rosoli, 1977) of return migration reaches the same conclusions as Bohning's (op. cit.), but it also highlights the importance of addressing the complexity of the migratory phenomenon not solely from a political and economic perspective, but also crucially in terms of the collective experience understood as the sum of migrant's sacrifices and challenge's once back (244).

Musillo (1981) survey also attempts to answer some of the questions concerning the benefits of migrants' return for their country of origin, but it does so through the analysis of employment rates and interviews of a small (200) sample of Italian migrants who returned from Switzerland to southern Italy between 1969 and 1970. Specifically, this work aims to investigate the causes of return migration in relation to the reception policies implemented by the Italian national and regional authorities. The results again suggest a significant discrepancy between return policies and any concrete benefit.

In relation to the investigation of the reasons for returning, Rogers (1983) analyzes census and survey data from several countries and cross-examines them with findings from previous studies. Her results highlight some of the gaps in research on return migration at the time, including the lack of cultural indicators in the surveys, sample limitedness, and the inadequacy of a persistent model that sees the migrant who returns essentially as a failure to integrate into the host society. Though still in relation to the time when the study was carried out, the conclusions point to the need for a more nuanced investigation of the phenomenon of return migration that would go beyond the binary conceptual framework of whether it is good or bad.

Gentileschi and Simoncelli (1983) address the specific question if and to what extent return migration rebalanced regional and local differences. From the review of relevant data for Italy of the previous twenty years, the book concludes that return migration has not had any significant effect on the territorial imbalance. The authors also suggest that research on the topic should consider the migrant's family rather than the individual, thus opening interesting avenues for a more cognitive and cultural angle. Kubát and Center for Migration Studies (1984) analyze return migration in Europe from a socio-political perspective. The book attempts to map the impact of return migration on the sending countries in relation to restrictive measures in the receiving countries as well as integration policies in the sending countries aimed at encouraging return. From data on Algeria, Tunisia, Spain, Portugal, Italy, Greece, former Yugoslavia, France and Turkey among others, the book concludes that "emigration has failed to provide a discernible developmental impetus in any of the countries" (ibid., 266). More recently, Wyman (1993) draws on official U.S. and European statistics on returned migrants, previous studies on the experience of return migrants, and his own archival work to shed light on the migrants' motivations to return. The author highlights commonalities and differences between national groups such as the fact that for many Europeans, the intention had never been to settle in America, and so returning was not the result of a failure, but an undisputed decision.

As this brief literature review shows, these important contributions all acknowledge the relevance of integrating the study of return migration with a more comprehensive analysis of the cognitive and subjective aspects of the migratory experience. Methodologically however, these works reflect a general tendency to either take a macro approach based on the analysis of historical events, statistics and socio-economic indicators or to adopt a zoom-in perspective based on a limited amount of small case studies. There seems to be a frustrating lack of large-scale investigations that would devote attention to the cultural perception of return migration –both by the country of origin and by migrants themselves –as an alternative entry point to the question of why they returned. In other words, historical analyzes that focus on the inner voice of migrants and on how they perceived returning beyond the reasons for returning remain comparatively rare. The reasons for that are complex but on the whole scarcity of primary sources may be at the root of the gap. Moreover, for the most part remigration is rigidly envisioned as a one-way movement from the host country to the country of origin and, perhaps even more critically, as the final stage of the migration experience.

This article explores a novel way to understand historical return migration as a more complex and discursive process. It does so by using migrants' own narratives of migration both as a data collection tool and an analytical object (Baynham and De Fina, 2005; De Fina, 2017; Viola, 2021). Without losing the quantitative advantage of large-scale data, the aim is to obtain a more intimate investigation of return migration than statistics, analysis of public discourse alone or small, post-facto interviews can offer that could allow us to explore unanswered questions about the mass migratory movements of the past century. The case study is Italian American migrants at the turn of the twentieth century. The study uses as its main primary source ChroniclItaly 3.0 (Viola and Fiscarelli, 2021), a collection of Italian diasporic newspapers published in the United States between 1897 and 1936. By using diasporic newspapers –newspapers written by migrants for the migrants –the analysis explores migrants' stories of their own migratory experience (Viola, 2021).

The study rests on two hypotheses. The first one is that the Italian domestic discourse about remigration was significantly different from the Italian American one. We base our hypothesis on the fact that discourses of migration (emigration, immigration, and remigration) are always built on justification and legitimation (especially when societies feel threatened) as well as moral values and fluctuate with the development of economic needs and political agendas (Viola and Verheul, 2020b; Oberbichler and Viola, 2023). The immigrant press, on the contrary –though certainly mediated –may offer a more authentic voice of the migratory experience in that it was produced by the diasporic community itself. To test this hypothesis, the analysis will compare the narratives constructed in the Italian American newspapers and used by migrants to communicate (return) migration with the baseline of public discourse of emigration and remigration in Italy.

The second hypothesis is that remigration is a movement both from the host country to the country of origin and vice versa. Historical scholarship of migration has widely acknowledged the phenomenon of the so-called "birds of passage," that is temporary migration of mostly young males who traveled back and forth depending on the accumulated capital when abroad. This pattern is shared by many, if not all national groups, including of course Italy. This means that the final decision of settling in America or returning to the home country was often the result of several transatlantic trips, sometimes in the span of years during which these individuals, whilst living through larger historical changes, learned from the experience and were profoundly changed by it. Within this framework, the displacement process of remigration, it is argued here, can equally be conceptualized as a crisis of identity, in both collective and personal identities. This is justified by the fact that social processes of displacement and transformation deeply affect people's inner notions of identity and belonging. For example, in the context of Italian Transatlantic migration, historical statistics show that Italians were one of the groups most actively engaged in Transatlantic remigration but more recent readings indicate that many of these returnees eventually re-emigrated to the US (Boyd Betty, 1973). Indeed, one of the open questions about (r)emigration is the discrepancy between migrants' statements at the departing ports and their actual behaviors. By exploring this more subjective element, this study may expand on the current conceptualization of return migration as the ending point of the experience.

For example, according to Cerase (1967), there are four different paradigms of return migration: failure, conservation, investment, and retirement. The first one would describe a failure of the migration project, for instance a failure of integration in the host country, and it would entail a rapid re-integration in the country of origin. The second one, the migrant of conservation would be the one who has kept ties with the country of origin, for instance through several trips and who, once accumulated enough capital, perhaps after five to ten to years, decides to return to the society of origin, but we no ability to impact it in any way. The migrant of investment would describe a type of migrant who has internalized new values in the host country and once achieved their personal goals, is eager to contribute them to the old society. Finally, the migrant of retirement would be the type of migrant who only returns once the accumulated capital allows them to live comfortably in the country of origin without having to work and who, consequently, would not be interested in contributing to the society of return. Cerase's model may therefore explain whether remigration is beneficial or detrimental for the sending country in relation to the reasons for returning. However, even though it recognizes the 'birds of passage' pattern, it rigidly understands return migration as the final stage of the migration experience. In other words, it fails to acknowledge that the experience often did not end with repatriation. Indeed, returnees did not always meet favorable circumstances on their return and in the mid-, long term, the difficult readjustment disappointed their expectations. The skills supposedly acquired abroad very rarely fitted the economies of the places of origin and any impulse of innovation often found major socio-cultural and political obstacles (Rosoli, 1977). It is therefore not unreasonable to hypothesize that particularly for the migrant of investment –the only one truly capable of representing an innovative contribution to the society of origin –remigration abroad became the only alternative. This study wants to explore this more complicated pattern of migration. Resting on the foundation that discourses on migration also inevitably go along with a sense of crisis (Viola and Musolff, 2019, 3), we posit that the more nuanced experience of migration provided by immigrants' newspapers will narrate the intricate relationships between specific manifestations and negotiations of identity and the wider experience of migration as embedded in the larger social context of both the American and the Italian society of the time.

Another distinctive feature of this study is the integration of Critical Discourse Analysis theory (Fairclough and Wodak, 1997) into the quantitative investigations (collocations and ngrams). *ChroniclItaly 3.0*, the archive used for the investigations, contains 21,454,455 words, therefore using close reading methods alone would not allow us to make full use of the record. The proposed methodology will first provide a comprehensive, zoom-out overview of how words are distributed in the newspapers and second, will identify relevant narratives. CDA will then be used as the applied theory for the analysis of the narratives of (r)emigration. Despite the ever-large amount of available digital sources, particularly for historical research, to my knowledge only two studies exist to date that have conducted digital discourse-based analysis of remigration on large quantities of textual data (Oberbichler and Pfanzelter, 2022; Oberbichler and Viola, 2023). The methodology proposed in this research may therefore have wider relevance for historians and other humanities researchers who are increasingly confronted with the challenge of having to navigate the complexity of sources abundance, particularly historical.

Finally, as an additional innovative contribution, this study will be the first to use immigrant newspapers in remigration scholarship to explore the tension between questions of policy and governance as manifested through language in the public discourse and the individual's actual experience as it emerges from their own discourse of migration. Though certainly considered as an important source for migration history, immigrant newspapers have comparatively received less attention than national media, especially in digital research, because they were believed to be considerably less worthy of scrutiny. To the contrary, as recent digital research on migration and identity has shown (Viola and Verheul, 2019a,b, 2020a; Viola, 2021) and as it will be further demonstrated here, immigrant newspapers offer unique emic perspectives on how migration was perceived by migrants through how it was narrated to migrants themselves. In this way, this study also actively engages in the debate about language diversity representation and archival biases in digital practices.

Italian Transatlantic (r)emigration

For the period of mass migrations (1860–1920), official statistics indicate that about 40% of all European migrants eventually returned. More recent studies, however, evidence that due to differences in how the data were gathered between sending and receiving countries and several errors in how individuals were counted, the return migration rate from the host country to the country of origin may have in fact been as high as 70% (Gould, 1980; Bandiera et al., 2013; Dustmann and Görlach, 2016; Abramitzky et al., 2019). As for Italy, the most accurate data on return migration were gathered by the *Commissariato Generale dell' Emigrazione* (general commissariat for emigration) in the years 1905–1906 (Beneduce, 1910). The survey also compared rates of return in 1905–6 with rates of emigration in 1901–1905, this is because available statistics at the time indicated that most Italians

were staying in America between two to five years. According to these data, *I Ritornati* (the returnees), as they were called, were about 40%, they were predominantly males (90%), young (between 14 and 44), uneducated and unskilled. Women were less likely to return because they would typically emigrate to join their husbands who had already decided to settle abroad. Although certainly valuable, these data suffer from severe limitations; above all, they exclude from the count the returnees entering Italy through ports other than Genoa, Naples, Palermo, and Messina, or by railroads.

In the United States, data on return migration were gathered from 1907 to 1908 by the Commissioner of Immigration. As noted by Gould (1980), however, migration statistics based on these numbers give an inaccurate picture of the actual intensity of return migration, or migration in general for that matter, because the data collection was based on the statement of the individual about their intention to stay abroad or to return –not actual figures. Latin American surveys are even less reliable. Nevertheless, both Italian and American statistics relatively consistently show that although Italians were the largest community of immigrants, they were also the group most actively engaged in return migration. There seems to be consensus now in the literature that between 1900 and 1920, at least 50% of individuals returned, though again this percentage varies greatly according to the year and the region in Italy (Boyd Betty, 1973).

Another percentage that is of particular interest is the one indicating the number of individuals the Dillingham Commission[1] referred to as *non-emigrant aliens*, that is non-US individuals who had declared that they would return to the United States within a year. These figures show a steady decreasing trend therefore suggesting that most returnees intended to resettle permanently in Italy (United States Bureau of Immigration and Naturalization, 1908). However, crossing data from Italian departures and American arrival statistics, the numbers tell a different story. They show that even if most returnees had initially intended to resettle permanently in Italy, they eventually re-emigrated to the US as indeed, second or even third departures were a common phenomenon happening at intervals of five to ten years (Cerase, 1967). By investigating Italian migrants' own narratives of migration, this study explores the potential reasons for re-emigrating to the US.

The Italian debate of (r)emigration at the turn of twentieth century

The political unification of Italy (1861) had dramatically worsened the already critical economic situation of the South, making the task of creating a nationally integrated economy virtually impossible (Viola, 2019). The rapid impoverishment of the South, the *Questione meridionale* (the southern question) as it was called, soon became the government's top priority, therefore dominating the public debate (Wong, 2006). However, despite several targeted programs and special legislations geared towards solving it, three decades later the delay of the South was worse than ever.

---

1 The U.S. Congressional commission responsible for investigating immigration.

It is in these years that discourses on migration (both emigration and return migration) entered the public debate. Emigration was generally considered as a negative phenomenon, seen as a disruption of the social and moral order, a drain of capital, and a side-effect of the process of national integration which ultimately made Italy look bad abroad (di Cosentino, 1873). Accordingly, the Italian Government was against emigration (Manzotti, 1962; Ostuni, 2001; Choate, 2008). However, despite efforts to limit emigration, towards the end of the century migrating movements, especially from the South to Europe and America, reached numbers so impressive that the Government had to change its strategy. And because the highest emigration flows were departing from the South, naturally the topic of emigration became entangled with the *Questione meridionale* (Manzotti, 1962) and more precisely, framed as a way to solve it. After decades of failed attempts at rescuing the South, politicians started to argue that there was nothing that the Government could do to help the cause because, in fact, the South could not be helped (Wong, 2006). Concurrently, some economists began to highlight the positives of emigration. Emigration was relieving the South of Italy –and therefore the country as a whole –from the demographic burden of unemployment. Moreover, thanks to migrants' remittances, Italy was experiencing an unprecedented flow of cash which in the long term would have made the South –and therefore the country as a whole –richer. Perhaps emigration was not as disrupting as observers had originally argued only a few years before. Perhaps emigration and remittances were in fact the only way the southern question could finally be solved.

It is important to state that at the turn of the twentieth century, within the national debate of emigration Italian migrants were expected to return. This expectation was in line with the general European trend for which migrants were indeed returning. The argument was that for as long as emigration was necessary, Italians would continue to leave but these migrants would alternate on a rotating basis every three to 5 years. In other words, virtually every migrant would have eventually come back (Bertani, 1886). Moreover, so the argument went, these migrants would have returned with money and new acquired skills and their enriching experience would have contributed to the cultural modernization of the Italian society. Naturally, migrants who were not conforming were victim of harsh criticisms; if they decided to settle permanently abroad, they were framed primarily as amoral individuals or disloyal to the nation and its values, and as a failure or as parasites if they did not promptly invest their savings once back in Italy (the so called "returnees of retirement") (Jacini, 1885; Cerase, 1967, 1974).

As the debate was shifting more towards emigration as being beneficial for Italy, the role of the Government became at the centre of the discourse. One of the main arguments was that the Government had the obligation to regulate emigration so that its benefits could be maximized, and its disrupting effects contained. In other words, regulation of emigration was in the national interest (Carpi, 1871, 1878). The discussion mostly focused on remittances and their intelligent exploitation (Balletta, 1968). Indeed, remittances were considered a true national asset, crucial to the Italian economic integration in the international capitalistic system. At the individual level, remittances may have been small, but in the aggregate, they represented substantial amounts of money. For this reason, the Government had to protect

remittances every step of the journey from the rest of the world to Italy. Better yet, it had to regulate the way remittances were channeled so that immigrants' savings could be used productively (Gatto, 2021). As the late Prime Minister Francesco Crispi claimed, after all national and individual immigrants' goals were the same and assisted emigration was the key to reach them both (Crispi, 1915).

These arguments culminated in the 1901 law, the first Italian law on emigration. The law aimed to assist migrants in all matters concerning the migratory experience, including issuing the necessary documents, setting ships' sanitary standards, creating guidelines for transatlantic fares, and establishing employment offices in the major American destinations. From that moment on, all these activities were to be managed by the newly established *Commissariato Generale dell' Emigrazione*. Remittances would also start to be regulated. With this law, the Government officially gave the Banco di Napoli the task to collect, protect, save, and transfer migrants' remittances from all over the world to Italy (Gatto, 2021). The aim was to minimize the number of intermediate actors in the transferring process, to keep commissions and exchange fees low and guarantee that the highest possible amount of money would reach Italy. The centralization of the remittances would have then protected migrants' savings, and thus the country's interests. However, due to both structural and cultural barriers, the plan to centralize the channeling of remittances never reached the Government's unrealistic expectations, as most migrants would still send their money through *banchisti*,[2] friends, relatives, or post offices. It has been calculated that in the end, only one-fourth of the remittances reached Italy through the Banco di Napoli (Commissariato Generale dell'Emigrazione, 1925).

The 1901 law remained practically the same until fascism when the regime firmly rejected emigration as weak and unpatriotic (Gentile, 1986; Choate, 2008; Braun-Strumfels et al., 2023). Fascism opposed the narrative that Italians were forced to emigrate, and it replaced the word *emigrante* (emigrant) with the more pleasant title *lavoratore italiano all'estero* (Italian worker abroad). Along with the same narrative, in 1927 the *Commissariato Generale dell' Emigrazione* was abolished. But for more than three decades prior to fascism, the national discourse had praised migration, return migration, and remittances for single-handedly being able to achieve what decades of governmental measures had failed to accomplish. The reason behind such a naïve and opportunistic vision of (r)emigration and the hype about remittances as a miracle cure may be found precisely in the failure of all previous governmental programs. Manzotti (1962) for instance argues that it was with the Second World War that the deeper fractures in the Italian economy became ever more apparent, demonstrating that migration was only one aspect of a much wider economic crisis.

---

2   The term banchisti refers to a specific group of Italian Americans who claimed to be bankers but profited from the situation of uneducated and naive Italian migrants. These individuals may have exploited their fellow immigrants through deceptive financial practices or scams. In many cases, these individuals took advantage of vulnerable newcomers who were unfamiliar with the financial system and its regulations in their new country. This type of exploitation is not unique to Italian immigrants but has been observed in various immigrant communities.

Sources and methodology

The immigrant press to access narratives of migration

One of this study's most distinctive features is the use of immigrant newspapers (*ChroniclItaly 3.0*-Viola and Fiscarelli, 2021) to explore questions of belonging and identity in relation to the experience of migration and return migration. Beyond the binary conceptual framework about the function of the immigrant press in the USA, i.e., either assimilating or retarding assimilation (Hickerson and Gustafson, 2016), in the absence of large-scale accounts of migrants' personal experience, this study explores immigrant newspapers as a way to *unsilence* the voice of migrants.

The immigrant press constitutes the first historical stage of the ethnic press and it is a phenomenon associated with the mass migration from Europe to the Americas between the 1880s and 1920s (Viola and Verheul, 2019a). During this period, it is estimated that in the USA alone ~1,300 foreign language newspapers were read by about 2.6 million people (Rhodes, 2010; Bjork, 2013). As for the Italian immigrant press, recent calculations estimate that in the period of reference, between 150 and 264 Italian language newspapers were published in the USA, of which 98 managed to publish uninterruptedly (Deschamps, 2011, 81; Vellon, 2017). The newspapers of the years 1880–1920 are generally divided into two main categories: *prominenti* and *sovversivi*.[3] The *prominenti* were mainstream newspapers whereas the *sovversivi* were radical publications of socialist and anarchic orientation.[4] In terms of reach, their circulation ranged from few hundreds to many thousands (Vecoli, 1998). In his 1922 investigation of the role of the immigrant press in the USA, urban sociologist Robert E. Park reported that in 1900, 691,353 Italian newspapers were sold across the United States (Park, 1922, 304) and in New York alone, the circulation ratio of the Italian daily press was one paper for every 3.3 Italian New Yorkers (Vellon, 2017, p. 10). These already impressive numbers should however be doubled or even tripled, since most Italians were illiterate at the time and newspapers were often read aloud (Viola and Verheul, 2019b).

Despite the high distribution and circulation figures, some scholars have argued that the influence the Italian language press exerted on the immigrant community was rather limited (Russo, 1972; Vecoli, 1998, 2006). The reason for that would lie in the fact that immigrant newspapers could not impose their definition of social reality because ultimately, migrants had to make sense of it within the context of their own migratory experience. Others have stated that because they were often pushing specific individual agendas, these media in fact damaged Italians (Pozzetta, 1973). Though true to an extent, it is undisputable that

---

3   Although these two categories include the majority of the Italian American newspapers of the period, there were also some self-claimed politically independent publications such as *The Patriot, La Sentinella del West* and *The Independent. ChroniclItaly 3.0* includes them all.

4   For a more in-depth discussion of these two types of publications, see Vellon (2017), Viola and Verheul (2019a,b), Deschamps (2020), Viola (2021).
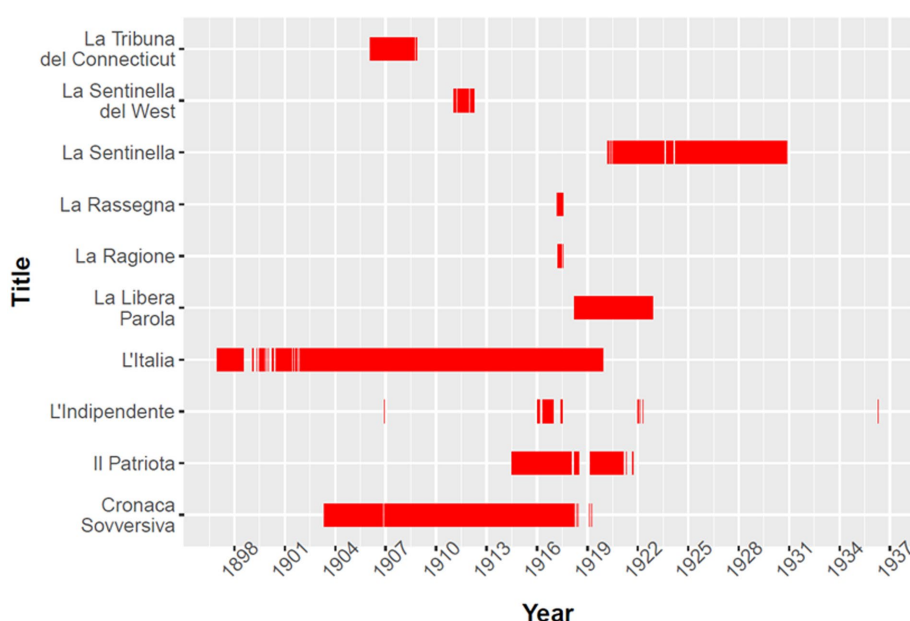
by functioning as tools of language retention and national identity construction and preservation (Vellon, 2017; Viola and Verheul, 2019a,b), Italian newspapers became a powerful instrument for community building (Park, 1922; Vellon, 2017; Viola and Verheul, 2019a,b; Deschamps, 2020). Moreover, beyond individual agendas, Italian immigrant newspapers would play an important social role for example by offering practical and cultural information whilst at the same time reporting news from the homeland. Finally, they championed for the rights of the Italian immigrant community by supporting nationalistic campaigns, entering pleas for convicted Italians, holding fundraisings for natural disasters in Italy, and voicing protests against mistreatments of Italians (Vecoli, 1998).

ChroniclItaly 3.0

The collection *ChroniclItaly 3.0* used in this study features ten titles of *prominenti*, *sovversivi*, and independent newspapers published between 1897 and 1936. *ChroniclItaly 3.0*[5] is fully Open Access. The titles included in *ChroniclItaly 3.0* are: *L'Italia, Cronaca Sovversiva, Il Patriota, La Libera Parola, La Rassegna, La Ragione, L'Indipendente, La Sentinella, La Sentinella del West,* and *La Tribuna del Connecticut* for a total of 8,653 issues and 21,454,455 words.

L'Italia is the title with the highest number of issues in the corpus (6,489) and by far, the one that covers the longest timespan (1897–1919) (see Figure 1). The title was founded in 1886 by a group of Italian *prominenti* from the fusion of two failing papers. Initially, the newspaper was published bi-weekly, but from 1889,

it was published daily. In 1895, the editor-in-chief of *L'Italia* was Pio Morbio, co-founder of *Il Corriere della Sera*, one of the main newspapers in Italy at the time, while in 1897 Ettore Patrizi and Giovanni Almagia became co-editors. In 1898, Patrizi became the sole owner and publisher of the newspaper. For about two decades, under Patrizi's lead, *L'Italia* voiced a leftist ideology, close to the Italian labor class, and defended Italians against defamation and discrimination. However, after 1909, Patrizi overtly embraced Mussolini's ideology and the newspaper's orientation became ardently nationalistic.

Cronaca sovversiva, a *sovversivi* newspaper, is the second largest title in *ChroniclItaly 3.0*; it includes 771 issues from 1903 to 1919. It was founded by the anarchist Luigi Galleani in 1903 who had escaped extradition a few years before and had settled in Barre, Vermont, where an Italian community of stonemasons was living. Galleani published the anarchist newsletter for 15 years until the United States government forced him to stop under the Sedition Act of 1918. *Cronaca Sovversiva* typically discussed a variety of radical topics, including arguments against the existence of God and against historical and contemporary establishment. Like all radical press, *Cronaca sovversiva* not only served as the main means of communication for the *sovversivi*'s community but it acted as the movement's financial centre (Bencivenni, 2016). Deschamps (2020) argues that within the immigrant press, it was especially the radical press that truly functioned as a transnational tool since it was published in the US but it was mainly targeted at a readership in the country of origin.

La Sentinella includes 518 issues from 1920 to 1930. *La Sentinella* included more pages than other Italian-language newspapers of the time, with some issues extending to a full eight pages. Interestingly, during the early 1920s, each issue included a page dedicated to news from the Italian American community of Port Chester, New York. The exact foundation date is unknown, but it might have been in 1913 or

1914.[6] Politically, *La Sentinella* was a Republican-leaning paper and it overtly supported the Fascist ideology (Bucki, 1995).

Il Patriota contains 339 issues from 1914 to 1921. It was founded in 1914 in Indiana, Pennsylvania by Francesco Biamonte. It claimed to be a politically independent newspaper aiming to inform Italians in the region and to offer immigrants advice on adjusting to American life. *Il Patriota* encouraged Italians to become naturalized citizens and it permanently featured a column listing questions probably taken from the citizenship test.

La Libera Parola – which was originally called *La Voce del Popolo* – features 240 issues from 1918 to 1922. It was founded in 1906 in Philadelphia by the two brothers Arpino and Giovanni Di Silvestro. It was a weekly newspaper which publicized the activities of the Pennsylvania Chapter of the Order of the Sons of Italy and overall had a nationalistic orientation. For example, *La Libera Parola* supported Italy's participation in the war and criticized Pope Benedict XV for opposing Italy's involvement in the conflict. The paper also encouraged Italian-Americans to become American citizens, enlist in the military, and buy Liberty Bonds to help finance the Allied war effort.

La Sentinella del West Virginia contains 53 issues from 1911 to 1912. It was West Virginia's only Italian periodical. It was founded by Rocco D. Benedetto in 1905 and by 1906, its circulation had peaked at 3,500 copies. Although Benedetto was an active Republican, the newspaper claimed to be politically independent. The publication mainly informed immigrants about Italy, but also about their new homeland. The paper discussed American politics and current events, but it mostly chronicled the stories of the Italian immigrants in West Virginia.

La Tribuna del Connecticut contains 130 issues from 1906 to 1908. It was a weekly newspaper published in Bridgeport, Connecticut where an Italian *colonia* (colony) was rapidly growing. The paper claimed to be independent and voiced support for socialism, striking laborers, and the International Workers of the World ("Wobblies," for short). *La Tribuna del Connecticut* published reflections on America by Russian writer Maxim Gorky; gave prominent coverage to notable socialist intellectuals visiting Bridgeport, like Italian editor Carlo Tresca; and responded to articles and ideas then being discussed in Italian socialist newspapers. It also informed the Italian community about dances, concerts and other social activities. Of note is that *La Tribuna del Connecticut* also served Danbury, Connecticut and Port Chester, New York; it also had a wide network of stable regional correspondents who were located in more than a dozen cities in the Nutmeg and Empire States (Connecticut and New York). In 1913 Altieri became the editor of *La Sentinella*.[7]

L'Indipendente includes 48 issues from 1907 to 1936. It began publication as a weekly in 1904. It was headquartered in Wooster Square, one of two neighborhoods (along with the "Hill") which housed New Haven's large Italian immigrant community. It claimed to be the "first and only Italian daily in New England." Its aim all was to support and protect Italians in New Haven. Although it claimed to be pro working classes, the tone of the coverage was pro-capitalist and pro-American which was not uncommon in the early 20th century Italian American press.[8] Unlike other Italian American newspapers that presented it as a source of national pride, *L'Indipendente* overtly opposed the Italian colonial invasion of Ethiopia of 1936.

ChroniclItaly 3.0 also features the whole 40 issues of *La Ragione* from 25 April to 23 August 1917, as this newspaper only survived eight editions. It was published in Philadelphia and its main aim was to expose corrupted personalities within the Italian community such as *prominenti* and dishonest bankers (*banchisti*). *La Rassegna* was also a short-lived newspaper published in Philadelphia in 1917. The archive includes 25 issues from 7 April to 25 August 1917. It focused on issues affecting Italian immigrants in Philadelphia and chronicled major historical events such as World War I and Italy's nationalistic claims to Dalmatia. In addition to defending *prominenti*, *La Rassegna* encouraged Italian immigrants to seek naturalization.

Due to economic struggles, immigrant newspapers were often abruptly discontinued; this is reflected in the composition of the collection which presents both gaps across titles and differences in the number of issues per title as described above and shown in Figure 1. However, the factor that most heavily conditioned which titles and which issues are included in *ChroniclItaly 3.0* is the existence of a complete, or largely complete, microfilm "object of record" with priority given to higher-quality microfilms (Viola, 2023). This requirement was set by the American National Digital Newspaper Program (NDNP), the American National Endowment for the Humanities (NEH), and the Library of Congress in 2005, when the mass digitization program *Chronicling America*[9] started. This criterion is still adopted for reasons of efficiency and cost; however, as in the past microfilming practices in the United States were entrenched in a complex web of interrelated factors (Baker, 2002), including economic and political interests, the material in the directory incorporates issues such as previous decisions of what was worth microfilming and more importantly, what was not (Rumsey and Digital Library Federation, 2001; Fagan, 2016; Viola, 2023). Though titles and issues are constantly added to the database, this element has also influenced the periodization of the collection, i.e., 1897–1936, in that it reflects the titles available in *Chronicling America* at the time when *ChroniclItaly 3.0* was harvested. In other words, *ChroniclItaly 3.0* was not created with a specific periodization in mind nor for the specific goal of studying Italian American migration. The aim was to provide a digital source of Italian American newspapers for scholars of nineteenth-century periodicals and intellectual and digital history, but more widely for historians, linguists, media and communication scholars interested in topics such as conceptual change, continuity and replacement, and representation of actors and events in public discourses. More widely still, for anybody concerned with text mining in the humanities (Viola, 2021). Nevertheless, the author acknowledges the far-reaching network of influencing factors and actors involved in digital research which have impacted the creation of *ChroniclItaly 3.0*.

Despite these limitations, the collection arguably maintains an acceptable degree of balance between the representation of titles of different political orientation, geographical distribution, and numbers of issues throughout the period. Figure 2 shows the different places of publications of the newspapers collected in *ChroniclItaly 3.0*. The

---

6   https://chroniclingamerica.loc.gov/lccn/sn84020351/

7   http://www.loc.gov/chroniclingamerica/lccn/sn92051386/issues

8   https://chroniclingamerica.loc.gov/lccn/sn93053873/holdings/

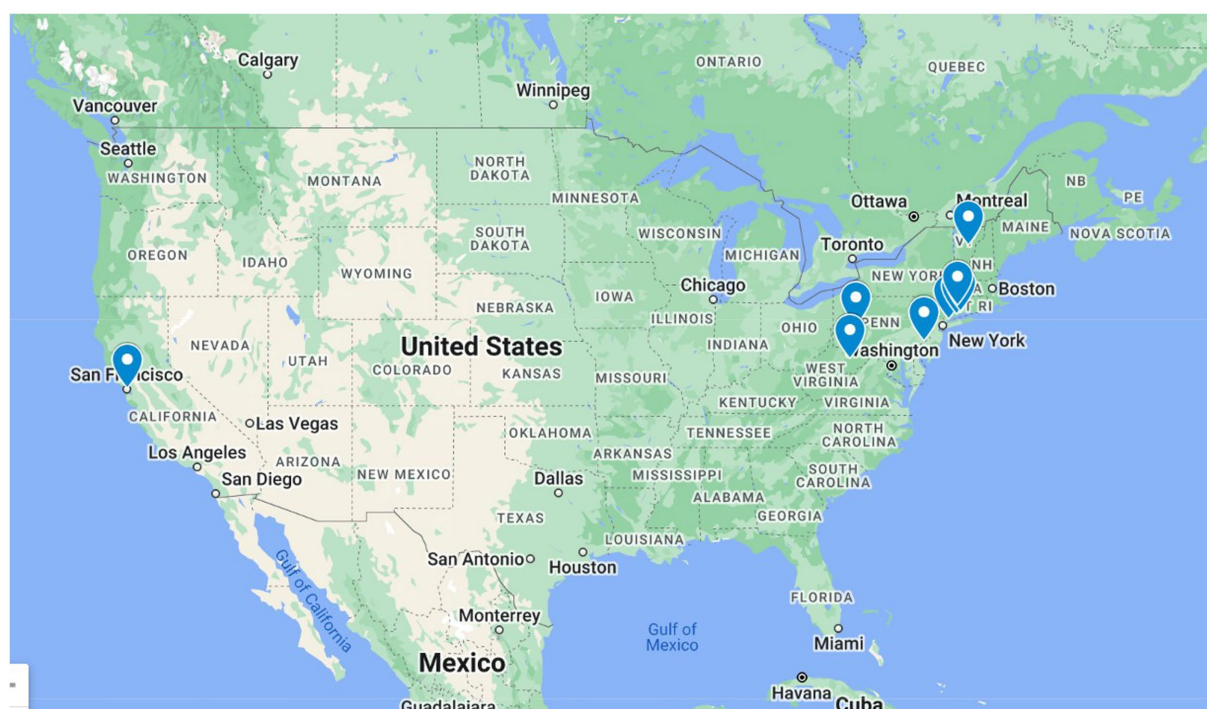9   https://chroniclingamerica.loc.gov/

FIGURE 2
Distribution of places of publication of the newspapers in *ChroniclItaly 3.0* per title.

author nevertheless acknowledges that although the resource provides a reasonably comprehensive picture in the period of reference, discourse of migration produced by the Italian American community, issues such as over-or under-representation of some titles and potential polarization of topics may arise.

Methodology

This study combines text mining techniques such as collocations and ngrams with Critical Discourse Analysis (CDA). Collocations and ngrams are used to identify relevant articles and passages in the corpus. Collocations are words that appear frequently within a certain distance of the search term and provide a picture of which words co-occur with other words in a corpus; ngrams are a contiguous sequence of two or more (n) items in a text. Ngram language models are probabilistic models –known as Markov chain (Gagniuc, 2017) – for predicting the next item in a sequence of elements in which the probability of each item depends on the proceeding one. These two methods complement each other; together, they provide a comprehensive overview of how words are distributed in the analyzed material and therefore they are especially helpful for exploring large quantities of unstructured textual data. Particularly for cases in which keywords searches are challenging, for instance because some concepts are difficult to define linguistically (such as return migration) (Oberbichler and Pfanzelter, 2022), these methods may reveal important insights into the concepts migrants associated with the notion of remigration as well as answer questions of discourse proliferation and awareness.

Guided by CDA theory, the narratives of remigration are analyzed in identified excerpts. Specific attention is given to how retuned migrants are discursively represented in the diasporic media and through which communicative strategies their identity is constructed. As in any crisis, crisis of identity are also always attempts at maintaining power (Viola, 2022). For example, when a host society feels threatened, concepts of "foreign" and "belonging" are used to justify specific ideologies and thus usually become explicit; but in contexts of return migration, individuals can paradoxically become strangers in their own country of origin. The analysis will therefore focus especially on understanding how processes of social inclusion and identity construction and representation unfold in diasporic contexts as well as within the same community. The first hypothesis is that contrary to the Italian national debate that simplistically framed emigration, return migration and remittances as the single solution to a problematic, social situation, the discourse of remigration in the immigrant press –though still mediated –was more ambivalent in that it was produced by the Italian diasporic community and therefore enacted as a consequence of its own migratory experience. The intention is to obtain a more intimate representation of the experience of return migration as a changing, sometimes traumatic experience, both for the returnees and the Italians who had never left. By analyzing the conflicting vectors of public discourse on migrants and discourse by migrants, the aim is to open up avenues for a critical reflection on emerging common as well as diverging themes and the impact these may have had on the general perception of return migration as a crisis of identity (second hypothesis).

FIGURE 3
Frequency of distribution of N-grams in *ChroniclItaly 3.0* with selected words.
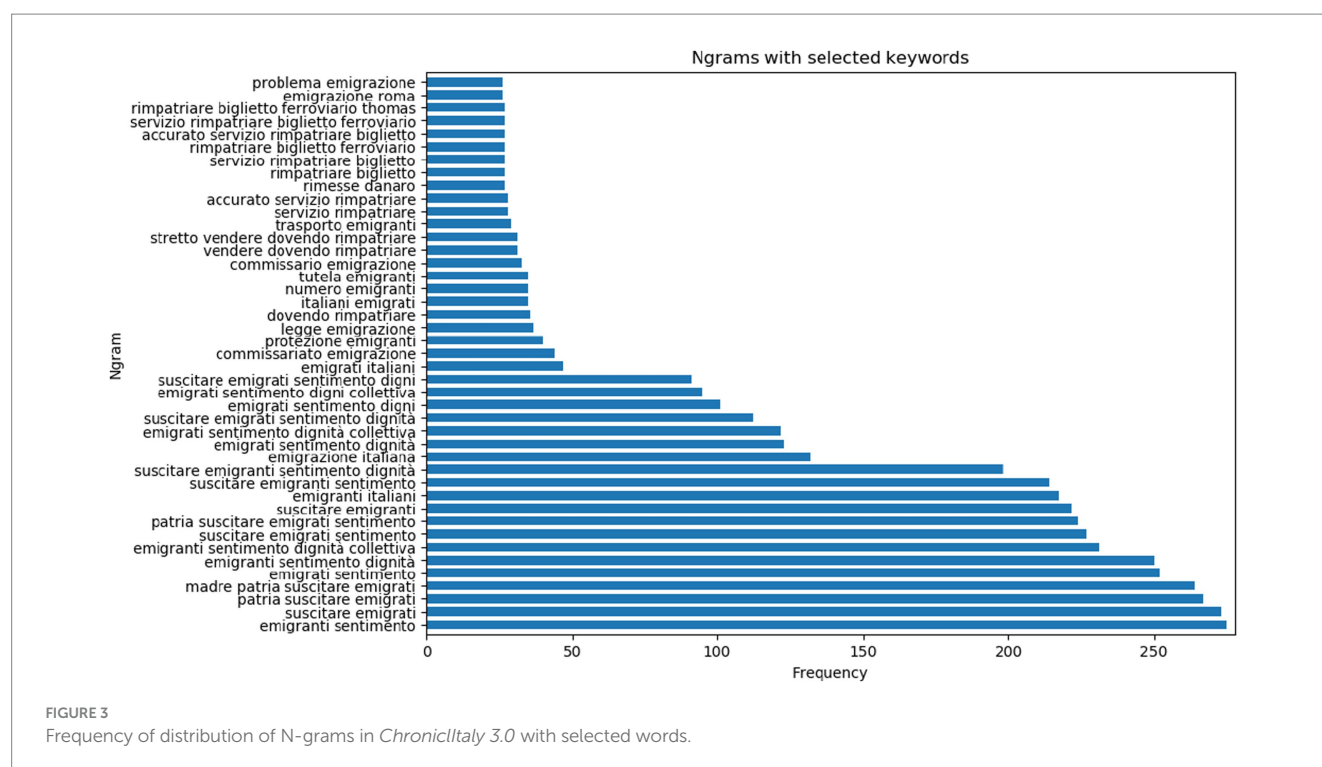
Analysis and results

This section shows the results of the analysis following the described methodology; in particular, the first section shows the results of the ngrams analysis while the second one analyzes identified excerpts using CDA.

Ngrams

Ngrams (bi-grams, tri-grams and four-grams) are computed for salient words so as to gain insights on their semantic distributions in the corpus. The used words are *emigrazione* (emigration), *emigrant* (emigrants), *rimpatriare* (to repatriate), *ritornare* (to return), *rimesse* (remittances) in all possible morphological combinations (e.g., *ritornare*, *ritornati*, *ritornato*). The results are displayed in Figure 3.

A few interesting considerations can already be drawn from the results. First, ngrams containing the word 'emigranti' (emigrants) or 'emigrati' (emigrated) have the highest frequency of occurrence whereas ngrams in combination with 'rimpatriare' (to repatriate) occur much less frequently. This would suggest that the topic of emigration was discussed much more frequently than the topic of returning in turn suggesting that Italian migrants were more preoccupied with matters concerning their migratory experience rather than with returning to Italy. The finding indicates a contrast with the Italian public debate of migration which almost exclusively focused on the economic contribution migrants would provide to Italy upon return and on solving the southern question through remittances (here with the lowest rate of occurrence – 'rimesse').

Second, ngrams containing 'emigranti' are almost always in combination with 'sentimento' (sentiment) and 'patria' (homeland).

As it was custom at the time, ethnic newspapers' front page often reported their mission statement; this would change periodically depending on the newspaper's political orientation or indeed the historical moment. One of the newspapers in *ChroniclItaly 3.0* is *L'Italia*, a *prominenti* newspaper with the longest record of publication. Figure 4 shows the top part of *L'Italia* frontpage of 25 December 1909. The mission statement can be found right below the date. In the years 1909–1910, *L'Italia*'s mission statement was the following (emphasis mine):

> *In questa terra cosmopolita, dove insieme col popolo americano vivono popoli di ogni razza e d'ogni paese; Qui, dove lo spirito di nazionalità si acuisce fortemente nella gara del lavoro tra i figli di tutte le Patrie europee e americane: Qui, il giornale L' ITALIA vive e lotta a difesa del nome Italiano, per tener vivi ed alti l'amore e l'attaccamento verso la Madre Patria, per suscitare nel nostri **emigranti un sentimento** di dignità collettiva Italiana, educando la coscienza loro alla grandezza e virtù della razza Italiana, alle glorie nazionali del passato, al virile e civile risveglio del presente, alle liete e ragionevoli speranze nell' avvenire.*[10]

---

10   "In this cosmopolitan land, where people of any race and from any country live together with the American people; here, where among the sons of all European and American countries the spirit of nationality becomes stronger in the hunt for a job: here, the newspaper L'ITALIA lives and fights to defend the Italian name, to keep the love for and attachment to the Homeland high and alive, to inspire in our **emigrants a sentiment** of Italian collective dignity, educating their consciousness in the greatness and virtue of the Italian race, in the glories of the past, in the virile and civil awakening of the present, in the

FIGURE 4
Frontpage of *L'Italia* 25 December 1909.

The mission statement pushes a narrative aiming at empowering the Italian migrant community by resorting to the cardinal concepts of an Italian heritage of civilization and a glorious past of grandeur. As it has been pointed out in the academic discussion (Cibotti, 1994; Deschamps, 2011; Vellon, 2017; Viola and Verheul, 2019a), such exaltation of the Italian nationalistic sentiment was not isolating the Italian diasporic community; rather, it was part of a wider strategy to construct an ideological concept of Italian identity that would create unity and exert political force to negotiate inclusion (Viola and Verheul, 2019b). Immigrant newspapers used such nationalistic strategies to unify one's in-group and educate the Italian community accordingly. The mass migrations were profoundly impacting the socio-cultural landscape of the United States, most notably visible in the process of redefinition that was affecting social categories such as race, citizenship, and whiteness (Barrett and Roediger, 1997; Foley, 1997; Brodkin, 1998; Jacobson, 1998; Guglielmo and Salerno, 2003; Guglielmo, 2004). Italian immigrants were not spared from this process; they would often be victims of social discrimination, exploitation, physical violence, and even lynching (LaGumina, 1999, 2018; Connell and Gardaphé, 2010). The narrative found in *L'Italia*'s mission statement reflects this struggle to negotiate inclusion in the host society and it shows how this formed a substantial part of the Italian migratory experience.

Third, ngrams containing the word "rimpatriare" (to repatriate) score rather low in the corpus; again, this would suggest that the topic of resettling in Italy was not frequently discussed by the immigrant press. Even more interestingly, "rimpatriare" is found in combination with "dovendo" (being forced to), "servizio" (service), and "biglietto" (ticket). This could indicate that contrary to the

national discourse of migration in Italy for which migrants were expected to return, in the immigrant press repatriation was framed as a forced, rather than a voluntary decision to return. It would also indicate that immigrant newspapers were mostly discussing remigration in the context of giving the Italian community practical information about returning to Italy. The CDA of selected excerpts in the next section will offer a finer grained picture of the different narratives in the repository.

Critical discourse analysis

Collocations lists are generated from the same search terms used to identify salient ngrams and relevant excerpts are retrieved from these lists. CDA is then applied to analyze these linguistic data as discursive "events" (Reisigl and Wodak, 2001), that is as realities in which the social, political, and historical context in which they are embedded is accounted as co-producer of the narratives. The aim is to obtain a richer, more nuanced and intimate perspective of the experience of migration and remigration from the point of view of the migrants as it was shaped by the social context in which the diasporic communities were entrenched. Excerpt 1 is taken from *L'Italia* on 18 August 1904 (emphasis mine).

1. *È da augurarsi che la burocrazia degli uffici ministeriali **tanto lenta quanto è imbecille**, non frapponga inciampi e ostacoli all'esecuzione di disegni e di proposte, mercè cui la **difesa della emigrazione** e la sua tutela **contro le male arti delle arpie** nazionali e straniere esca dal regno della retorica, e diventi realtà immanente ed efficace. […] Non dimentichino in Italia che la popolazione Italiana del Nord America si accosta al milione e che, specialmente nel Mezzogiorno della Penisola, se interi paesi pagano le tasse o non muoiono di fame é **in virtù delle rimesse degli emigranti!***

good and reasonable hopes for the future." In this article, unless otherwise stated all translations are by the author.

2. Let us all hope that the bureaucracy of the ministerial offices, **as slow as it is idiotic**, will not create hurdles and obstacles interfering with bills and proposals, and that the **defense of emigration** and its protection against the **evils of national and foreign sharks** will leave the kingdom of rhetoric to soon become an efficient reality. […] May Italy not forget that the Italians in Northern America are about one million and that, especially in the *Mezzoggiorno*, if entire towns are able to pay the taxes or not starve is **by virtue of the emigrants' remittances**!

More than 3 years after the law on emigration was passed, migrants' conditions had not improved much. Excerpt 1 shows a feeling of frustration, resentment, and powerlessness by the Italian American community towards the Italian Government about the lack of progress towards effective reforms. One of the main novelties introduced by the law had been the centralization of all matters regarding emigration through the establishment of the *Commissariato Generale per l'Emigrazione*. The passage above refers precisely to a visit to the United States by Adolfo Rossi, Inspector of the general commissariat for emigration at the time. The purpose of the visit was to assess the conditions of the Italian colonies, especially outside of the biggest cities, to explore the possibility of establishing employment offices that would give Italian migrants accurate job information and protect their interests on their behalf. Indeed, like millions other migrants, Italians were often exploited as cheap labor and forced to work in extremely poor conditions. The argument in favor of the law had claimed that the reform would have assisted migrants in all stages of their migratory experience, including providing support in the host countries. But in reality, changes had been slow to implement. The passage highlights how Italian Americans were disillusioned by the moderate impact of these reforms and how they were feeling neglected and essentially misunderstood.

The reference to remittances is also significant; it shows that the Italian diasporic community felt exploited by the Italian Government. Emigrants thought that they were significantly contributing to the Italian national economy by sending remittances to their families in Italy. Yet despite their important contribution –perhaps exaggerated in the article –their efforts were not appreciated by the Government which was lost in rhetoric and slow bureaucracy.

Excerpt 2 is taken from an issue published by *L'Italia* 2 years after, 14 November 1906 (emphasis mine).

1. *L'intensità e l'eccesso del lavoro imposti dalla* **povertà** *che esiste fra i nostri connazionali in ragione* **diretta delle loro rimesse postali** *e che da queste viene tragicamente mascherata.*
2. The intensity and excess of work imposed by **poverty** on our fellow Italians is **directly due to their remittances** and by them is tragically concealed.

The excerpt above provides a sad account of the harsh living conditions of Italian Americans in the period of mass migrations; it also gives a more bitter view on remittances than the exalted narrative the national discourse was spreading. The passage shatters the illusion that migrants were becoming rich and that they were supposedly able to save large amounts of money easily and without sacrifices. Migrants knew that this fabricated belief tragically concealed a much darker reality. The pressure of saving as much money as possible to send

remittances to Italy forced them to work excessively and to live in extreme poverty. The sharp contrast with the national discourse suggests that Italian migrants felt their struggles were minimized or even denied and that their experiences were misunderstood.

The following excerpt taken from *La Tribuna del Connecticut* of 27 April 1907 describes the constant tension of the migrant's condition, the longing for a beloved, yet resented country that negates them everything and forces them to leave. This ambivalence remains sustained even when, crushed by nostalgia, the migrant decides to return (emphasis mine).

1. *Poverino!.* **Tutto gli ha negato la patria**! *Il lavoro, il pane, il tugurio, il vestito! Ed egli segue la corrente, che fuori della patria trascina tante preziose esistenze, che* **toglie alla patria i migliori dei suoi figli**! *Parte per raggiungere l'ignoto, pieno di fiducia, di speranza, perche' ha inteso che la', lontano, lontano, oltre l'oceano, c'e' una terra dove* **l'oro si guadagna a palate**. *E parte … Mentre il piroscafo solca le onde, […] guarda, e protende minaccioso il pugno verso* **l'ingrata** *che gli ha negato il pane! Il povero emigrante si volta ancora, non vede più' quel punto, che […] adesso gli e' caro; e il ricordo […] gli da' il capogiro, il cuore palpita, le lagrime gli fanno velo agli occhi. […] E guardando sempre scorge finalmente la terra, e' l'***America**! *L'America, la meta agognata, la* **terra che gli deve procurare quello che la patria gli ha negato**. *Arriva, sbarca, lavora, accumula capitali, spedisce denaro ai suoi, ha il benessere, ha tutto. Ma… un punto oscuro e' sempre nella sua mente, il suo cuore ha un palpito perenne'. Egli dovrebbe esser contento e' non lo e'! […]* **E' il ricordo della terra natia**. *[…] Ma che m'importa della terra natia, egli dice […] E' la lontananza dei miei cari. Se li avessi vicino a me ritornerebbe la calma. E spinto da una repentina risoluzione s'imbarca e inaspettato tocca la terra natia. Rivede i suoi cari, e al suo cuore ritorna la calma.* **Ma la patria e' sempre quell'ingrata**, **ed egli insieme ai suoi ritorna in America**. *Ora li ha tutti vicini a se, ora é tranquillo. Ma egli s'inganna, egli mente a se' stesso, sente sempre dentro di se qualcosa che lo conturba. La malinconia l'invade sempre.* **Lo perseguita il ricordo della terra natia** *[…] Non può' più mentire a sé stesso.* **E' la patria lontana, è la nostalgia**.
2. Poor thing! **His homeland denied him everything**! Work, bread, hovel, clothes! And he follows the current, which carries so many precious lives out of the homeland, which **takes away its best children**! He leaves to reach the unknown, full of trust, of hope, because he has understood that over there, far, far away, beyond the ocean, there is a land **where gold can be earned by the shovelful**. And he leaves … While the steamer cuts through the waves, every now and then he looks back […] and holds out his fist threateningly towards the **ungrateful land** who denied him bread! […] The poor emigrant turns around again, he no longer sees that point, which […] is now dear to him; and the memory […] makes him dizzy, his heart flutters, the tears veil his eyes. […] As he keeps looking, he finally sees the land, it's **America**! America, the craved land, **the land that will give him what his homeland denied him**. He arrives, he disembarks, he works, he accumulates capital, he sends money to his family, he lives well, he has everything. But a dark thought is always on his mind, his heart flutters perpetually. He should

be happy and he's not! […] It is the memory of the native land. […] But what do I care about my native land, he says […] It's the distance from my loved ones. If I had them close to me, I'd be at peace. And driven by a sudden resolution he embarks and unexpectedly reaches his native land. He sees his loved ones again, and calm returns to his heart. But **the homeland is always ungrateful**, and **he and his family return to America**. Now he has them all close to him, now he is calm. But **he deceives himself**, he lies to himself, he always feels something inside that disturbs him. Melancholy always invades him. The memory of his native land haunts him […] He can no longer lie to himself. **It's the distant homeland**, it's nostalgia.

The passage above well exemplifies the tormented condition of the migrant, the perennial crisis of identity that does not allow them to fully enjoy even a wealthy life. The migrant adores the native country but at the same time, they resent it, they feel rejected by it, their country "denied them everything." As they leave, they are overwhelmed by these two conflicting feelings. This tension is also conveyed linguistically in the text. Whereas America is expressly mentioned, Italy is always referred to as "the homeland." This not only highlights a more intimate relationship with the home country, but also a clear opposition with the host country. This narrative thread is sustained throughout the entire passage: the homeland is "ungrateful," America is 'craved'; the homeland denies the migrant everything, America is the promise of gold. When the migrant returns, they sadly discover that the homeland has not changed at all, it has remained "ungrateful." They must leave again, they must **return**, but this time to America.

This passage illustrates the second hypothesis of this study, that migration experience did not end with repatriation. The migrant described in the excerpt may well be the migrant of investment (Cerase, 1974), the one who has done well abroad and could innovatively contribute to the society of origin when returning to it. But upon return, the migrant experiences an old sense of crisis; just like before the departure, the homeland does not welcome the migrant's contributions, nor is it grateful for the sacrifices the migrant has made. For this migrant to leave is, once again, the only alternative. This time, however, the migrant takes the family with them, marking the process as definitive. This more nuanced experience of migration narrates a complicated negotiation of identity between the migrant and the homeland within the wider experience of migration. The migrant will always long for their homeland but something has changed. The migrant understands that their torment is not a desire to go back, it is nostalgia.

Excerpt 4 is taken from *L'Italia* of 16 November 1908 (emphasis mine).

1. *L' **italiano ritornato** in patria, di regola è praticamente male avvezzato. Egli ha **perduto le sue buone qualità italiane**, guadagnando le cattive americane, conservando sempre le peggiori, e non le migliori qualità dei due paesi.*

2. **Italians who return** home have normally bad habits. They have **lost their good Italian qualities** and learned the bad American ones, always keeping the worst, and not the best qualities of both countries.

The excerpt describes repatriated Italians as having been worsened by migration, especially morally. The narrative is built along the 'Us vs. Them' paradigm synthesized in the culturalized images of country-specific qualities, which are positive if referred to the in-group and negative when referred to the Other. This discourse path highlights the differences between the two groups, rejects the values of the Other, and builds confidence in the "us" group. But the passage shows another interesting phenomenon: in the context of return migration, the returnees become themselves "the Other." Because they have lost what indeed made them Italian (the good Italian qualities) in favor of 'bad' American habits, now their right to belong is contested. The "us" vs. "them" narrative emphasizes the in-group identity which in fact refers to the Italians who never left. The process of transformation migrants have gone through is oversimplified and reduced to a linear, binary opposition: Italian = good; American = bad. This description of how returnees were perceived in Italy suggests that their process of social inclusion was complex even within their community of origin. Within this discourse frame, Italians who left were no longer considered Italian, suggesting that once returned, migrants struggled again with their identity negotiation.

Excerpt 5 is taken from *L'Italia* of 19 July 1912 (emphasis mine).

1. *Facendosi l'esame di coscienza, l'Italia deve confessare di **non aver ancora pagato il proprio tributo all'emigrazione**. Quando molti si domandano se l'emigrazione sia un bene o un male, si puo' rispondere paradossalmente — ma non meno veracemente — che l'emigrazione è un bene per la patria, ma **un male per i suoi figli**. È l'individuo che si sacrifica per la collettività. Le **vittime dell'emigrazione** in tempo di pace sono assai piu' numerose delle vittime della guerra.*

2. If Italy examined its own conscience it should confess to **not having paid its dues to emigration** yet. When many ask if emigration is good or bad, one can answer paradoxically-but not less sincerely –that **emigration** is good for the country, but **bad for its children**. It is the individuals that sacrifice themselves for the sake of the collectivity. The **victims of emigration** in peacetime far outnumber war victims.

The article specifically refers to the victims of tuberculosis which at the time was affecting the migrant communities. According to the ship's logs of ships from America to Italy, about 50% of all Italian passengers were affected by tuberculosis (Padovani, 1909). This was because once they would fall ill, migrants could not work anymore, and because typically they could not afford medicines, they would either voluntarily try to repatriate to be cured by their families or they would be expelled by the host country. One out of three of these migrants would die during the journey (*ibid.*). Those who would survive, however, would become a liability both for national and local governments as they represented an economic burden and a threat to national health security. Their families also considered them as a failure as they had become reason for embarrassment in the community (Molinari, 2017). The excerpt presents a similar narrative of examples 1 and 2 analyzed above: migrants frame themselves as victims. Using the war metaphor, migrants are in fact war heroes who sacrifice themselves for their country. The article indicates acceptance of emigration as a necessary evil: Italy is a mother who must lose her children to survive. This feeling of acceptance is however far from

being uncritical; the passage indeed reveals an element of resentment towards Italy for not recognizing the enormous sacrifices Italian migrants were enduring for the greater good. Again, the conveyed message is that of conflicting worldviews in which migrants felt exploited and misunderstood.

Excerpt 6 is taken from *La Sentinella* of 26 October 1929 (emphasis mine).

1. *Un carattere precipuo distingue gl'italiani immigrati in America dagl'italiani emigrati altrove. All'inizio erano tutti accomunati da una sola speranza: **fare o rifare le proprie finanze e rimpatriare**. […] Fra gli italiani d'America **il fine del rimpatrio** è, salvo le eccezioni, **cessato**, per necessità di cose, per logica e diritto di famiglia. […] Qui non vi sono più, e non vi possono essere "italiani d'America, "ma v'è un Gruppo italo-americano, che ha trovato inevitabile costituirsi parte integrante della popolazione "at large," composta di altri sessantasette gruppi di razza. La nostra **italianità**, quindi, è divenuta un sentimento, un **fenomeno, più ampio**. **Noi praticamente non siamo più italiani** da un punto di vista nazionale, ma dal punto di vista della razza, che comprende la nazionalità. I nostri figli, nati in America, sono già più numerosi di noi, e poiché, come dice il Diritto Romano, che, in fondo, sanziona un sentimento naturale, "amor descendit, "**nazionalmente noi ci sentiamo attaccati più alla terra dei nostri figli che a quella dei nostri padri**.*

2. A specific character distinguishes Italians who immigrated to America from Italians who emigrated elsewhere. At the beginning they all had one hope in common: **to raise their finances and repatriate**. […] Among Italian Americans the **purpose of repatriating has**, with a few exceptions, **ceased**, because of necessity, logic and family law. […] Here there are no longer, and there can no longer be "Italians from America, "but there is an Italian-American group, which found it inevitable to form an integral part of the population "at large," made up of sixty-seven other groups of race. Our **Italianness**, therefore, has become a **broader feeling**, a phenomenon. **We practically are no longer Italians** from a national point of view, but from the point of view of race, which includes nationality. Our children, born in America, are already more numerous than us, and since, as Roman Law says "amor descendit, "after all sanctioning a natural feeling, nationally we feel more attached to the land of our children than to that of our fathers.

The excerpt illustrates the identity negotiation struggle of Italian migrants. After a few years in America, migrants' sense of identity has been changed by the migratory experience. They are no longer 'Italians who live in America'; the sentence highlights the opposition between Italy intended as manifestation of a cultural identity ('Italians') and America, intended as a place. The process of transformation has deeply affected their Italianness which has transformed into something bigger. Now migrants' identity incorporates American identity too ('there is an Italian-American group') and this allows them to negotiate inclusion within the wider American society ('integral part of the population at large'). It is interesting to notice how the process of identity negotiation also affected migrants' decision to repatriate. Initially they all wanted to return to Italy, but now this path is chosen only by a few exceptions. This new identity is also passed to the new

generations, thus strengthening a 'natural feeling'. This discourse frame abandons the "us" vs. "them" narrative of the previous years. In this passage, there is no opposition between the in-group and the out-group identity; the complex process of social inclusion is mediated by a newly formed identity which is no longer in crisis.

Discussion

The combination of text mining and CDA allowed for the identification of recurring themes and characteristics in how the experience of migration and remigration was perceived by Italian American migrants. The ngrams analysis of the semantic distribution of salient words such as *emigrazione*, *emigranti*, *rimpatriare*, *ritornare*, *rimesse* suggested that in the ethnic press, repatriation was less frequently discussed than matters of emigration and life in America. Ngrams containing the word 'remittances' scored low in the corpus suggesting a contrasting trend with the Italian public debate of migration in which remittances were at the centre. Moreover, these are found in combination with words such as 'service' or 'ticket' which could indicate an emphasis on practical information and advice in relation to repatriation. This would suggest a stronger preoccupation with the enormous social challenges migrants had to face in America rather than with repatriating. Migrants often struggled to navigate the complexities of a constantly evolving migration landscape, both in Italy (e.g., the law of 1901) and the United States. The finding would therefore also be in line with the historical mission of the ethnic press: primarily to help Italians cope with life in the host country whilst maintaining a bond with their heritage.

The CDA allowed for the triangulation of linguistic data within their social contexts and concurrent socio-historical events. The analysis of excerpts provided deeper insights into how the Italian diasporic community in the United States was narrating its own migratory experience. Specifically, it showed a common pattern of a sense of crisis and identity negotiation manifested through feelings of being misunderstood, rejected and unappreciated and constructed around three main themes: 1) the Italian Government was profiting from emigration; this narrative was found in reference to remittances; 2) Italian migrants were heroes who were sacrificing themselves for the greater good; this narrative was found in reference to the struggles in the host country, including poverty and tuberculosis, and 3) Italian migrants had changed; this narrative was found in reference to returnees or returning to Italy. Generally, the adopted discourse strategy was constructed around an "us" vs. "them" narrative, i.e., Italian migrants vs. the Italian Government but also Italians vs. Americans (particularly in early years) and Italian migrants vs. Italians. The criticism of Italy and the Italian Government as Other-identity was found to be a common communicative device used by Italian migrants to position themselves as the 'sacrificial victims', the heroes who were singlehandedly rescuing the country. Therefore, if in the national debate migrants and emigration were framed as beneficial for the country and migration was oversimplified as an easy way to make money, in the ethnic press the same frame is used to criticize such praises as plainly rhetorical, to characterize migration as an extremely painful process for the individual, and the Italian Government as being unable to factually demonstrate appreciation for the migrants' sacrifice.

The analysis also highlighted how migrants were struggling with processes of inclusion and identity when returning to Italy. Returnees were framed as having lost their 'italianness' and as having been worsened by the migratory experience. They were described as hybrid creatures made up of the worse traits of the two countries. The derogatory image polarizes differences between Italian and American values first and between the so-transformed Italians and Italians after. This narrative is once again in contrast with the national discourse for which repatriated migrants were believed to being improved by the migratory experience and in turn praised for improving Italy. Migrants also described themselves as feeling rejected by the society of origin when returning. The homeland was characterized as ungrateful, again not appreciative of the migrants' sacrifices and feelings. Returnees were on the contrary represented as 'foreign', suggesting that upon return, migrants faced new, great challenges and that these individuals may have experienced remigration as a second crisis of identity. It also supports the second hypothesis of this study, that for many individuals remigration was a bi-directional movement which often did not stop with repatriation.

Processes of identity negotiation and inclusion were found to be a constant preoccupation of the migratory experience. References to Italian and American identity, Italianness, and Italian American identity indicate a continuous search for acceptance and suggest a less material reading of the reasons behind migration, that is beyond economic reasons. Ethnic newspapers revealed that from a cognitive point of view, what made the displacement process particularly painful for the migrant –even when they could enjoy good social and living conditions –was a constant sense of identity crisis, a feeling of rejection, be it from their homeland or the host society. It is only when the migrant made peace with this feeling, only when they finally embraced a new identity that they were no longer in crisis and migration –intended as a painful process –was finally concluded.

Conclusion

At the turn of the twentieth century, the national debate of migration in Italy became entangled with the so-called Southern Question, that is the economic integration of the South. The predominant –though not exclusive –view was that mass emigration was fatalistically necessary to relieve the South from unemployment; moreover, thanks to remittances, emigration was beneficial for the country, as well as remigration, since returning migrants were coming back with money and skills. The exaggerated value of remittances became at the centre of the discourse, almost creating the myth that remittances alone would solve all Italy's problems. This article offered a novel perspective of the migratory experience of Italian migrants by using for the first time ethnic newspapers to compare migrants' discourse of emigration, remigration, and remittances against the national debate of emigration and remigration in Italy. This insider's perspective showed that while the Italian public discourse praised and actively encouraged remigration as positive for the country, both economically and socially, Italian Americans' views were more complex and ambivalent. This more inner perspective – though still mediated –unveiled the identity construction mechanisms part of those strategies which, historically, were

implemented by the Italian diasporic community to construct identity, negotiate inclusion, and maintain power in a hostile environment, may this have been the host or the country of origin. In this way, the study provided a more nuanced and discursive conceptualization of return migration as a bidirectional phenomenon deeply entrenched in identity negotiation processes.

The study also offered a methodological contribution to digital migration studies. A mixed-method approach of distant (ngrams and collocations) and close reading (CDA) was used to explore how migrants themselves were experiencing migration, how their positioning was constructed in relation to the homeland, and how they coped with identity negotiation struggles. Text mining and semantic modeling methods facilitated a more immediate identification of relevant passages, whereas the semantic similarity clustering (ngrams) allowed for a general overview of the distribution of salient words in the corpus. Finally, CDA allowed for a finer-grained analysis of the narratives.

The findings of this study provided a comprehensive understanding of the dynamics, motivations, and challenges associated with migrants returning to their home countries. This knowledge may contribute to effective migration management and policy development, to ensure the well-being of both migrants and the societies they return to, making it highly relevant in the EU and the broader context of global migration. In this respect, the study demonstrated the value of the ethnic press as a novel source that can add an intimate dimension to the study of migration of the past century, of the migratory experiences of those migrant communities and their process of identity negotiation. It highlighted the considerable discrepancy between the exalted domestic discourse about remigration and the more nuanced experience of Italian migrants. Such discrepancy, the article argued, may be seen as an indication that the Italian Government's view of remigration –mainly through the lens of domestic economic advantage –deeply underestimated the complexity of migration as a social phenomenon and as a profoundly changing psychological experience, also for the Italians who never migrated. In the long run, this error of judgment deeply damaged Italy as many of those *ritornati* felt misunderstood, rejected, and disillusioned and crossed the Atlantic again, this time never to return.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://zenodo.org/record/4596345.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fsoc.2023.1239585/full#supplementary-material

References

Abramitzky, R., Boustan, L., and Eriksson, K. (2019). To the New World and Back again: return migrants in the age of mass Mifration. *Ind. Labor Relat. Rev.* 72, 300–322. doi: 10.1177/0019793917726981

Baker, N. (2002). *Double fold: Libraries and the assault on paper. 1. Vintage.* New York, NY: Vintage Books, *27*.

Balletta, F. (1968). *L'emigrazione Italiana Negli Stati Uniti d'America: Le Rimesse Del Risparmio (1914–1925).* Napoli: Università degli Studi di Napoli.

Bandiera, O., Rasul, I., and Viarengo, M. (2013). The making of modern America: migratory flows in the age of mass migration. *J. Dev. Econ.* 102, 23–47. doi: 10.1016/j.jdeveco.2012.11.005

Barrett, J. R., and Roediger, D. (1997). Inbetween peoples: race, nationality and the 'new immigrant' working class. *J. Am. Ethn. Hist.* 16, 3–44. doi: 10.1007/978-1-349-25584-9_9

Baynham, M., and Fina, A.De, eds. (2005). *Dislocations/relocations: Narratives of displacement. Encounters 6.* Manchester: St. Jerome Publ.

Bencivenni, M. (2016). *Italian immigrant radical culture: The idealism of the Sovversivi in the United States, 1890–1940.* New York: New York University Press.

Beneduce, A. (1910). Sul Movimento Dei Rimpatriati Dalle Americhe. *Giornale Degli Economisti e Rivista Di Statistica* 41, 225–258. Available at: http://www.jstor.org/stable/23223474.

Bertani, A. (1886). Atti della Giunta per l'inchiesta agraria sulle condizioni della classe agricola (rist. anast. 1881–86). Vol. 10: Relazione (porto Maurizio, Genova, Massa Carrara). Available at: https://www.unilibro.it/libro/bertani-a-cur-/atti-giunta-inchiesta-agraria-sulle-condizioni-classe-agricola-rist-anast-1881-86-vol-10-relazione-porto-maurizio-genova-massa-carrara-/9788827159606

Bjork, U. (2013). "Ethnic press" in *History of the mass media in the US: An encyclopedia.* ed. M. A. Blanchard (Hoboken: Taylor & Francis), 207–209.

Böhning, W. R. (1975). Basic aspects of immigration and return migration in Western Europe. ILO Working Papers. Available at: https://ideas.repec.org//p/ilo/ilowps/991611463402676.html

Bohning, W. R. (1975). Some thoughts on emigration from the Mediterranean Basin. *Int. Labour Rev.* 111, 251–277.

Bonifazi, C., and Heins, F. (1996). Le Migrazioni Di Ritorno Nel Sistema Emigratorio Italiano: Un Riesame. *Studi Emigrazione* 33, 273–303.

Boyd Betty, C. (1973). *Italian repatriation from the United States, 1900–1914.* New York: Center for Migration Studies.

Braun-Strumfels, L., Marinari, M., and Fiorentino, D., eds. (2023). *Managing migration in Italy and the United States. 1st. 3.* Boston: De Gruyter Oldenbourg.

Brodkin, K. (1998). *How Jews became white folks and what that says about race in America.* New Brunswick, NJ: Rutgers Univ. Press.

Bucki, C. F. (1995). Workers and politics in the Immigrant City in the early twentieth-century United States. *Int. Labor Work Class Hist.* 48, 28–48. doi: 10.1017/S0147547900005329

Carpi, L. (1871). *Dell'emigrazione italiana all'estero nei suoi rapporti coll'agricoltura, coll'industria e col commercio.* Stabilimento Giuseppe Civelli. Africa

Carpi, L. (1878). *Statistica illustrata della emigrazione all'estero del triennio 1874–76 nei suoi rapporti economico-sociali monografia di Leone Carpi* tip. del Popolo romano. Roma

Cerase, F. P. (1967). A study of Italian migrants returning from the U.S.a. *Int. Migr. Rev* 1, 67–74. doi: 10.2307/3002741

Cerase, F. P. (1974). Expectations and reality: a case study of return migration from the United States to southern Italy. *Int. Migr. Rev* 8, 245–262. doi: 10.2307/3002783

Choate, M. I. (2008). *Emigrant nation: The making of Italy abroad.* Cambridge, Mass: Harvard University Press.

Cibotti, E. (1994). Giornalismo Politico e Politica Giornalistica. La Formazione Pubblica Di Una Opinione Italiana Nella Buenos Aires Di Fine Secolo. *Ventesimo Secolo* 4, 351–376.

Commissariato Generale dell'Emigrazione. (1925). *Le Rimesse Degli Emigranti. Bollettino Dell'Emigrazione.* Italy Commissariato Generale dell'Emigrazione.

Connell, W. J., and Gardaphé, F. L., eds. (2010). *Anti-Italianism: Essays on a prejudice.* New York: Palgrave Macmillan.

Crispi, F. (1915). *Discorsi parlamentari* Tip. della Camera dei deputati. Roma

De Fina, A., and Tseng, A. (2017). "Narrative in the study of migrants" in *The Routledge handbook of migration and language.* ed. S. Canagarajah (London: Routledge), 381–396.

Deschamps, B. (2011). "In the cultures of Italian migration" in *Diverse trajectories and discrete perspectives.* eds. G. Parati and A. J. Tamburri (Madison [N.J.]; Lanham, Md: Fairleigh Dickinson University Press; Rowman & Littlefield), 75–94.

Deschamps, B. (2020). *Histoire de la presse italo-américaine: du Risorgimento à la Grande guerre. Aire anglophone.* Paris: l'Harmattan.

di Cosentino, M. (1873). *Uno Sguardo Alla Emigrazione Italiana Ed Estera.* Rome: Pallotta.

Dustmann, C., and Görlach, J.-S. (2016). The economics of temporary migrations. *J. Econ. Lit.* 54, 98–136. doi: 10.1257/jel.54.1.98

Fagan, B. (2016). Chronicling white America. *American Period. J. Hist. Criticism* 26, 10–13. doi: 10.17613/M6VD5X

Fairclough, N., and Wodak, R. (1997). "Critical discourse analysis." In *Discourse studies: A multidisciplinary introduction* (Vol. *2*)., edited by DijkT. van. London: SAGE.

Foley, N. (1997). *The white scourge: Mexicans, blacks, and poor whites in Texas cotton culture.* Berkeley: Univ. of California Press.

Gagniuc, P. A. (2017). *Markov chains: From theory to implementation and experimentation.* Hoboken, NJ: John Wiley & Sons.

Gatto, A. (2021). Rimesse Dagli Emigrati Italiani Negli USA, Sviluppo e Cicli Economici: Dalle Fonti Archivistiche Del Banco Di Napoli Ai Dati Della Banca Mondiale (1861-2017). *RiMe. Rivista Dell'Istituto Di Storia Dell'Europa Mediterranea 8/ II*, 281–294. doi: 10.7410/1469

Gentile, E. (1986). L'emigrazione italiana in Argentina nella politica di espansione del nazionalismo e del fascismo. *Stor. Contemp.* 3, 355–396.

Gentileschi, M. L., and Simoncelli, R. (1983). *Rientro Degli Emigrati e Territorio: Risultati Di Inchieste Regionali.* Napoli: Istituto Grafico Italiano

Gould, J. D. (1980). European inter-continental emigration. The road home: return migration from the USA. *J. Eur. Econ. Hist.* 9:41. Available at: https://api.semanticscholar.org/CorpusID:126785083.

Guglielmo, T. A. (2004). *White on arrival: Italians, race, color, and power in Chicago, 1890–1945.* New York: Oxford University Press.

Guglielmo, J., and Salerno, S. (2003). *Are Italians white?: How race is made in America.* New York: Routledge.

Hickerson, A., and Gustafson, K. L. (2016). Revisiting the immigrant press. *J. Theory Pract. Criticism* 17, 943–960. doi: 10.1177/1464884914542742

Jacini, S. (1885). *Inchiesta Jacini. Atti della giunta per la inchiesta agraria e sulle condizioni delle classi agricole.* Università degli Studi di Firene Florence.

Jacobson, M. F. (1998). *Whiteness of a different color: European immigrants and the alchemy of race.* Cambridge; London: Harvard University Press.

Kubát, D., and Center for Migration Studies (U.S.), eds. (1984). *The politics of return: international return migration in Europe: proceedings of the first European conference on international return migration (Rome, November 11-14, 1981)*. Roma, New York; Center for Migration Studies.

LaGumina, S. J. (1999). *WOP! A documentary history of anti-Italian discrimination in the United States*. Toronto: Guernica.

LaGumina, S. J. (2018). "Discrimination, prejudice and Italian American history" in *The Routledge history of Italian Americans*. eds. W. J. Connell and G. Stanislao (New York and London: Routledge).

Manzotti, F. (1962). La Polemica Sull'emigrazione Nell'Italia Unita: (Fino Alla Prima Guerra Mondiale). Milano: Soc. Ed. D. Alighieri. https://books.google.it/ books?id=aWYRAQAACAAJ

Molinari, A. (2017). Le traversate delle migrazioni storiche italiane tra evento e racconto. *História (São Paulo)* 36:16121. doi: 10.1590/1980-436920160000000112

Musillo, I. (1981). Retour et Emploi Des Migrants Dans Le Mezzogiorno: Enquête Sur Un Échantillon de Migrants Italiens. ILO Working Papers. Available at: https://ideas. repec.org//p/ilo/ilowps/992106593402676.html

Oberbichler, S., and Pfanzelter, E. (2022). "Tracing discourses in digital newspaper collections" in *Digitised newspapers – A new Eldorado for historians?* eds. E. Bunout, M. Ehrmann and F. Clavert (Berlin: De Gruyter), 125–152.

Oberbichler, S., and Viola, L. (2023). "National Narratives and immigrant voices: transatlantic return migration to Italy and Austria, 1850-1950" in *Return and circular migration in contemporary European history*. ed. V. Larcher (Berlin: De Gruyter Mouton).

Ostuni, M. R.. (2001). "Leggi e Politiche Di Governo Nell'Italia Liberale e Fascista." In *Storia Dell'emigrazione Italiana*, edited by P. Bevilacqua, ClementiA. De and I. Emilio Franzina 309–319. Roma: Donzelli.

Padovani, E. (1909). *Le Condizioni Igienico-Sanitarie Dell'emigrazione Transoceanica Italiana. Rivista Di Emigrazione*.

Park, R. E. (1922). *The immigrant press and its control. New York*; London: Harper & Brothers.

Pozzetta, G. E. (1973). The Italian immigrant Press of New York City; the early years, 1880–1915. *J. Ethnic Stud.* 1, 32–46.

Reisigl, M., and Wodak, R. (2001). "The discourse-historical approach" in *Methods of critical discourse analysis*. eds. R. Wodak and M. Meyer (London: Sage), 63–94.

Rhodes, L. D. (2010). *The ethnic press: Shaping the American dream*. New York: Lang.

Rogers, R. (1983). Return migration in comparative perspective. *Int. Migr. Rev.* 17, 277–299. doi: 10.1177/019791838301701S40

Rosoli, G. (1977). L'emigrazione Di Ritorno: Alla Ricerca Di Una Impostazione. *Studi Emigrazione* 14, 235–246.

Rumsey, A. S. Digital Library Federation. (2001). *Strategies for building digitized collections*. Washington, D.C: Digital Library Federation, Council on Library and Information Resources.

Russo, P. (1972). "La Stampa Periodica Italo-Americana" in *Gli Italiani Negli Stati Uniti. L'emigrazione e l'opera Degli Italiani Negli Stati Uniti d'America*. ed. R. J. Vecoli (Florence: Istituto di studi americani), 495–498.

Saloutos, T. (1956). *They remember America: The story of the repatriated Greek-Americans*. UCL Press, London.

United States Bureau of Immigration and Naturalization. (1908). *Annual report of the commissioner-general of immigration to the secretary of commerce and labor for the fiscal year ended*. Washington, DC. U.S. Government Printing Office.

Vecoli, R. J. (1998). "The Italian immigrant press and the construction of social reality" in *Print culture in a diverse America*. eds. J. P. Danky and W. A. Wayne (Urbana, Chicago Springfield: University of Illinois), 17–33.

Vecoli, R. J. (2006). Italian immigrants and working-class movements in the United States: a personal reflection on class and ethnicity. *J. Can. Hist. Rev.* 4, 293–305. doi: 10.7202/031067ar

Vellon, P. G. (2017). *A great conspiracy against our race: Italian immigrant newspapers and the construction of whiteness in the early twentieth century*. New York and London: New York University Press.

Viola, L. (2019). "Polentone vs Terrone: a discourse-historical analysis of media representation of Italian internal migration" in *Migration and media: Discourses about identities in crisis*. eds. L. Viola and A. Musolff (Amsterdam: John Benjamins), 45–90.

Viola, L. (2021). Chronicl Italy and Chronicl Italy 2.0: digital heritage to access narratives of migration. *Int. J. Humanit. Arts Comput.* 15, 170–185. doi: 10.3366/ijhac.2021.0268

Viola, L. (2022). 'Italy, for example, is just incredibly stupid now'. European crisis narrations in relation to Italy's response to COVID-19. *Front. Commun.* 7:757847. doi: 10.3389/fcomm.2022.757847

Viola, L. (2023). *The humanities in the digital: Beyond critical digital humanities*. London: Palgrave Macmillan.

Viola, L., and Fiscarelli, A. M. (2021). "*Chronicl Italy 3.0. A deep-learning, contextually enriched digital heritage collection of Italian immigrant newspapers published in the USA, 1898-1936.*" Zenodo Switzerland.

Viola, L., and Musolff, A. (2019). *Migration and media: Discourses about identities in crisis*. Amsteram: John Benjamins.

Viola, L., and Verheul, J. (2019a). Mining ethnicity: discourse-driven topic modelling of immigrant discourses in the USA, 1898–1920. *Digit. Scholarsh. Humanit.* 35, 921–943. doi: 10.1093/llc/fqz068

Viola, L., and Verheul, J. (2019b). The media construction of Italian identity: a transatlantic, digital humanities analysis of Italianità, ethnicity, and whiteness, 1867-1920. *Identity* 19, 294–312. doi: 10.1080/15283488.2019.1681271

Viola, L., and Verheul, J. (2020a). "*Machine learning to geographically enrich understudied sources: A conceptual approach:*" in proceedings of the 12th international conference on agents and artificial intelligence, 469–75. Valletta, Malta: SCITEPRESS-Science and Technology Publications.

Viola, L., and Verheul, J. (2020b). One hundred years of migration discourse in the times: a discourse-historical word vector space approach to the construction of meaning. *Front. Artif. Intell.* 3:64. doi: 10.3389/frai.2020.00064

Wong, A. S. (2006). "Race and the nation in Liberal Italy, 1861-1911: Meridionalism, empire, and diaspora" in *Italian and Italian American studies*. ed. S. G. Pugliese. *1st* ed (New York: Palgrave Macmillan)

Wyman, M. (1993). *Round-trip to America: The immigrants return to Europe, 1880–1930*. Ithaca, N. Y.: Cornell University Press.