

Chapter 6

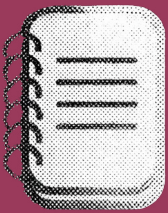
Data traces and the inevitable visibility of irregular migration

Alejandra Rodríguez Sánchez

To cite: Rodríguez-Sánchez, A. (2025). Data traces and the inevitable visibility of irregular migration. In D. Kierans and A. Kraler (eds), *Handbook on Irregular Migration Data. Concepts, Methods and Practices*. Krems: University of Krems Press.
<https://doi.org/10.48341/g31s-vq79-06>

Keywords: Data traces, Metadata, Digital footprints, Indirect measurement, Visibility

Data traces and the inevitable visibility of irregular migration



Key points

- Irregular migrants, like all people, leave behind data traces – through surveys to digital activity and use of public services. These traces can indirectly signal their presence, even when their legal status remains unknown or unrecorded.
- This chapter explores how metadata and alternative data sources, such as social media, mobility records, and mortality registers can complement traditional statistics and offer new ways to estimate irregular migration.
- Pilot studies from the MIRreM project show that while these approaches are limited and ethically complex, they can be adapted and combined to improve understanding of irregular migrants, including their presence, behaviours and needs.

Introduction

One of the greatest challenges in researching irregular migration is the lack of reliable and up to date information on migrants and their legal status. For example, existing estimating methods to learn about irregular migration stocks and flows and the estimates they produce lack a reference value that can serve as a “gold standard” precisely because data on irregular migrants is incomplete. In addition to the inherent difficulty of studying mobile populations, researching individuals with irregular status is further complicated by the risks they face if their identity or legal status is revealed. The threat of deportation and the dangers of being forced to return to a place they may have

fled for economic, political, or personal reasons create strong, legitimate disincentives to sharing information on their legal status.

Nevertheless, despite these concerns, data on migrants with irregular status is passively extracted, and collected by researchers, civil organizations, private companies, and immigration authorities, on a daily basis. This chapter offers a brief reflection on the differences and similarities in which this kind of data is being extracted, the potential uses researchers can make of this information, and especially the approaches we employed in MIRreM.

Metadata, data about data that offers context rather than direct description, can, in the context of migration, though not originally intended for migration research, serve as valuable indirect indicators of irregular migration patterns, behaviours, and trends. In this way, metadata can act as a complement to traditional sources like surveys, censuses and population registers to enhance our understanding of this complex issue. Like everyone else, irregular migrants leave traces of their presence within a country, especially in today's highly digital and technologically mediated environments. Whether making purchases,

applying for a driver's license, visiting a doctor, responding to a government survey, or logging onto social media, irregular migrants generate records that document their presence—often without their knowledge or intent. To better understand the population size of irregular migrants, as well as their needs and living conditions, researchers can tap into the data generated by such interactions, in combination with official data. Essentially, this viewpoint aligns in part with indirect demographic methods but extends beyond them by embedding the indirectness within the data itself, rather than relying on underlying assumptions or parameters.

Box 6.1: Metadata in the context of migration

Alejandra Rodríguez Sánchez

Metadata refers to data about data—information that provides context, structure, or insight into how data is produced, rather than describing the subject directly. In the context of migration, metadata includes digital traces such as social media logins, geolocations, IP addresses, mobile network records, and internet searches. These data can signal movement patterns, border crossings, or changes in residence. Though not originally intended for migration research, such metadata—often derived from social media platforms, telecommunications networks, or administrative systems—can serve as valuable indirect indicators of irregular migration flows, behaviours, and trends. Even traditional sources like censuses, population registers, or household surveys can be viewed through this lens, offering a way to connect diverse data approaches and better understand the complex challenge of estimating irregular migration.

Mixed approaches combining direct and indirect data on migrants

The presence of irregular migrants is generally unknown to authorities. Contact between irregular migrants and immigration authorities may occur through various means, such as encounters at entry points using forged documents, random border checks, sea rescues, police operations, or interactions with other government agencies. However, the completeness of data collected through these operations has been criticized for relying heavily on the number of border personnel, thereby conflating enforcement activity with the actual presence of migrants (Savatic et al., 2024).

Population registers are another important source of official data on migrants. A key example is Spain's

registration system, known as *empadronamiento*, which all residents are required to complete to access social services (Eniquez, 2019). However, due to recent legal changes, many irregular migrants may choose not to register, limiting the coverage of this data source, thus highlighting the voluntary aspect of participation (Velasco, 2021). Additionally, the use of data from registration systems is often restricted by data protection laws, which can prohibit the use of personally identifiable information or prevent the lawful matching of records across systems.

One of the most common ways researchers gather information on irregular migrants is through

sampling and household or individual surveys—an approach that has been used in past studies (Rodríguez-Sánchez & Tjaden, 2023; Jandl, 2011). These efforts often rely on institution-led data collection, such as censuses or household surveys. However, the legal status of migrant respondents—who belong to a hard-to-reach demographic—is frequently unknown, making the use of probabilistic sampling difficult. Surveys rarely include direct questions about legal status, as doing so may lower response rates or result in incomplete data (Bachmeier et al., 2014;

Young & Madrigal, 2017). Notwithstanding these difficulties, other methods, such as the indirect method, which is used in a standard fashion to estimate the number of unauthorized migrants in the United States, relies on data from the American Community Survey (ACS), a large-scale household survey that includes all migrants. However, ACS underrepresents immigrants, especially newly arrived ones. Although efforts are made to adjust for this underrepresentation in the estimation of irregular migrants employing the residual method (van Hook et al., 2015).

Metadata, indirect approaches, and irregularity

Metadata can help in detecting and tracking population groups without requiring their direct participation or consent. For populations that actively avoid detection or are unlikely to participate in traditional surveys, such as irregular migrants or other marginalized communities, metadata may provide the only viable method for demographic observation and measurement.

Viewed from a certain perspective, this kind of information is already in use, though it is rarely labelled in these terms. For example, when the legal status of immigrants is not known, the use of statistical imputation has been employed to learn about the size and characteristics of irregular migrants. Statistical and machine learning methods can be used to predict what would be the most likely status of respondents based on the responses migrants give to other questions deemed to be less sensitive (Van Hook et al., 2015; Borjas & Cassidy, 2019), especially when such algorithms are trained in complete data (i.e., data capturing more direct evidence of irregularity). Therefore, even when not intending to answer questions about legal status, information provided by the correlation between legal status and other attributes of respondents can lead to plausible estimations.

The search for potential correlates or statistical signals of irregularity within existing data is an interesting approach. A compelling example of why it is helpful to view data on irregular migrants as metadata emerges when analysing the “digital traces” left by migration behaviour. Data from social media, internet searches, call detail records, economic transactions, and similar sources—often referred to in the recent past as “big data”

sources to emphasize their greater volume, variety, variability and velocity of generation (Sirbu et al., 2021), especially in migration (Tjaden, 2021)—are not originally created to study migration. Rather, this type of digital data arises from commercial or operational activities primarily intended for business or statistical purposes.

For example, META, a company owning some of the largest social media platforms in the world, offers the possibility to run targeted ads campaigns for marketing, political or civic purposes on Facebook, Instagram, Messenger or the Audience Network. To get these ads to the people most likely to be interested in the products or services advertised, META classifies its user base in various demographics, often without the awareness of the users themselves.

Attributes such as gender, age, location, interests, and places where the user has been or lived in the past, etc. are used to create targeted groups. It is this last attribute that researchers have employed to derive additional information on the stock of migrants in a given country (Zaghenei, Weber & Gummadi, 2017). Recently, based on individual-level data, Chi et al. (2025) used Facebook profile locations to identify users who changed countries. If the change lasted most of the year, the authors classified users as migrants to the specific destination, while going beyond existing work on stocks by pairing the definition of migrant using duration of stay as a further criterion. In this way, researchers can circumvent asking respondents to disclose information about them and instead gather this information from behavioural clues.

Pilot studies in MIrreM

In a series of pilot studies developed in the context of MIrreM, we have made use of a similar intuition regarding metadata to develop innovative methodologies to estimate irregular migration. We used data from the following sources: Facebook users (Rodríguez-Sánchez & Tjaden, 2025a); digital surveys on US based Mexicans and Venezuelan users of social media platforms (Tjaden & Rodríguez-Sánchez, 2025); air passenger data collected on all flights across the world towards and outside of Europe (Bernasconi & Recchi, 2025); information on employment conditions of immigrants in the UK (Salihoğlu & Vargas-Silva, 2025); the effects of changes in laws governing access to healthcare through the National Health Institute (NHS) in the UK (Rodríguez-Sánchez & Tjaden, 2025b); and a matching of official registers capturing mortality and population in Belgium (Surkyn & Bircan, 2025), a method that shares many more commonalities with traditional approaches as discussed in Chapter 5).

The results of these pilot studies advance the literature in this domain and connect to a growing literature employing innovative data sources and methodologies to tackle the challenges of estimating irregular migration (Rodríguez-Sánchez & Tjaden, 2023).

Tjaden and Rodríguez-Sánchez (2025) demonstrated that Facebook ads can effectively reach irregular migrants and that list experiments provide more reliable estimates of legal status than direct questions in the United States context, particularly among well-established immigrant groups such as Mexicans residing in the US.

Salihoğlu and Vargas-Silva (2025) demonstrated that some of the conceptual and measurement challenges in studying irregular migrants can be addressed through analysis of the informal

economy and migrants' characteristics, using a clear conceptual framework and probabilistic tools such as national labour force surveys from Turkey and the UK to estimate their presence within it.

Rodríguez-Sánchez and Tjaden (2025b), in turn, showed that healthcare reforms in the UK resulted in a slight decline in new GP registrations at practices serving large migrant populations, which, when combined with arrival data at local levels, could be used to estimate undocumented migration flows using a multiplier approach.

Bernasconi and Recchi (2025) analysed net air travel flows in the Schengen area, using passenger data and adjusted official net migration figures, providing estimates of irregular inflows by region of origin for 2019, closely matching the notion of visa overstayers.

Drawing on Belgian population register data, Surkyn and Bircan's study (2025) shows that mortality rates can serve as a robust indicator for estimating the size and changes of irregular migrants over time, providing detailed insights by gender, age group, and even region of origin.

Finally, Rodríguez-Sánchez and Tjaden (2025a), in turn, showed that Facebook stocks of migrants, when examined by means of predictive modelling and machine learning, can provide hints at the hidden numbers not measured by official migrant stocks, offering a global comparison of irregular migrant stocks.

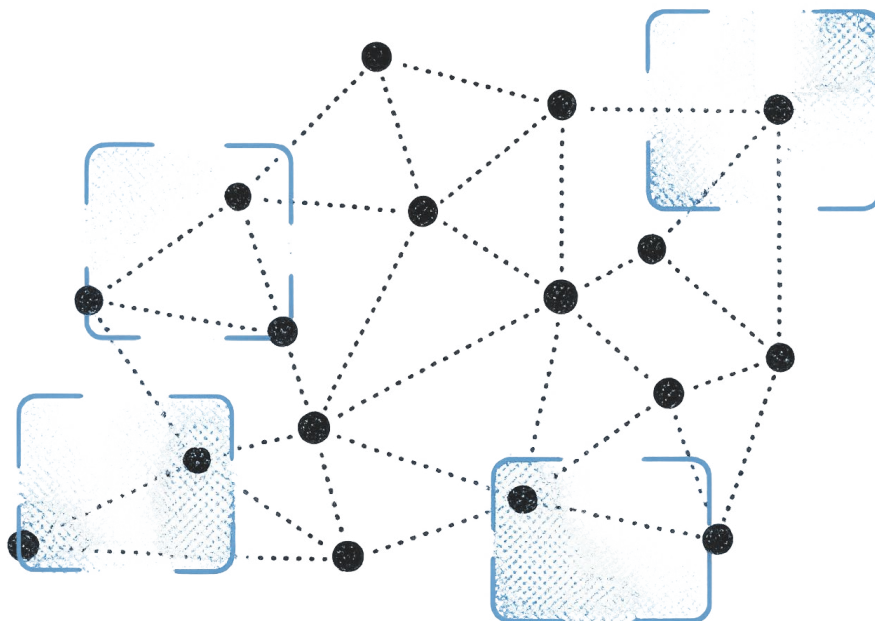
Conclusion

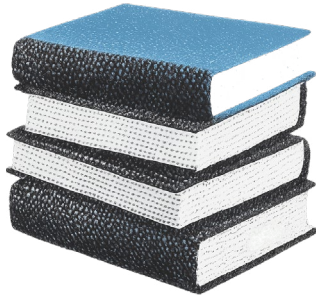
Computational approaches, of which digital data is an important component, have already enriched migration research (Drouhot et al., 2023). Thinking about data about migrants without a legal status as meta-data is helpful in understanding the information and in the development of further methods to estimate the number of irregular migrants. This type of data also comes with important limitations. The definition of what constitutes “migration”, “place of birth”, “migration status” etc. may vary across the different sources of metadata that exist and, importantly, may not be comparable to official or research standard definitions. These alternative data sources have not been created for research purposes.

Despite the insights generated by metadata, one major limitation is the inability to learn something more about the demographics and living conditions of the population of irregular migrants thus

estimated. Moreover, there are important risks associated with employing what previous research as defined as “footprints” when not put into the larger context of statistical information on other population statistics (Gelatt, Fix, & van Hook, 2018), such as information on birth, death, school enrolment, housing, and other records.

Assessing when metadata is generated, whether it results from voluntary or involuntary actions, and understanding the potential coverage of such alternative data sources are key steps in determining how much insight can be gained from using innovative data in migration research. While each of these data traces can only offer a partial view of migrants with an irregular status, together these different sources of information underscore the inescapable visibility of irregular migrants and the potential to better understand their presence and the challenges such communities are facing.





References

- Bachmeier, J. D., Van Hook, J., & Bean, F. D. (2014). Can we measure immigrants' legal status? Lessons from two US surveys. *International Migration Review*, 48(2), 538-566. <https://doi.org/10.1111/imre.12059>
- Borjas, G. J., & Cassidy, H. (2019). The wage penalty to undocumented immigration. *Labour Economics*, 61, 101757. <https://doi.org/10.1016/j.labeco.2019.101757>
- Drouhot, L. G., Deutschmann, E., Zuccotti, C. V., & Zagheni, E. (2023). Computational approaches to migration and integration research: promises and challenges. *Journal of Ethnic and Migration Studies*, 49(2), 389-407. <https://doi.org/10.1080/1369183X.2022.2100542>
- Enríquez, C. G. (2019). Inmigración en España: una nueva fase de llegadas. *Análisis del Real Instituto Elcano (ARI)*, (28), 1.
- Savatic, F., Thiollet, H., Mesnard, A., Senne, J. N., & Jaulin, T. (2024). Borders Start With Numbers: How Migration Data Create "Fake Illegals". *International Migration Review*, 59(3), 1432-1463. <https://doi.org/10.1177/01979183231222169>
- Rodriguez Sanchez, A., & Tjaden, J. (2023). Estimating Irregular Migration – A Review of Traditional and Innovative Methods. In *MIRreM Working Paper No.4 (version 2)*. Krems: University for Continuing Education Krems (Danube University Krems). <https://doi.org/10.5281/zenodo.8380854>
- Sîrbu, A., Andrienko, G., Andrienko, N., Boldrini, C., Conti, M., Giannotti, F., ... & Sharma, R. (2021). Human migration: The big data perspective. *International Journal of Data Science and Analytics*, 11, 341-360. <https://doi.org/10.1007/s41060-020-00213-5>
- Tjaden, J. (2021). Measuring migration 2.0: a review of digital data sources. *Comparative Migration Studies*, 9(1), 59. <https://doi.org/10.1186/s40878-021-00273-x>
- Van Hook, J., Bachmeier, J. D., Coffman, D. L., et al. (2015). Can we spin straw into gold? An evaluation of immigrant legal status imputation approaches. *Demography*, 52, 329-354. <https://doi.org/10.1007/s13524-014-0358-x>

Velasco, F. R. (2021). El arduo camino hacia la universalidad de la asistencia sanitaria de los inmigrantes irregulares en España. *e-Revista Internacional de la Protección Social*, 6(1), 343-369. <http://dx.doi.org/10.12795/e-RIPS.2021.i01.16>

Young, M. E. D. T., & Madrigal, D. S. (2017). Documenting legal status: A systematic review of measurement of undocumented status in health research. *Public Health Reviews*, 38, 1–25. <https://doi.org/10.1186/s40985-017-0073-4>

Zagheni, E., Weber, I., & Gummadi, K. (2017). Leveraging Facebook's advertising platform to monitor stocks of migrants. *Population and Development Review*, 721-734. <https://doi.org/10.1111/padr.12102>