MIRREM

Measuring Irregular Migration

www.irregularmigration.eu

WORKING PAPER ON IRREGULAR MIGRATION FLOWS

MIrreM Working Paper 9/2024

MAIN AUTHORS:

Lalaine Siruno

Arjen Leerkes

Contributors:

Maegan Hendow

Ema Brunovská

Funded by the European Union

Co-funded by:

UK Research and Innovation

Toronto Metropolitan University — Canada Excellence Research Chair in Migration & Integration

Deliverable Information:

Project Acronym:	Measuring irregular migration and related policies (MIrreM)
Project No.	101061314
WP	WP5 – Irregular migration flows
Deliverable Type:	R – Document, report
Deliverable Name	D5.1 Working paper on irregular migration flows
Version:	1.0
Date:	22/04/2024
Responsible Partner:	Maastricht University (UM)
Contributing Partner:	International Centre for Migration Policy Development (ICMPD)
Authors	Lalaine Siruno (UM) and Arjen Leerkes (UM)
Contributors	Maegan Hendow (ICMPD) and Ema Brunovská (ICMPD)
Reviewers:	Ann Singleton (University of Bristol) and Denis Kierans (UOXF)
Dissemination Level:	Public

Revision History:

Version	Date	Author	Organisation	Description
1	25/02/2024	Lalaine Siruno	UM	Initial draft
1	05/03/2024	Lalaine Siruno Arjen Leerkes	UM	Second draft
1	22/03/2024	Ann Singleton Maegan Hendow Denis Kierans	University of Bristol ICMPD UOXF	Review
1	25/03/2024	Lalaine Siruno Adriana Harm	UM UWK	Revision Formatting check
1	28/03/2024	Lalaine Siruno Arjen Leerkes	UM	Revision and final version

Table of contents

LIST OF TABLES

LIST OF FIGURES

THE MIRREM PROJECT

MIrreM examines estimates and statistical indicators on the irregular migrant population in Europe as well as related policies, including the regularisation of migrants in irregular situations.

MIrreM analyses policies defining migrant irregularity, stakeholders' data needs and usage, and assesses existing estimates and statistical indicators on irregular migration in the countries under study and at the EU level. Using several coordinated pilots, the project develops new and innovative methods for measuring irregular migration and explores if and how these instruments can be applied in other socio-economic or institutional contexts. Based on a broad mapping of regularisation practices in the EU as well as detailed case studies, MIrreM will develop 'regularisation scenarios' to better understand conditions under which regularisation should be considered as a policy option. Together with expert groups that will be set up on irregular migration data and regularisation, respectively, the project will synthesise findings into a Handbook on data on irregular migration and a Handbook on pathways out of irregularity. The project's research covers 20 countries, including 12 EU countries and the United Kingdom.

TO CITE:

Siruno, L., Leerkes, A., Hendow, M. & Brunovksá, E. 2024. Working Paper on Irregular Migration Flows. MIrreM Working Paper No. 9. Krems: University for Continuing Education Krems (Danube University Krems). https://doi.org/10.5281/zenodo. 10702228.

KEYWORDS

Irregular migration flows; flow estimates; flow indicators, Eurostat, data and policymaking

ACKNOWLEDGEMENTS

This working paper was written by Lalaine Siruno and Arjen Leerkes, with contributions from Meagan Hendow and Ema Brunovska. We gratefully acknowledge support from the rapporteurs who collected country-level data, which will be fully available in the forthcoming *MIrreM Public Database on Irregular Migration Flow Estimates and Indicators*, as well as from Ann Singleton who provided comments on this piece; and Denis Kierans, for additional feedback. We also thank Arsh Ajmera and Giulia Mori, interns at UNU-MERIT/University of Maastricht who provided research assistance. Finally, we would like to express appreciation to everyone who participated in an interview and shared valuable insights on measuring irregular migration flows and the use of data in policymaking. We strive for accuracy, but the complexity of the subject matter may result in unintentional inaccuracies or oversights. Any errors or omissions remain the sole responsibility of the main authors and do not reflect the views of any of the institutions or organisations mentioned in this paper.

FUNDING ACKNOWLEDGEMENT

Funded by the European Union. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or the Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Executive Summary

The **objective** of this working paper is to provide a methodological assessment of available data on irregular migration flows in Europe and other contexts. It is aimed at the public, including those who may not have a technical understanding of migration statistics. It also serves as a companion document to the *MIrreM Public Database on Irregular Migrant Flow Estimates and Indicators* (MIrreM D5.2).[1] In MIrreM, **irregular migration** is operationally defined as a form of migration that is not "regular," "unlawful," or not according to the rules (Kraler & Ahrens, 2023).

The paper addresses this **research question**: How are irregular migration flows measured and estimated in the EU and selected other countries and what can be said about the quality of the available indicators and estimates? Based on the qualitative and quantitative analysis of a broad range of data sources, the **main conclusions** are as follows:

1) Compared to stocks, there are still almost no available estimates of irregular migration flows. However, in view of findings from the CLANDESTINO Project, there are now more irregular flow indicators, particularly for geographic flows, and to some extent, also asylum-related status flows.
2) Eurostat data as a whole is better than (individual) country-level data in terms of accessibility, documentation, and validity and reliability. Among the countries covered in MIrreM, data from the US is arguably the most valid and reliable, which also includes some flow estimates. Data from European countries are generally comprehensive, but they are difficult to compare at the country level. For "transit countries," data appear to be highly fragmented except for Bosnia and Herzegovina.
3) Zooming in on Eurostat data, asylum data, though with their own limitations, seem to exhibit the highest level of validity and reliability. Despite reporting lags due to the unpredictable timing of first instance and final decisions, negative decisions on asylum applications serve as a meaningful status-related flow indicator.
4) However, there are still significant issues concerning the overall quality of Eurostat data, not so much concerning the FAIR (Findability, Accessibility, Interoperability, and Reusability) Data Principles, but with validity and reliability.
5) Policymakers use irregular migration data for purposes extending beyond the monitoring of irregular migratory flows. However, the ambiguity surrounding definitions and the lack of consensus on what precisely constitutes irregular migration even among migration scholars, practitioners, and other stakeholders, increases the risks associated with the use of irregular migration data for political purposes.
6) Good quality data is essential to migration governance. However, the interest in enhancing data collection on irregular migration and generating estimates must be carefully weighed against privacy considerations and societal interests.

[1] Siruno, L., Leerkes, A., Badre, A., Bircan, T., Brunovská, E., Cacciapaglia, M., Carvalho, J., Cassain, L., Cyrus, N., Desmond, A., Fihel, A., Finotelli, C., Ghio, D., Hendow, M., Heylin, R., Jauhiainen, J.S., Jovanovic, K., Kierans, D., Mohan, S.S., Nikolova, M., Oruc, N., Ramos, M.P.G., Rössl, L., Sağiroğlu, A.Z., Santos, S., Schütze, T., & Sohst, R.R. (2024). MIrreM Public Database on Irregular Migration Flow Estimates and Indicators. Krems: University for Continuing Education Krems (Danube University Krems). https://doi.org/10.5281/zenodo.10813413.

Going forward, there is a need to continue improving the quality of data for specific flow indicators, to acknowledge that supplemental qualitative information is essential for the validation and triangulation of quantitative data, and to consider using accessible informational resources to mitigate the misuse of migration data for political purposes.

1. INTRODUCTION

1.1 MIGRATION GOVERNANCE AND MEASURING IRREGULAR MIGRATION FLOWS

This working paper has been produced in the context of the EU-funded Horizon Europe MIrreM Project (Measuring Irregular Migration and Related Policies)[2]. MIrreM builds on the seminal CLANDESTINO Project (Undocumented Migration: Counting the Uncountable: Data and Trends Across Europe)[3] which was also funded by the EU. CLANDESTINO collected and classified data on irregular migration for 12 EU countries covering the period 2000 to 2007, and it concluded its final report as follows (Jandl et al., 2008, p. 17):

> The review of efforts to estimate the size of irregular migration on a European level has shown that the numbers indicated are based on very rough estimates. Often, we do not know which groups of irregular migrants are in [sic] included in a stock estimate, nor we do not know whether a flow estimate is meant to measure net inflows or gross inflows (without substraction [sic] of outflows).

Jandl (2008, p. 20) further pointed out that:

> ...Given the highly volatile nature of migration flows, the scarcity of reliable indicators on illegal migration flows, and the dearth of appropriate methods for estimating such flows, most efforts have concentrated on estimating stocks of undocumented migrants rather than flows.

Now over a decade since CLANDESTINO, and with controlling irregular migration flows a mainstay policy priority in the EU and other countries, this working paper analyses more recent data related to migrant irregularity.[4] Together with MIrreM Work Package 4 (WP4) on irregular migrant stocks, the objective is to assess estimates of irregular migration flows, if available, and to determine the quality of available statistical indicators on multiple geographical scales.

Immigration, particularly irregular migration, has become highly politicised due to its intersection with issues such as national security, economic concerns, and cultural identity. Public discourse often focuses on the perceived "flows" of migrants, framing migration as a

[2] https://irregularmigration.eu/
[3] https://irregular-migration.net/
[4] The notion of migrant irregularity is often contested. For a comprehensive discussion, please refer to Kraler, A., & Ahrens, J. (2023). *Conceptualising migrant irregularity for measurement purposes, MIrreM Working Paper 2/2023*. Zenodo. https://doi.org/10.5281/zenodo.7868237.

problem to be managed rather than a complex social phenomenon. Stocks and flows are the most commonly used migration data (Yildiz & Abel, 2021). Migrant **stocks** refer to the total number of migrants residing in a particular location at a specific point in time, offering a snapshot of the migrant population. Migrant **flows**, on the other hand, represent the movement of migrants over a defined period, capturing arrivals, departures, and net migration, providing a dynamic perspective on migration patterns and trends. In other words, flows are the events that increase (**inflows**) or decrease (**outflows**) migrant stock within a given territory during a certain period (Rodriguez Sanchez & Tjaden, 2023).

Good data is crucial in migration governance for several reasons, not least because of the aforementioned highly politicised nature of irregular migration. Firstly, it can help policymakers make informed decisions by providing accurate insights into migration patterns, trends, and demographics. Secondly, it can aid in designing effective policies and programs that address the specific needs of migrants and host communities, promoting integration and social cohesion. Additionally, reliable data facilitates monitoring and evaluation of migration policies, allowing for adjustments and improvements based on evidence rather than assumptions. Overall, good data plays a fundamental role in ensuring that migration governance is responsive, fair, and beneficial for all stakeholders involved.

However, the production and communication of numbers not only can channel socio-political behaviour; it also has the predilection to "remake what it measures." (Espeland & Stevens, 2008, p. 431). There has been widespread criticism of the framing of the large-scale movements of nationals from Syria, Afghanistan and other countries into the EU in 2015 as a "crisis" (see for example, Raineri & Strazzari, 2021). Still, the EU's approach to migration policy still operates "in crisis mode" where sudden increases in irregular migration numbers bring about "a sense of loss of control" which, together with narratives suffused with images of terrorist threats and hostile invasion make for fear-based political instrumentalisation (Angenendt et al., 2023, p. 2).

Rodriguez Sanchez and Tjaden (2023) reviewed traditional and innovative methodological approaches to measure irregular migration stocks and flows as part of MIrreM's WP3. Despite gaps in available data, there are more efforts directed at measuring irregular migration stocks than flows. Still, as already underscored in the CLANDESTINO Project (2009a), the public perception of irregular migration is centred on flows, and it is dominated by one particular flow – the irregular inflow over land or sea borders. The number of irregular border crossings as reported by FRONTEX, the European Border and Coast Guard Agency, is widely used as an indication of irregular flows. Despite not being the sole instigator, FRONTEX has been instrumental in reinforcing narratives that frame migrants as potential risks, invoking correlations between migration and security challenges such as crime and terrorism (Perkowski et al., 2023). These perceptions have contributed to the securitisation of migration at the EU level, and the portrayal of migrants as inherently threatening subjects (Kalkman, 2021; Léonard, 2010). The "normalisation of crisis" is used to justify increased border controls (Wærp, 2024, para. 8) and it has provided fertile ground for FRONTEX to expand its powers and authority. This is also the case for the US, where border crossings particularly at the border with Mexico are routinely framed in public discourse not just as a "crisis" but also as an "invasion" (Hinck et al., 2023, p. 320).

Against this backdrop, this working paper addresses the **research question**: How are irregular migration flows measured and estimated in the EU and selected other countries (UK, US, Canada, and "transit countries"), and what can be said about the quality of the available indicators and estimates? Additionally, it explores the use of these measures in migration governance and policymaking. Underlying the overarching question are the following **sub-questions**:

- What data and methods are used to measure irregular migration flows in the EU and selected other countries?
- What were the main flow trends for EU Member States as a whole in the period 2008-2022[5] based on selected Eurostat flow indicators?
- What can be said about the quality of the available flow indicators and estimates in terms of validity, reliability, and other relevant quality criteria such as accessibility and documentation?
- What are the main gaps and challenges in obtaining high quality flow indicators and estimates in the EU in particular, and how can these challenges be addressed?

Our analysis makes a distinction between international and country level data. **International level data** refers to data with a broader geographic focus that spans multiple countries like data from Eurostat, while **country level data** pertains to data specific to a country which, in the present case, are the data collected by MIrreM's national rapporteurs. As the primary focus of this paper is EU level data, the level of analysis is thus, international, but following the research questions just outlined, it also covers country-level data.[6]

We also make a distinction between irregular flow estimates and indicators. **Estimates** refer to statistical calculations or approximations that quantify both observed and non-observed or unknown irregular migration flows. **Indicators** as used in this working paper, on the other hand, refer to metrics or variables that relate only to observed or known irregular migration flows. In other words, indicators of irregular migration flows show the number of actual observations or cases, such as border apprehensions, whereas estimates use indicators to come to conclusions about a broader trend, including non-observed components, such as the *total* number of adults, detected and undetected, who crossed into a country without the legal right to do so. Eurostat's compilation of statistics on asylum and the enforcement of migration legislation serves as an example of indicators covering irregular flows such as refusal of entry at the external borders, orders to leave, and returns.

Our methodological approach is twofold: a quantitative analysis of a select number of Eurostat indicators; and a qualitative analysis of interview data with EU policymakers, data from other international sources, and data on the country level. When analysing EU level data, we focus on Eurostat data as Eurostat is the primary repository of publicly accessible and well-documented migration statistics. Their comprehensive datasets offer ample opportunities for quantitative analysis and in-depth insights into migration trends within the EU. For a deeper understanding of migration dynamics at the EU level, we complement the

[5] At the time of writing, complete data for 2023 is not yet available.
[6] Local level data pertaining to data in lower administrative tiers such as the city or municipal level is discussed elsewhere in MIrreM i.e., in WP3 (D3.4), and WP4 (D4.2, and D4.3).

quantitative data with interview data with EU policymakers. Finally, we also look into the country level data collected by MIrreM's national rapporteurs.

Following the MIrreM conceptualisation of migrant irregularity (Kraler & Ahrens, 2023), flows can be differentiated into **geographic flows** (in- and out-movements across borders), **demographic flows** (births and deaths in irregularity), and **status-related flows** (falling into irregularity or acquisition of legal status). The measurement of different types of irregular migration flows provides a more dynamic and comprehensive understanding of migration patterns. We focus on geographic and status-related flows as there is hardly any data available, historically and presently, on demographic flows. Each type of flow can be further distinguished into **inflows** or **outflows**, which are the more detailed mechanisms or pathways into and out of irregularity.

The **scientific relevance** of this effort lies in its systematic approach to assessing the quality of available data related to irregular migration flows. Whereas the CLANDESTINO Project assessed and classified estimates of irregular migration stocks and flows into four categories (high, medium, low, and low with plausibility warning) based on the data, documentation, and estimation technique (Jandl et al., 2008), MIrreM takes a more disaggregated approach to allow for more nuance in quality assessment. We assess the quality of estimates across four areas: accessibility, documentation, reliability, and other methodological aspects such as internal and external validity. As there are hardly any estimates on irregular migration flows then and now, we collect and assess data on **indicators** of irregular migration (for example, border apprehensions, return decisions, etc.) and assess them in terms of accessibility, documentation, and validity and reliability.

We also analyse the **policy relevance** of data on irregular migration flows. Accurate and reliable data is essential for policymakers and governments to develop effective and evidence-based policies. This is particularly important in the present era of "datafication" whereby the automation of border procedures, utilisation of biometrics, and implementation of extensive statistical methodologies are becoming progressively integral to migration management (Bellanova & Glouftsios, 2022; Bigo, 2020; Cantens, 2018; Frowd, 2024; Perret & Aradau, 2023). It is also becoming increasingly common for statisticians to function as "back-office policymakers" as they find themselves obligated to make numerous methodological choices while translating abstract statistical principles and definitions into operational procedures, thereby actively shaping policies rather than simply executing them (Ustek-Spilda, 2020, p. 290). A thorough understanding of the quality of existing data can help policymakers make informed decisions regarding immigration laws, border control measures, and humanitarian interventions. Oftentimes, data is also used to justify and measure the effectiveness of policies related to irregular migration. This paper looks into the different facets of irregular migration data and discusses the intricate relationship between quantification and policymaking.

In terms of broader **public relevance**, this undertaking allows for a critical analysis of available irregular migration data and provides accessible and reliable information on irregular migration flows, particularly at a time when there is a lot of concern about irregular arrivals. Providing the public with reliable information fosters a more informed understanding of the complexities surrounding irregular migration, reducing the likelihood of misinformation and prejudice.

1.2 CONCEPTUAL FRAMEWORK AND QUALITY ASSESSMENT CRITERIA

There is not one clear and distinct definition of irregular migration. There are also different terminologies used for this phenomenon (e.g., undocumented, unauthorised, clandestine, *sans papier*, etc.). In MIrreM, **irregular migration** is operationally defined as a form of migration that is not "regular," "unlawful," or not according to the rules (Kraler & Ahrens, 2023). It covers both unlawful *entry* and *stay.* It builds on the CLANDESTINO definition, and in the EU context, irregular residents are third-country nationals (TCNs) that are either:

- Without any legal residence status in the country they are residing in; or
- Those although possessing an authorisation of some sort whose presence in the territory – if detected – may be subject to termination through an order to leave and/or an expulsion order because of their activities (e.g., visa-free citizens engaging in work, students working more than allowed, or persons with falsified documents).

The differentiation between stocks and flows is most important for measurement purposes (Jandl et al., 2008). Following this and the MIrreM operational definition of irregular migration, this paper looks into a particular form of irregularity, that is, flows into and out of unlawful stay. Again, **stocks** refer to the total population size in a given territory during a particular year, while **flows** are the events that increase (**inflows**) or decrease (**outflows**) migrant stock within a given territory during a certain period.

While there is abundant information on migrant stocks, data on migration flows is generally more limited. Conceptually, this is largely because flows are characterised by far greater volatility than stocks, and thus, much larger uncertainty of flow forecasts (Wiśniowski, 2021, p. 378). As Bijak et al. (2019, p. 40) point out, "there is no single forecasting approach that would be well suited for different flows." Practically, there are several reasons. Firstly, countries have been predominantly focusing on recording inflows, with limited attention to outflows. Moreover, most countries rely on administrative events related to immigration status as proxies for migration flows (i.e., issuance/withdrawal of residence permits). Developing countries particularly face challenges in collecting, administering, and reporting data on migration flows due to limited resources and infrastructure, since tracking flows demands significant resources and information technology systems. Physical geographies in many countries, especially those with archipelagic or isolated borders, pose additional challenges for data collection on migration flows, particularly concerning data on irregular migration flows (McAuliffe & Ruhs, 2017). As Jandl (2008, p. 19) pointed out in CLANDESTINO, all serious methods to estimate the *unknown* part of irregular migrant populations must be based on some form of "hard data" on *known and reported* cases (quotations and emphasis in the original). Many countries compile and report statistics on migrant stocks and flows but often "fail to cover irregular migrants" (Ahmad-Yar & Bircan, 2021, p. 18). There are also inherent issues associated with irregular migration data as the phenomenon of irregular migration itself often takes place outside the confines of formal

administrative processes. Furthermore, as migration dynamics are generally difficult to capture, stocks are relatively easier to measure than flows (Kraler & Reichel, 2022).

This said, the measurement of irregular migration flows is essential to a more dynamic and comprehensive understanding of migration patterns. Given very limited methods used to quantify irregular migration flows (Kraler & Reichel, 2011) and thus, hardly any estimates available, the MIrreM project takes a kind of "back-to-basics" approach by zooming in on indicators (Table 1) and assessing their quality based on several criteria (Table 2 and Table 3). Eurostat's compilation of statistics on asylum and the enforcement of migration legislation serves as an example of indicators covering irregular flows such as refusal of entry at the external borders, orders to leave, and returns. We thus assess the quality of the "hard data" and offer guidance in terms of their accessibility, documentation, and validity and reliability. We deem this an important first step given how data on irregular migration can become politicised and significantly influence both policy decisions and public opinion in various ways. Indeed, "aggregate indicators are never sufficient and meaningful in themselves" (Kraler & Reichel, 2022, p. 458). Thus, this effort to assess the quality of available flow indicators and to identify how they are used in policymaking is aimed ultimately at providing guidance on how to avoid the misuse of migration data.

Table 1. Irregular migration flows and indicators

Type of flow	Indicators	
	Inflow	**Outflow**
Demographic flows	• Births in irregularity (parents without status) • Failure to obtain a status for the child	• Deaths in irregularity (may or may not be related to migration)
Geographic flows	• **Refusals of entry at the external border** • **Border apprehensions either at external border areas or inland within the national territories**	• **Return decisions (otherwise known as orders to leave or expulsion orders)** • **Returns after an order to leave (voluntary or forced)** • Onward migration
Status-related flows	• **Negative asylum decisions** • **Withdrawal of status** • Visa overstaying	• Regularisation • Change in personal circumstances entitling to stay • Acquisition of provisional status

Notes: Adapted from the MIrreM taxonomy of migrants with a precarious status (Kraler & Ahrens, 2023). Items in **bold** are indicators covered in Eurostat.

Table 2. MIrreM criteria for quality evaluation of indicators

Criteria	High (3 points)	Medium (2 points)	Low (1 point)
Accessibility	Data is publicly available and electronically accessible with no permissions required	Data is available on request from relevant authorities	Data is available, but access and use are exclusive to authorities
Documentation	Sufficient and transparent information on data and methods are available and	Limited information on data, methods, and quality are	Information on data, methods, and quality are neither

	accessible; a comprehensive quality report is also available	available and accessible	available nor accessible
Validity and reliability	Data is representative of the phenomenon it is supposed to measure and adequately reflects the type of irregular migration being measured; data is relatively complete (not highly selective) and does not indicate internal contradictions	Data is selective and points to some internal contradictions	Data is neither valid nor reliable

The first two criteria, accessibility and documentation, are based on the FAIR Data Principles (Findability, Accessibility, Interoperability, and Reusability) which provide guidelines for making data and related resources more accessible and reusable (Wilkinson et al., 2016). Here is a brief explanation of each principle:

1) Findability: Data and resources should be easy to find for both humans and machines through clear metadata and identifiers.
2) Accessibility: Data and resources should be readily accessible, preferably with open access, and permissions should be clearly stated.
3) Interoperability: Data and resources should be structured in a way that allows for easy integration with other datasets and tools.
4) Reusability: Data and resources should be well-described and properly formatted to facilitate reuse, ensuring that they can be used for multiple purposes by different individuals or groups.

To simplify, we have combined findability and accessibility under the rubric of **accessibility**, and interoperability and reusability under **documentation**. When establishing the quality of quantitative data, validity and reliability are two essential considerations. **Validity** is the extent to which the data accurately reflects the type of irregular migration flow it is supposed to measure. In statistics, a distinction is made between internal validity (how well a measure reflects the phenomenon under study) and external validity (the extent to which the measure is representative of the universe and applies to other situations and settings). Meanwhile, **reliability** is the consistency and stability of the data collected and reported over time. For the quality assessment of estimates, there is another criterion, **methodology**, which looks into the appropriateness and adequacy of the method used in the estimation and the extent to which it can be replicated.

Table 3. MIrreM criteria for the quality evaluation of estimates

Criteria	High (3 points)	Medium (2 points)	Low (1 point)
Accessibility	All raw data used to construct the estimate is publicly available and electronically accessible with no permissions required.	At least some of the raw data used to construct the estimate is only available on request from relevant authorities. If some of the data is not available at all, then give 1 point.	At least some of the raw data used to construct the estimate is not available for most potential users.
Documentation	Full documentation about data and methods are available and accessible. The level of	Limited information on data, estimation methods, and quality are available and	Information on data and estimation methods is neither

	information allows for replication of the estimates.	accessible. Insufficient details to replicate the estimates.	available nor accessible.
Reliability	Analysis includes demonstrated reliability indicators, with limitations clearly specified (e.g. ranges, alternative calculations, characterisation as minimum or maximum estimate).	Some discussion of reliability, but no indicators in quantitative terms.	Missing a discussion of reliability.
Methodology	Methodology is adequate and comprehensive including, but not limited to, rigorously implemented multiplier or residual studies.	Methodology is adequate, even if not comprehensive, including but not limited to: (1) Simple multiplier calculations; (2) Simple residual estimates; (3) Adjustment of older estimates with partly insufficient data; (4) Aggregate estimates for different groups, partly relying on plausibility calculations.	Inadequate method and application of the method; resulting estimate lacks foundation

In the following sections, we discuss the methodology in more detail, followed by a presentation of our findings on indicators and estimates across different levels of analysis. We conclude with a synthesis of the findings addressing the research questions and some recommendations on how to address the identified gaps in data on irregular migration flows.

2. METHODS

2.1 BACKGROUND AND SCOPE

The MIrreM project is a follow-up to CLANDESTINO, which covered the period 2000-2007. MIrreM extends this to the subsequent period 2008-2023, although some of the statistics covered in the present report are only until 2022 pending complete reports for 2023. The goal of this exercise is to produce an inventory and provide a critical appraisal of estimates and indicators related to irregular migration flows. This is done in conjunction with WP4, which is doing the same for irregular migration stocks. Together, the resulting databases will contain estimates of the size and characteristics of the irregular migrant populations in a given country (stocks) and the changes in that population (flows).

As shown in the following Table, data from 12 European countries were collected in CLANDESTINO. MIrreM expands the scope and covers 13 European countries (12 EU Member States and the UK), with the addition of Canada and the US, as well as Bosnia and Herzegovina, Morocco, Serbia, Tunisia, and Türkiye, making for **20 countries** in total. The inclusion of Canada and the US not only allows comparisons beyond the European context but also makes MIrreM results generally more relevant for other international cooperation frameworks, such as the OECD or in the context of the Global Compact on Migration. Originally, the intention was to collect and analyse data from five "transit countries"– Bosnia and Herzegovina, Morocco, Serbia, Tunisia, and Türkiye. However, data from Serbia is not presently included as it had not been made available in time for the preparation of this working paper. The categorisation of certain states as "transit countries" has been critiqued for its politicised nature and oversimplification of migration dynamics (İçduygu & Yükseker, 2012). Critics also argue that such labels underestimate the nuanced reality of migration, which often involves both temporary and long-term settlement in transit countries (Collyer & De Haas, 2012). Despite these criticisms, the concept of *transit* remains valuable in understanding irregular migration flows. It enables scholars and policymakers to move beyond the traditional origin-destination framework and acknowledge the fluidity of migration aspirations and decisions over time and space (Cummings et al., 2015).

Table 4. Countries covered in CLANDESTINO and in MIrreM

	CLANDESTINO	MIrreM
EU countries	Austria	Austria
	Czech Republic	Belgium
	Germany	France
	Greece	Finland
	France	Germany
	Hungary	Greece

	Italy Netherlands Poland Slovakia Spain UK	Ireland Italy Netherlands Poland Portugal Spain
Other countries	N/A	Bosnia and Herzegovina Canada Morocco Serbia Tunisia Türkiye UK USA

2.2 DATA COLLECTION AND ANALYSIS

Eurostat data

On the EU level, different events have generated a lot of interest in predicting irregular migration (Angenendt et al., 2023). A detailed overview of the EU irregular migration data landscape is provided in Annex 1. Our primary data source is Eurostat which provides publicly accessible and well-documented migration statistics. In the absence of irregular flow estimates to assess, we look instead into a select number of these flow indicators as presented in the following Table (definitions are provided in Annex 4). As Kraler and Reichel (2011) have underscored previously, despite limitations with the available data, they can still be used as indicators of certain trends.

Table 5. Selected irregular flow indicators for EU-level analysis

Type of flow	Indicator – inflow	Indicator – outflow
Demographic		
Geographic	(1) TCNs refused entry at the external borders	(3) TCNs ordered to leave
	(2) TCNs found to be illegally present inland because of illegal entry	(4) TCNs returned to a third country following an order to leave
Status-related	(5) Negative first instance asylum decisions	
	(6) Decisions withdrawing status granted at first instance decision	

As there are no systematic and reliable data available for births and deaths and irregularity, demographic flow indicators are not included in the present analysis, and as mentioned in the Introduction previously, we focus on geographic and status-related flows.

The selection of the six indicators is based primarily on their frequency of use in discussions related to irregular migration flows. When reporting enforcement of migration legislation

statistics for example, Eurostat focuses on four indicators: refused entry in the EU, found to be illegally present in the EU, ordered to leave an EU Member State, and returned following an order to leave (Eurostat, n.d.-e). Then and now, geographic flows are the most visible flows as these involve the physical movement of individuals across borders (CLANDESTINO Project, 2009a). FRONTEX, for example, has been collecting and reporting data on irregular migration flows into the EU since its establishment in 2004. Its *Migratory Map* and *Annual Risk Analysis* report data on geographic flows including the following indicators – illegal border crossings, refusals of entry, return decisions, and so on. Many destination countries have invested heavily in securing their borders and detecting and monitoring irregular migration flows (McAuliffe & Sawyer, 2021). For its part, the EU has allowed for the significant expansion of FRONTEX's mandate in the management of external borders (Kalkman, 2021). To better manage migration, increased capabilities to secure the borders and prevent illegal entry are complemented by similar stepped-up efforts to return as many irregular migrants as possible, whether voluntary or involuntary, to countries of origin or countries outside the EU (DeBono, 2016).

For this report, we thus zoom in on the number of TCNs refused entry at the external borders (**Indicator 1**). While Eurostat has only started collecting and reporting statistics on TCNs found to be illegally present by "place of apprehension" (external border area or inland area) and by "grounds of apprehension" (illegal entry, overstay, and other reasons including unknown reason) in 2021, and the dataset as a whole is ordinarily a stock rather than a flow indicator, we are including TCNs found to be illegally present inland because of illegal entry (**Indicator 2**) in the present analysis as it indicates secondary irregular movements within the EU.[7] Illegal entry refers to illegal border crossing either by avoiding border control or by successfully using a fraudulent document to cross the border. Apart from orders to leave or return decisions issued by EU Member States (**Indicator 3**), we also look at actual returns to third countries, either to the person's country of origin or another third country (**Indicator 4**). This indicator is particularly interesting as it points to actual departures or outflows from EU territory.

In addition to the four geographic flow indicators, we are also looking into two status-related flow indicators – negative first instance asylum decisions (**Indicator 5**) and decisions withdrawing status granted at first instance decision (**Indicator 6**) – which are collected by Eurostat under the rubric of asylum statistics. Negative asylum decisions are important indicators of irregular migration because a negative decision signifies that the authorities do not recognise the individual's claim for protection under asylum law, which can imply that they are in the country irregularly. In other words, negative asylum decisions often produce an inflow of irregular migration. This is also the case for decisions withdrawing status granted at the first instance. A person's temporary protected status can be revoked or terminated due to refusal to renew, and the request to extend the period of validity can be denied when the legal ground for granting protection status has ceased to exist. The withdrawal of status marks a transition from regularity to irregularity and is thus, also an inflow indicator. For both indicators related to asylum, we are looking only at **first instance decisions** which refer to decisions granted by the respective authority acting as a first instance of the

[7] We recognise the ethical issue pertaining to the use of the word "illegal" in the context of migration but here, we use the same terminology employed by Eurostat.

administrative/judicial asylum procedure in the receiving country. The decision concerns refugee status and subsidiary protection, as well as authorisation to stay for humanitarian reasons. This decision is subject to an appeal or review, resulting in a final decision. The asylum procedures and the numbers/levels of decision-making bodies, however, differ between Member States and there are significant time variations when reporting the true "final instance" so for this paper, we looked at first instance decisions.[8]

Quantitative analysis of selected irregular migration flow indicators

By focusing on the six flow indicators explained above, we aim to provide a picture of the data associated with the evolving patterns and dynamics of irregular migration in the EU. We start with a presentation of the general trends between 2008 and 2022. We then discuss the quality of the quantitative data using the assessment criteria discussed in Section 1.2 where special attention is given to the third criterion – validity and reliability. Using Eurostat's bulk download facility, the data were obtained and subsequently uploaded to SPSS for analysis. When establishing the quality of quantitative data, validity and reliability are two essential considerations. To reiterate, **validity** is the extent to which the data accurately reflects the type of irregular migration flow it is supposed to measure, while **reliability** is the consistency and stability of the data collected and reported over time. To establish external validity (the extent to which the measure is representative of the universe and is applicable to other situations and settings), we looked at the **percentage of missing values** for each indicator. More concretely, we looked into the different levels of disaggregation available for each indicator and determined the percentage of missing values. In Eurostat, missing values are denoted by the special value of ":" which means "not available;" a value of "0" (zero) is thus, not a missing value. For the first indicator, TCNs refused entry at the external borders for example, we analysed the percentage of missing values in terms of country of citizenship, the reason for the refusal of entry, and the type of border (land, sea, or air). Calculating the percentage of missing values provides valuable information about the completeness and quality of the dataset which are relevant to assessing the external validity of the data. While we cannot establish internal validity (how well a measure reflects the phenomenon under study) with this test, analysing the percentage of missing values ensures transparency and provides guidance on how well one can use the datasets when making claims or imputations about irregular migration flows. A higher percentage may indicate data quality issues or limitations in data collection processes.

To measure reliability or internal consistency, we look at **inter-item correlation** and the degree to which observations are relatively stable over time, providing insights into the consistency or coherence of the flow indicator. Higher inter-item correlations (i.e., strong positive correlations) suggest that the items are measuring the same underlying construct consistently, indicating high internal consistency or reliability of the indicator. Conversely, low inter-item correlations may indicate inconsistency or heterogeneity among the items,

[8] It should be noted however, that negative first instance decisions have limitations as status-related inflow indicators precisely because a significant number of negative decisions are successfully challenged on appeal. ECRE reports in 2022 that over the last five years, an average of 33% of first instance negative decisions across EU countries and Norway and Switzerland are overturned on appeal. For details, please see: ECRE. (2022). *Asylum statistics and the need for protection in Europe: Updated Factsheet*. https://ecre.org/wp-content/uploads/2022/12/Asylum-statistics-and-the-need-for-protection-in-Europe-final.pdf.

raising concerns about the reliability and, by implication, the validity of the indicator. Calculating the inter-item correlation was done as follows:

In the first step, we calculated the inter-inter correlation for each EU country when the number of observations on a flow indicator for each EU country was specified by third country and year for the period 2008 (or the first year for which Eurostat data is available for the flow indicator) up to and including 2022. For example, observations for calculating the inter-item correlation on return flows for Germany would be (1) the total number of Afghans who returned to a third country from Germany in 2008, (2) the total number of Armenians who returned to a third country from Germany in 2008, (3) the total number of Afghans returned to a third country from Germany in 2009, and so forth. In other words, the inter-item correlation for an EU country is the overall correlation between the year variables (i.e., 2008 up to and including 2022) where cases are observations by a third country on the flow indicator for these years. In subsequent steps, we then specified the data according to additional dimensions, if available, such as gender or age category. For example, we then separately calculated the total number of male and female Afghans who returned to a third country from Germany in year x, and so forth. The inter-item correlation for the EU country then still is the correlation between the different year variables, but the number of observations then increases with each dimension added (e.g., if we specify gender in addition to country of origin, the number of observations doubles, as gender "unknown" was coded as missing). The inter-item correlation for the EU "as a whole" on a given flow indicator is the unweighted average of the inter-item correlation for all countries that were EU countries in the period of observation, or part of that period, including the UK and Croatia. The SPSS syntax used for the analyses is available on request.

Interview data

Between September 2023 and February 2024, the teams at the University of Maastricht and ICMPD also conducted 10 key informant interviews (online and in person) with migration data experts and policymakers. The semi-structured interviews focused on data needs and usage, particularly in policymaking. The participants were three women and seven men from the following organisations (listed in alphabetical order):

- European Border and Coast Guard Agency (FRONTEX)
- European Commission, Department for Migration and Home Affairs (DG HOME)
- European Commission, Knowledge Centre on Migration and Demography (KCMD)
- European Union Agency for Asylum (EUAA)
- European Union Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice (eu-LISA)
- IOM Global Migration Data Analysis Centre (GMDAC)
- IT and Development Centre at the Estonian Ministry of the Interior (SMIT)

With consent from the participants, the interviews were recorded and then transcribed and coded using ATLAS.ti. Given the sensitivity of the topic of irregular migration, anonymity is maintained, and direct quotations are not used in this paper. Findings are reported instead in thematic form (Section 3.1).

Other data sources

To supplement Eurostat data, we looked into other producers of migration statistics. and consulted the following data sources:[9]

- FRONTEX's Migratory Map which presents the current migratory situation in Europe, and Annual Risk Analyses which identifies the most important thematic risks for European integrated border management and the most affected regions at the EU external border
- IOM's Missing Migrants Project (MMP) which, since 2014, records people who die in the process of migration towards an international destination

Both data sources provide data on geographic flows, and IOM additionally, collects statistics on migrant deaths and disappearances. While this is not exactly a demographic outflow indicator, they are relevant to the discussion as many migrant fatalities have been reported in the context of attempted, presumably unauthorised entries into the EU via the Mediterranean.

Country-level data

With the same national rapporteurs working on both stocks and flows, data collection for WP5 (flows) was administratively subsumed under WP4 (stocks). Together with the team at the University of Oxford leading WP4, the team at the University of Maastricht leading WP5 developed the *MIrreM guidelines for data collection of estimates of stocks and flows of irregular migrants and irregular migration indicators*. This comprehensive document reiterates the purpose of the exercise as well as operational definitions, and illustrates, with concrete examples, the reporting matrices for the collection and assessment of stock and flow estimates and flow indicators. Apart from the written guidelines, reporting templates (Excel sheets for the estimates and indicators, and a Word document with the country context questionnaire) were developed and shared with all the rapporteurs. In brief, national rapporteurs were instructed to:

- Compile relevant irregular migration data and record them using the templates provided;
- Assess the quality of the data based on the developed criteria;
- Provide background information, and identify how the data is used in policymaking (either through desk research or fieldwork[10]); and
- Finally, complete the country context questionnaire and reflect on the overall picture of irregular migration data within the specific country assignment.

The guidelines build on those developed in CLANDESTINO to maintain some consistency across the two projects, but the MIrreM version was modified to account for changes across

[9] We also initially considered analysing data from UNODC's Observatory on Smuggling of Migrants (number of people who arrived irregularly to Europe by the sea) and KCMD's EU Return Dashboard (nationalities that receive an order to leave) but excluded them in this paper as UNODC uses FRONTEX data, and KCMD uses Eurostat data.

[10] The national rapporteurs working on WP4 and WP5 also worked on WP3 (politics) and prepared country migration profiles and briefs which involved desk research and fieldwork (i.e., interviews and workshops with key stakeholders). As WP3 also covered the use of data in policy development, national rapporteurs were requested to also report information pertinent to WP4 and WP5 when available.

the different periods and overall purposes of the project. In addition, the approach to assessing the quality of estimates and indicators was refined, notably by explicitly distinguishing between statistical indicators, on the one hand, and estimates, on the other; developing different quality assessment criteria (discussed in Section 1.2 of this paper); and collecting information on the use of these data in policymaking. Beyond the immediate purpose of guiding data collection and analysis within the scope of MIrreM, these tools may also be useful for other researchers working on comparable topics characterised by a lack of robust research-driven data, hard-to-reach target groups, and limited and imperfect administrative data.

The WP4 and WP5 leaders jointly organised an information workshop on 15 May 2023 and invited all national rapporteurs to go over the guidelines and assess them for clarity and usability. The WP leaders prepared concrete examples from the UK and the Netherlands to demonstrate the intended use of the data collection instruments. Another meeting was held on 21 June 2023 as a check-in moment, and rapporteurs had the opportunity to raise questions and discuss issues with the data collection exercise.

We recognise that there is some element of subjectivity involved when doing the quality assessment of the collected data. Thus, beyond the numerical scores, rapporteurs were requested to provide sufficient explanation for the score given and as much relevant contextual information as possible. As noted previously, a quantitative overview of the available data on irregular migration flows together with the detailed national datasets, quality assessments, and corresponding country context documents will be made available in the MIrreM public dabatase. For this, please see:

> Siruno, L., Leerkes, A., Badre, A., Bircan, T., Brunovská, E., Cacciapaglia, M., Carvalho, J., Cassain, L., Cyrus, N., Desmond, A., Fihel, A., Finotelli, C., Ghio, D., Hendow, M., Heylin, R., Jauhiainen, J.S., Jovanovic, K., Kierans, D., Mohan, S.S., Nikolova, M., Oruc, N., Ramos, M.P.G., Rössl, L., Sağiroğlu, A.Z., Santos, S., Schütze, T., & Sohst, R.R. (2024). MIrreM Public Database on Irregular Migration Flow Estimates and Indicators. Krems: University for Continuing Education Krems (Danube University Krems). https://doi.org/10.5281/zenodo.10813413.

Contributions from the national rapporteurs were received between July 2023 and January 2024. Some of the rapporteurs participated in one-on-one feedback sessions with the main authors of this paper, which provided an opportunity to work on revisions and to collect additional information. There are significant variations in the volume of data received from the national rapporteurs as well as in the overall quality of each contribution. While this working paper looks primarily into EU-level data, it also sets the stage and provides contextual information for the upcoming MIrreM public database where country level data from the countries covered in MIrreM are presented in full.

3. RESULTS

3.1 EUROSTAT DATA

Irregular flow trends in the EU

As illustrated in Figure 1[11], we generally see fluctuations in the levels over the years for all selected indicators except for an increasing number of TCNs found to be illegally present inland because of illegal entry[12] (Indicator 2) of which data is only available for 2021 and 2022. Geographic inflows as indicated by the number of TCNs refused entry at the external borders (Indicator 1) show an almost steady decline between 2008 and 2015, only to increase again beginning in 2016 and reaching a peak higher than 2008 figures in 2019. This year, Spain reported more than double refusals of entry compared to the previous year. As explained in Eurostat (2020):

> A large majority of the refusals of entry were recorded in Spain (493 500), followed by Poland (65 400) and France (56 600): together these three EU Member States accounted for 86% of all non-EU citizens refused entry into the EU in 2019.
>
> The highest number of non-EU citizens refused entry into the EU in 2019 were Moroccans (489 900), followed by citizens of Ukraine (66 500) and Albania (22 800).

The following year 2020 saw a rapid decline in refusals of entry at the external borders. Downturns are also observed for orders to leave (Indicator 3) and returns to a third country following an order to leave (Indicator 4) which reflect the impacts of the COVID-19 pandemic on migration and mobility across the globe.

It is interesting to note that in 2022, the number of TCNs found to be illegally present inland because of illegal entry (Indicator 2) is higher than those refused entry at the external borders (Indicator 1), not least because Indicator 2 is based on statistics on apprehensions and only people who are apprehended or otherwise come to the attention of national immigration authorities are recorded in these statistics (Eurostat, n.d.-e). While not indicative of a trend given the very limited time coverage, Indicator 2 points to a comparable, if not significant secondary irregular migration flows within the EU relative to those from outside the territory as captured by Indicator 1. It suggests more "illegal" than "refused" entries, or in other words, more people able to avoid border control or are successful in using a fraudulent document to cross the border compared to those detected and refused entry at the external borders.

[11] The datasets used are available here: https://ec.europa.eu/eurostat/web/migration-asylum/managed-migration/database, and here: https://ec.europa.eu/eurostat/web/migration-asylum/asylum/database.
[12] As explained in Footnote 6, please note that we use the same terminology employed by Eurostat.

In terms of pivotal events, 2015 marked the start of large-scale movements into the EU of nationals from Syria, Afghanistan and other countries. As shown in Figure 2, this year saw record numbers of asylum applications – over a million, which is more than twice the level recorded in the previous year and the highest recorded since 2008 (EASO, 2015). Though in absolute terms, there was an increase in negative asylum decisions (Indicator 2), the figures were lower relative to positive decisions.[13] The number of negative decisions peaked in 2017 and since then, it has reverted to pre-2015 levels where negative decisions are higher than positive ones. Generally, status-related inflows as indicated by negative first instance asylum decisions have increased since 2008, with significant peaks in 2016 and 2017.

There are also fluctuations with regard to decisions withdrawing status granted at first instance decision (Indicator 6), with noticeable declines between 2011 and 2018, and surges between 2019 and 2022. The numbers are illustrated in Figure 3, disaggregated by the type of protection granted (and subsequently withdrawn). The majority of withdrawals pertain to status granted under the Geneva Convention, which is also the type of protection most commonly granted. Data on the reason for the decision withdrawing status – revocation, ending, or refusal to renew – is collected only since 2021. As shown in Figure 4, most withdrawal decisions for both 2021 and 2022 are due to revocation. This happens when the person should have been excluded from international protection earlier, when the status was obtained through misrepresenting or omitting facts, or when the beneficiary is considered to be a danger to the security of the state or the community (EUAA, n.d.). Regardless of the reason, withdrawal of status is a status-related inflow indicator, where people formerly with a regular migration status transition to an irregular situation following the withdrawal decision.

[13] It should also be noted here that not all applications lodged are processed and receive a decision in the same year and as such, the number of applications is not necessarily the sum of positive and negative decisions.

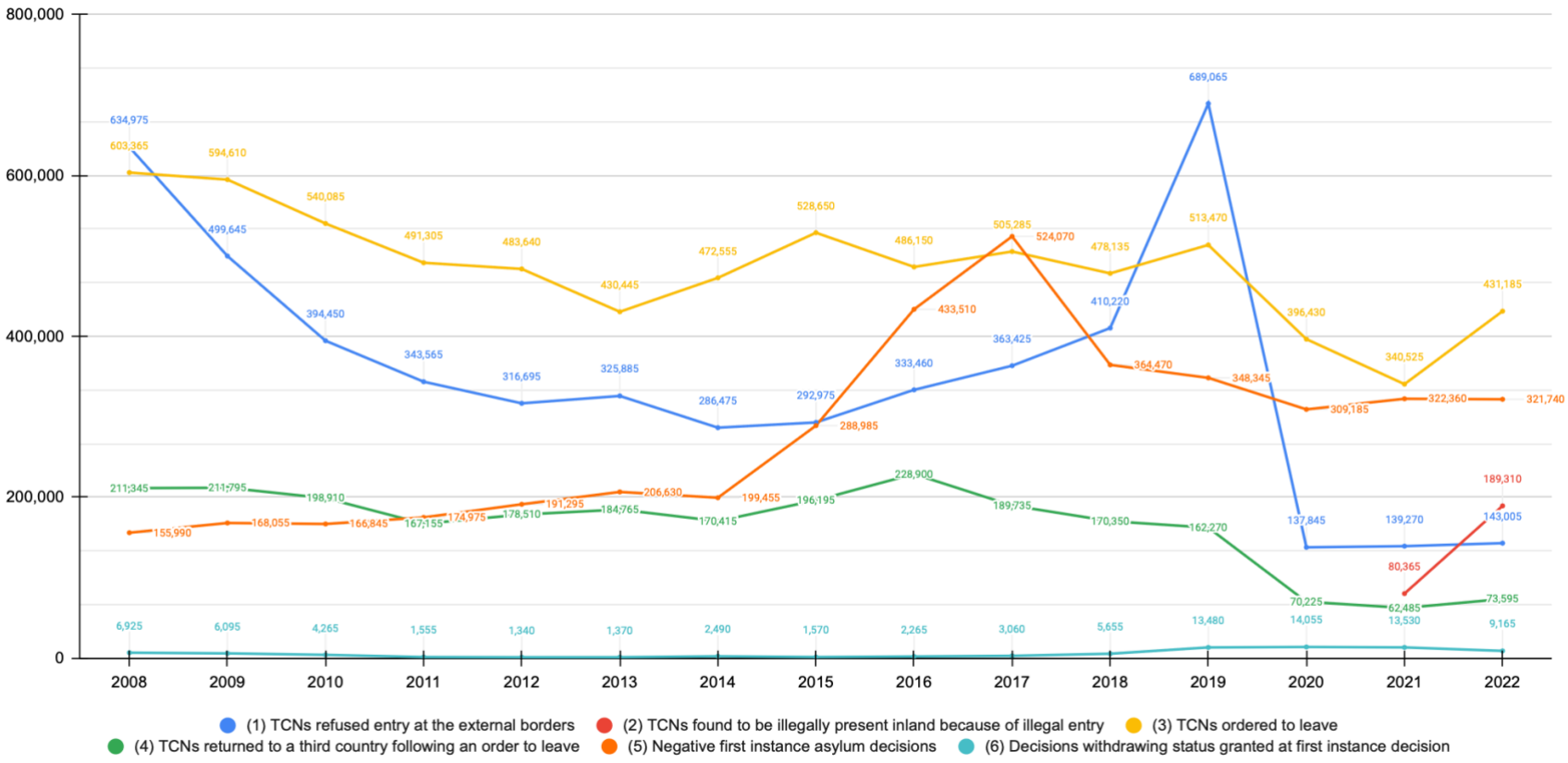


Figure 1. Trends of selected irregular flow indicators, total annual aggregated data from Eurostat

Notes: The numbers presented are totals of available data for EU27 (2008-2022, except for Croatia with data starting from 2013) and the UK (2008-2019). They are not necessarily comparable over time for all countries covered. Please refer to the analysis of missing values in the following section. For comparison purposes, a figure excluding data from the UK is presented in Annex 5. It shows the expected reduction in absolute numbers for all indicators, but trends overall appear to be consistent with the above.

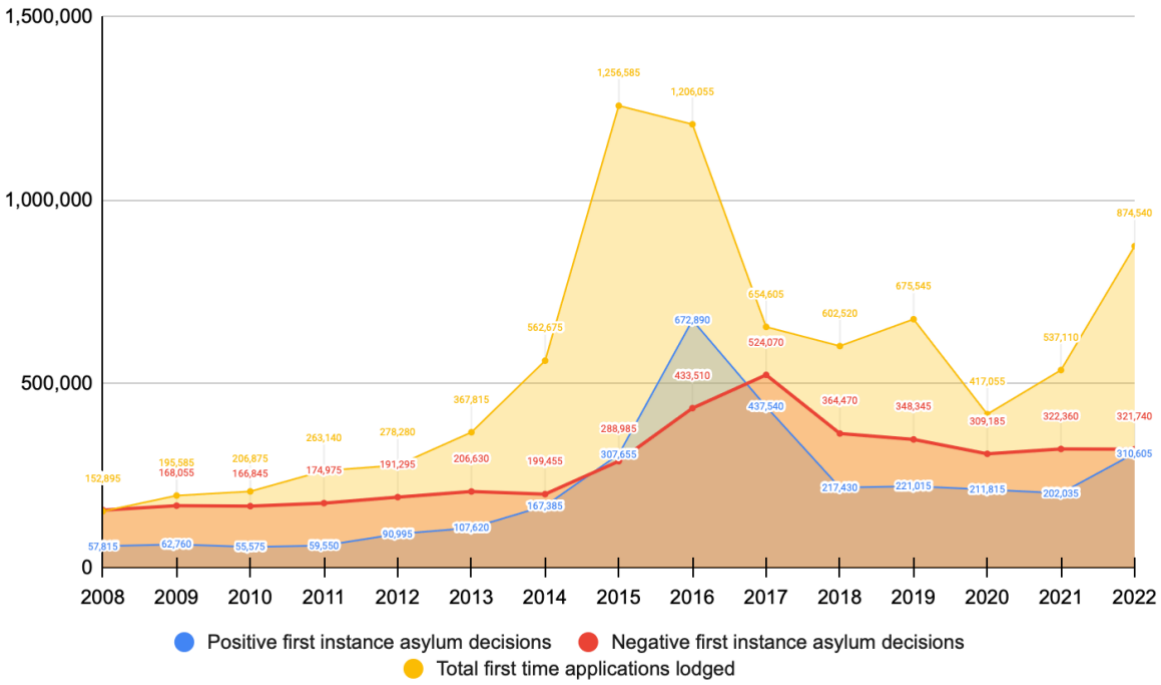


Figure 2. First time applications lodged, positive and negative asylum decisions (Indicator 5)

Note: Total annual aggregated data from Eurostat for EU27 (2008-2022) and the UK (2008-2019)

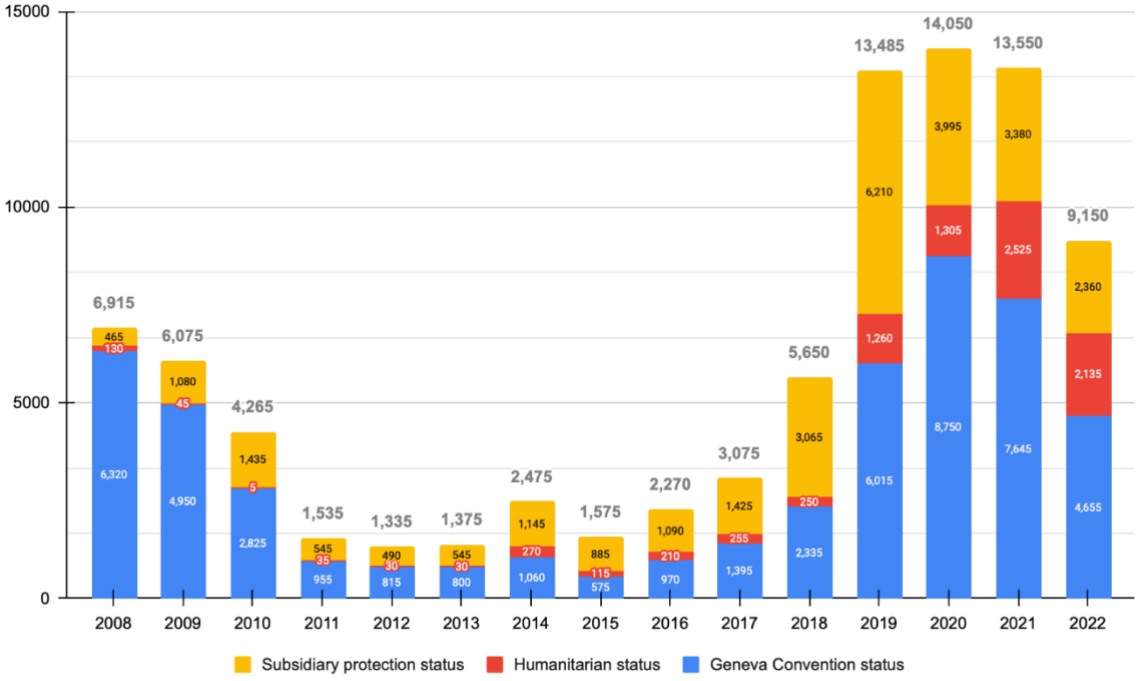


Figure 3. Decisions withdrawing status granted at first instance decision (Indicator 6) by type of protection

Note: Total annual aggregated data from Eurostat for EU27 (2008-2022) and the UK (2008-2019)

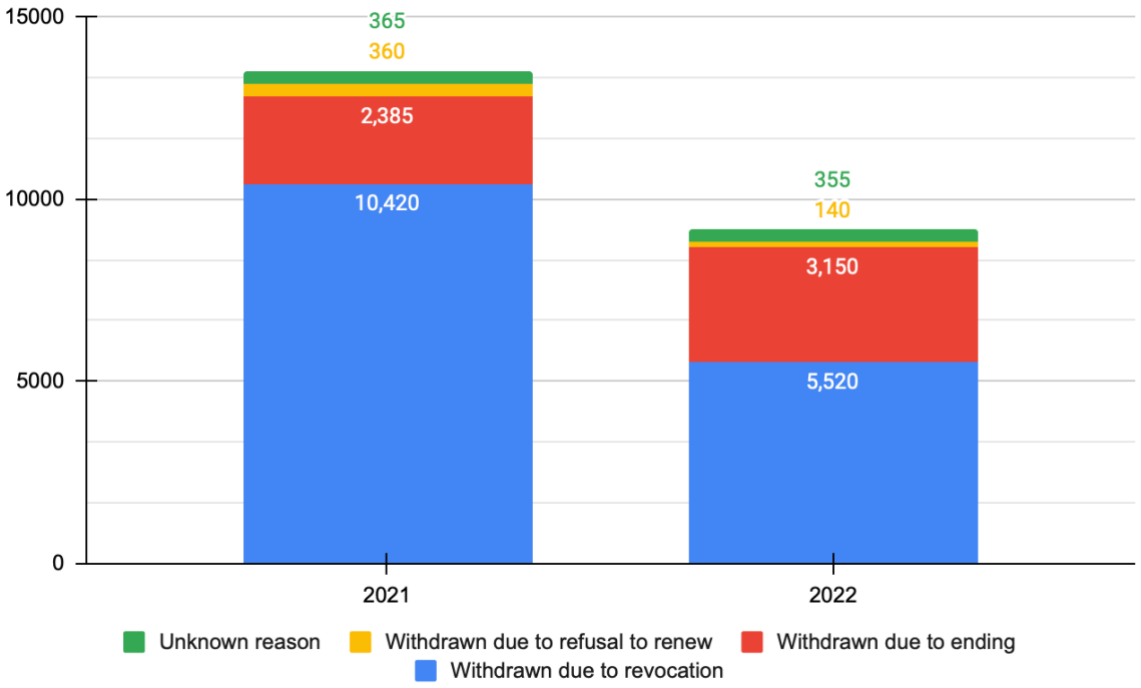


Figure 4. Reasons for decisions withdrawing status granted at first instance decision

Note: Total annual aggregated data from Eurostat for EU27 (2021-2022)

Quantitative analysis of selected irregular migration flow indicators

The results of the percentage of missing values and inter-item correlation analyses are presented in the following Tables.

Table 6. Percentage of missing values for selected irregular flow indicators

INDICATORS		Eurostat code	Mean
1	**TCNs refused entry at the external borders**	migr_eirfs	**2.1%**
Breakdown	country of origin		2.9%
	reason for refusal		2.9%
	type of border		8.7%
2	**TCNs found to be illegally present inland because of illegal entry**	migr_eipre	**3.6%**
Breakdown	country of origin		4.1%
	sex		36.1%
	age		46.7%
3	**TCNs ordered to leave**	migr_eiord	**4.8%**
Breakdown	country of origin		6.8%
	sex		82.1%

	age		85.4%
4	**Third country nationals returned to a third country**	migr_eirtn	**4.5%**
Breakdown	country of origin		6.6%
	sex		82.4%
	age		89.0%
5	**Negative first instance asylum decisions**	migr_asydcfsta	**2.1%**
Breakdown	country of origin		2.2%
	sex		34.8%
	age		44.1%
6	**First instance withdrawal of status**	migr_asywitfsta	**5.7%**
Breakdown	country of origin		6.2%
	decision to grant status		6.2%
	reason for withdrawal		90.0%

Notes: Data used for the analysis are from Eurostat, covering aggregated annual data from EU27 countries (2008-2022 except for Croatia, 2013-2022) and the UK (2008-2019). To prevent inflation of results, countries with no external land or sea borders and hence, no data to report, have been excluded in the analysis of Indicator 1. Items marked 'unknown' in Eurostat for example, gender and age, have also been recoded. For Indicator 2, countries with external land borders were excluded because their inland apprehensions were contaminated with external flows. For Indicator 6, the 'reason for withdrawal' has been recorded only since 2021, thus the very high missing value of 90%.

The analysis points to a relatively complete dataset for all six indicators with average missing values ranging from 2.1% (Indicator 1) to 5.7% (Indicator 6) and thus, a generally good external validity. There are, however, noticeable increases in missing values when the indicators are disaggregated by sex and age.[14] While there is no established "cut-off" when it comes to missing data, it can have important consequences for statistical analysis (Dong & Peng, 2013). Caution should be exercised when using the same dataset to generate estimates, particularly for certain variables like sex and age. Indicators 3 (orders to leave) and 4 (returns to a third country) stand out in this regard, with missing values higher than 80%. When reporting geographic outflows using these two indicators, care is required when making assumptions or conclusions about the sex and age of irregular migrants involved.

In terms of reliability, a more varied picture emerges following the inter-item correlation test. As shown in Table 7, while the average inter-item correlation for each indicator is 0.70 and above suggesting a relatively high internal consistency of the data, there is a significant variation per country.

Table 7. Inter-item correlation of selected irregular migration flow indicators

INDICATORS		**Lowest value**	**Highest value**	**Mean**	**sd**
1	TCNs refused entry at the external borders	0.04 (DK)	1.00 (EL*)	0.70	0.22
2	TCNs found to be illegally present inland because of illegal entry	0.14 (CZ)	0.97 (SI)	0.70	0.24
3	TCNs ordered to leave	0.36 (MT)	0.97 (PT)	0.74	0.15

[14] The country of origin is recorded more frequently, which may point to the prominence of an "ethnic lens" approach also to the collection of migration statistics. However practically, this serves as essential information for authorities, particularly in view of returns.

4	Third country nationals returned to a third country	0.27 (NL)	0.93 (EL)	0.70	0.17
5	Negative first instance asylum decisions	0.86 (LT)	0.99 (FR)	0.94	0.03
6	First instance withdrawal of status	0.58 (LU)	0.98 (PL)	0.86	0.12

Notes: Data used for the analysis are from Eurostat, covering available aggregated annual data from EU27 countries (2008-2022 except for Croatia, 2013-2022) and the UK (2008-2019). The "lowest value" column records the inter-item correlation value for the country with the lowest resulting value (the minimum in the range of values), and the "highest value" column is for the country with the highest (the maximum in the range). The "mean" column is the average of all values based on the aggregated data for all countries and years covered in the analysis. *EL pertains to Greece (Elláda).

This is particularly significant for Indicators 1 and 2 with high standard deviations, and to some extent, also for Indicators 3 and 4. On the other hand, indicators 5 and 6 have low standard deviations. What this means is that with regard to the first four indicators, there is a greater degree of fluctuation or dispersion within the dataset. Meanwhile, the minimal variability observed in the final two indicators suggests a higher level of consistency or uniformity. While all six indicators may reliably be used to describe observed irregular migration flows, caution is again advised when using them for estimating flows.

Quality assessment of EU-level flow indicators

All of the above considered, and following the MIrreM criteria for the quality assessment of irregular flow indicators, we arrive at the following conclusion:

Table 8. Summary of quality assessment of selected EU-level indicators

Criteria	Quality	Explanation
Accessibility	High	All data is publicly available and electronically accessible from Eurostat with no permissions required. There is general alignment with the FAIR Data Principles.
Documentation	High	Sufficient and transparent information on data and methods are available and accessible; a comprehensive quality report is also available for almost all the countries covered (except for Ireland).
Validity and reliability	Medium	While the data is generally representative of the specific irregular migration flows being measured, there is some variation in terms of external validity and reliability following the percentage of missing values and inter-item correlation tests.

Among the six indicators analysed, negative first instance asylum decisions (Indicator 5) is arguably the most valid and reliable[15], followed by TCNs ordered to leave (Indicator 3), TCNs returned to a third country following an order to leave (Indicator 4), TCNs refused entry at the external borders (Indicator), Decisions withdrawing status granted at first instance decision (Indicator 6), and TCNs found to be illegally present inland because of illegal entry (Indicator 2) respectively. Data on asylum-related status flows has been collected by Eurostat since 1985 on a voluntary basis, and they are generally reliable as a flow indicator

[15] Taking note of course that there are different systems and outcomes for subsequent decisions.

as every asylum application lodged, apart from those withdrawn, receives a decision. It is, in this sense, a "closed" system with applications and outcomes, although cohort data is needed to be able to track applications from the beginning until the end of the asylum procedure. Asylum data also have links to return data as those who receive a negative decision on their asylum application are issued a return decision, a flow indicator in itself, which links to another indicator – actual returns. Indicators 5, 3 and 4 are therefore, procedurally linked.

This said, there are some important caveats when using asylum data and related indicators. We discuss this in more detail below, but to give an example, Savatic et al. (2024) recently concluded that "borders start with numbers." Based on a novel quantitative analysis of Eurostat's data on asylum decisions and FRONTEX's data on irregular border crossings, they found that between 2009 and 2021, the majority of individuals crossing borders and categorised as "irregular/illegal" (55.4%) were, in fact, "likely refugees." In 2015, they estimated that the same could be as high as 75.5%. Overall, their analysis demonstrates the potential for data to support political narratives regarding events at Europe's borders and the corresponding actions needed. It challenges prevailing discourses surrounding "irregular/illegal" migration flows and perceived crises across the Global North. This is an important finding because irregular border crossings as reported by FRONTEX are widely used as an indication of irregular flows.[16] As already underscored in the CLANDESTINO Project (2009a, p. 15), the public perception of flows is dominated by one particular flow – the irregular inflow over land or sea borders.

Decisions withdrawing status granted at first instance decision (Indicator 6) is also arguably a reliable status-related flow indicator, but only with respect to asylum, as it does not cover other possible withdrawn status for example, those pertaining to the termination or revocation of other residence permits. Finally, TCNs found to be illegally present inland because of illegal entry (Indicator 2) indicates secondary flows within the EU, but the dataset is likewise rather limited as it only covers irregular migrants apprehended by authorities.

Overall, there have been improvements in the collection and production of irregular migration data since CLANDESTINO. However, while there are more flow indicators available now, how the observations captured link to the "unobserved" which can serve as a basis for the estimation of irregular flows remains a challenge. To establish this link, more individual level and cohort data is needed. Given the lack of flow estimates, it is also not possible at the moment to make conclusions about the consistency of stock and flow measures i.e., whether the stock at $t+1$ is equal to the stock at time t + inflows – outflows in the same period.

Limitations of data available on Eurostat

There is a good amount of data on irregular flows, particularly geographic and status-related flows available on Eurostat. To ensure quality, controls and validation checks[17] on raw data

[16] And as it is, also in this working paper.
[17] The quality management report for EIL statistics are available here: https://ec.europa.eu/eurostat/cache/metadata/en/migr_eil_esms.htm#quality_mgmnt. Those for asylum decisions are here: https://ec.europa.eu/eurostat/cache/metadata/en/migr_asydec_esms.htm#quality_mgmnt.

transmitted by the national data providers are systematically carried out. More specifically, they check if the "totals" of a variable across different breakdowns are consistent. Still, there are important considerations when using the datasets available from the platform, owing primarily to limitations of received country level data (Eurostat, n.d.-e). Among them are (compiled from Eurostat, n.d.-b; Eurostat, n.d.-d , unless otherwise indicated):

- *Double counting at the EU level*

Data on TCNs illegally present[18] or ordered to leave are consolidated at the Member State level, not at the EU level. This means that the same TCN can be recorded as being illegally present or ordered to leave in several Member States during the same reference period. As a result, the EU total may be overestimated. In theory, this can also be the case for the total number of returns, but it is less likely that a TCN would have left several Member States after an order to leave during the same reference period, particularly if we consider that TCNs subject to the Dublin procedure are excluded from EIL statistics.

- *Undercounting at the country level*

EIL statistics provide the number of TCNs found to be illegally present in a Member State but do not include those in an irregular situation or who have not been detected. And though it can be assumed that a significant proportion of asylum seekers entered the EU illegally, EIL statistics do not provide statistical information on this.

- *Missing cohort data*

Eurostat's EIL statistics do not collect data for specific individuals. For data producers, users, and enforcement authorities, this poses a problem for returns, but also for status flows. It does not allow for tracking the different flows associated with a person's migration trajectory, say for example, from receiving a negative asylum decision (status-related inflow), being issued a return decision (geographic outflow), and successfully appealing and overturning a negative decision (status-related outflow).

- *Geographical comparability – variations at the country level*

Due to the recent implementation of the enforcement of immigration legislation data collection, some methodological and administrative differences still exist between the Member States. Some countries are in the process of harmonisation of the definitions, reducing conceptual disparities and changing data availability and completeness status for some categories of data. The national framework has an important impact on the resulting figures (i.e., the data comparability between the Member States is limited due to the national-specific rules and procedures). For asylum statistics, the geographical comparability across countries may be affected by the fact that in some cases an applicant may submit a request in more than one country during the same reference period.

- *Temporal comparability*

The enforcement of immigration legislation statistics should ideally be compiled based on the same methodology, and the outputs should be comparable between years. However, political developments impact temporal comparability, for example, the accession of Croatia to the EU in 2013 and the departure of the UK in 2020. Further, due to the ongoing

[18] Again, and as noted in Footnote 11, please note that Eurostat terminology is used here.

methodological improvements which occur at different reference periods, Member States may apply different rules for the same years for some categories of data, resulting in issues with comparability over time.

- *Time lag in recording asylum statistics*

In addition to the issue of comparability over time, the number of asylum applicants and the number of first instance decisions during the same reference period differs. This is due to the time lag between the date of asylum application and the date of decision. The duration of this time lag may vary considerably depending on the national asylum procedure and the administrative workload. Asylum applications lodged in one reference period may therefore result in decisions in a later period, while some asylum decisions reported for that period may relate to the applications lodged in previous reference periods (Eurostat, n.d.-f).

- *Interoperability across EU systems*

Frequently highlighted during the interviews conducted for this report is the issue of data sharing and interoperability across the EU's multiple data systems which were developed independently over time for different purposes. There are currently three large-scale IT systems deployed at the EU's external borders – the Schengen Information System (SIS); the European Dactyloscopy Database (Eurodac); and the Visa Information System (VIS). These systems are not directly linked and the European Commission acknowledges that "national authorities (border guards, customs authorities, police officers and judicial authorities) might have difficulties making connections between pieces of information, to get a clear picture" (European Commission, n.d.). Three other systems are set for implementation – the Entry/Exit System (EES); the European Travel Information and Authorisation System (ETIAS), and the European Criminal Records Information System on Third-Country Nationals (ECRIS-TCN). Efforts are underway to link these systems, led by eu-LISA under the interoperability framework. [19] It should be noted that while these systems generate data, they are primarily intended for border management and security purposes.

In summary, the limitations of Eurostat data reflect fundamental challenges in measuring irregular flows. Central to the issue is data comparability, as already highlighted in the CLANDESTINO Project (2009a) and subsequent studies (see for example, Raymer et al., 2013; Santamaria & Vespe, 2018). It is a challenge not just because national measurement frameworks differ but fundamentally because irregular migration is influenced by political and legal factors and different countries have different approaches towards irregular migration.

EU level data and policymaking

Policymakers interviewed for this report acknowledged that knowing the data is important for informed policy- and decision-making, and so is understanding the reasons underpinning the production of data. They highlighted the use of irregular migration data not just for border management and risk assessment but also for situational awareness, contingency planning,

[19] For details, including the underlying privacy and security principles of the *New EU Interoperability Framework*, please see: https://ec.europa.eu/isa2/sites/default/files/eif_brochure_final.pdf.

and emergency preparedness. This means that data is used in more ways beyond the monitoring of irregular migratory flows. A more detailed discussion on the needs and usage of irregular migration data by different stakeholders is presented in MIrreM D2.1 – *Briefing paper: The different uses of data on irregular migration* (Slootjes et al., 2023).[20] Here below, we outline the more salient insights shared during the interviews on the topic of irregular migration data and policymaking.

- *Data often indicates challenges, but they should be used for designing solutions*

Data on irregular migration is often taken as indicative of problems. A better approach is to use good quality data to understand the challenges better and to design appropriate solutions. Irregular migration data, in particular, can be used to improve cooperation among countries of origin and destination.

- *Data is vital to ensure the proper implementation of policies*

Policies can be based on facts, rational approaches, or beliefs. However, policies grounded solely in beliefs often prove to be incorrect. To ensure effective implementation, having evidence supporting the chosen course of action is crucial.

- *There is a continuous feedback loop in the production of data and policies*

The EU has the mandate to enact regulations and guidelines aimed at border control and migration governance more broadly. The policies enacted not only generate data but are also heavily influenced by existing data. These policies are not arbitrary but rather grounded in evidence-based practices, where decision-making is informed by a continuous cycle of data collection, analysis, and adaptation.

- *Fragmented data can be manipulated and weaponised*

Data can be manipulated to exaggerate local issues as global phenomena. This can divert attention from more significant problems with greater impact. Good quality data can help ensure that resources and efforts are not misallocated, and pressing issues are properly and timely addressed.

- *Forecasting migration flows requires accountability*

Given data and methodological limitations, forecasting migration flows should be subject to robust validation. A 'right' or 'wrong' forecast will have implications and as such, accountability should be integral to the whole process. There should be transparency, accuracy, and ethical conduct in predicting future migration patterns.

- *Forecasting irregular migration flows is a complex task*

Forecasting irregular migration flows is a complex task due to the many variables involved and the unpredictability of new crises or events. While forecasting techniques and data analysis have improved, it is still challenging to accurately predict future migration patterns.

- *Asylum flows are not necessarily irregular migration flows*

It is important to separate irregular migrants from asylum seekers. Irregular migration flows may include those who will claim asylum and they are entitled to a provisional status hence,

[20] The paper is available here: https://zenodo.org/records/7589494.

not irregular. Negative asylum decisions are important indicators, but more important than knowing the numbers is understanding the grounds for a positive or a negative decision, i.e., what exactly is the basis for the granting or denial of protection.

- *Reflecting on data needs and privacy rights is essential*

Stakeholders should reflect on what data is really needed for the purpose they have outlined, instead of adopting a "collect all the data" approach. They should also ensure that data needs are in line with privacy and data protection rights. On one hand, data can provide valuable insights and benefits for various purposes, for example, securing the borders and minimising risks. On the other hand, protecting individuals' privacy rights (regardless of migration background or status) is essential to prevent misuse or unauthorised access to sensitive information. Navigating the potentially competing interests between data production and privacy protection requires implementing safeguards, which the EU is working hard to address.

- *Technology has a lot of potential but risks should be carefully assessed*

Some stakeholders, including NGOs and CSOs supporting migrants, have expressed resistance to the use of technology in border management[21] and migration governance more broadly, citing concerns that it can lead to discrimination and compromise privacy. However, technology systems can be designed in such a way that the processes involved promote human rights by default. There is significant potential to leverage digital technologies, but the risks should be carefully assessed.

Quantification creates and reinforces categories and social boundaries; it also facilitates surveillance and control. Given its tremendous power to persuade and "enduring appeal and utility," the tendency to grant numbers authority to represent reality requires caution and ethical orientation (Espeland & Stevens, 2008, p. 432). These dynamics are certainly at play with regard to irregular migration data. Vollmer (2011, p. 330) for example, discusses the "number games" prominent in policy discourses on irregular migration in the EU and notes that different stakeholders – state and non-state actors alike – use numbers in different ways to put forward their strategic interests.

Overall, there is recognition of the importance of data in policymaking while underscoring current limitations and potential pitfalls when such data is used without full transparency and accountability. Forecasting (irregular) migration flows in particular, is a complex process that requires careful balancing of data needs and ethical considerations.[22] This is especially the case when using new digital technologies. Policies should be grounded in evidence to ensure effective implementation rather than solely on beliefs or assumptions. However, there is a two-way relationship between policies and data because policies also contribute to data production. Decision-making then should involve a continuous cycle of data collection, analysis, and adaptation.

[21] For a comprehensive discussion of border control technologies in their implications, please see: Amelung, N., & Galis, V. (2023). Border control technologies: introduction. *Science as Culture*, *32*(3), 323-343. https://doi.org/10.1080/09505431.2023.2234932

[22] For more information, please refer to: Cyrus, N. (2023). *Ethical Benchmarking in Irregular Migration Research, MIrreM Working Paper No. 03.* https://doi.org/10.5281/zenodo.8389189.

3.2 OTHER EU LEVEL DATA

FRONTEX: Annual Risk Analysis

In recent years, FRONTEX has emerged as the "authoritative" source of information on migration flows. Data published in its annual Risk Analysis Reports (ARA)[23] are frequently referenced in mainstream media (Kalkman, 2021). It collects data on geographic flows, using detected illegal border crossings[24] at the EU's external border as an indicator.

As illustrated in Figure 5, between 2008 and 2023, the number of detections peaked in 2015 at the height of the so-called European migration "crisis." It was on a steady decline until 2020, and figures started to climb starting 2021, indicating a consistent upward trend over the past three years. For 2023, the Central Mediterranean was the most active migratory route into the EU, accounting for two out of every five irregular crossings (41%) in 2023, followed by the Western Balkans (26%) and Eastern Mediterranean (16%) (FRONTEX, 2024b).

Despite the widespread accessibility and use of FRONTEX data and efforts to improve its methodological approach (please see Annex 6 for details), there are still significant limitations to its quality in terms of documentation, and validity and reliability. In addition to the explanation provided in Table 9, recording the number of incidents instead of persons calls attention to the issue of overcounting. If an individual attempts to cross the border multiple times within a short period, each attempt is likely recorded as a separate incident. This can inflate the number of recorded incidents compared to the actual number of individuals attempting to cross. FRONTEX may also record the same incident multiple times if it involves different authorities or occurs in multiple locations. Finally, as noted above, a recent study has demonstrated that the majority of individuals crossing borders and categorised as "irregular/illegal" were, in fact, "likely refugees" (Savatic et al., 2024). This confronts dominant narratives regarding "irregular/illegal" migration flows and perceived crises across the Global North.

[23] https://www.frontex.europa.eu/what-we-do/monitoring-and-risk-analysis/risk-analysis/risk-analysis/

[24] In some press releases, FRONTEX refers to "irregular border crossings" however, the term "illegal border crossings" is used in the ARA. It refers to cases of crossings of the external borders of the EU considered unauthorised at the time of the crossing under the Schengen Borders Code.

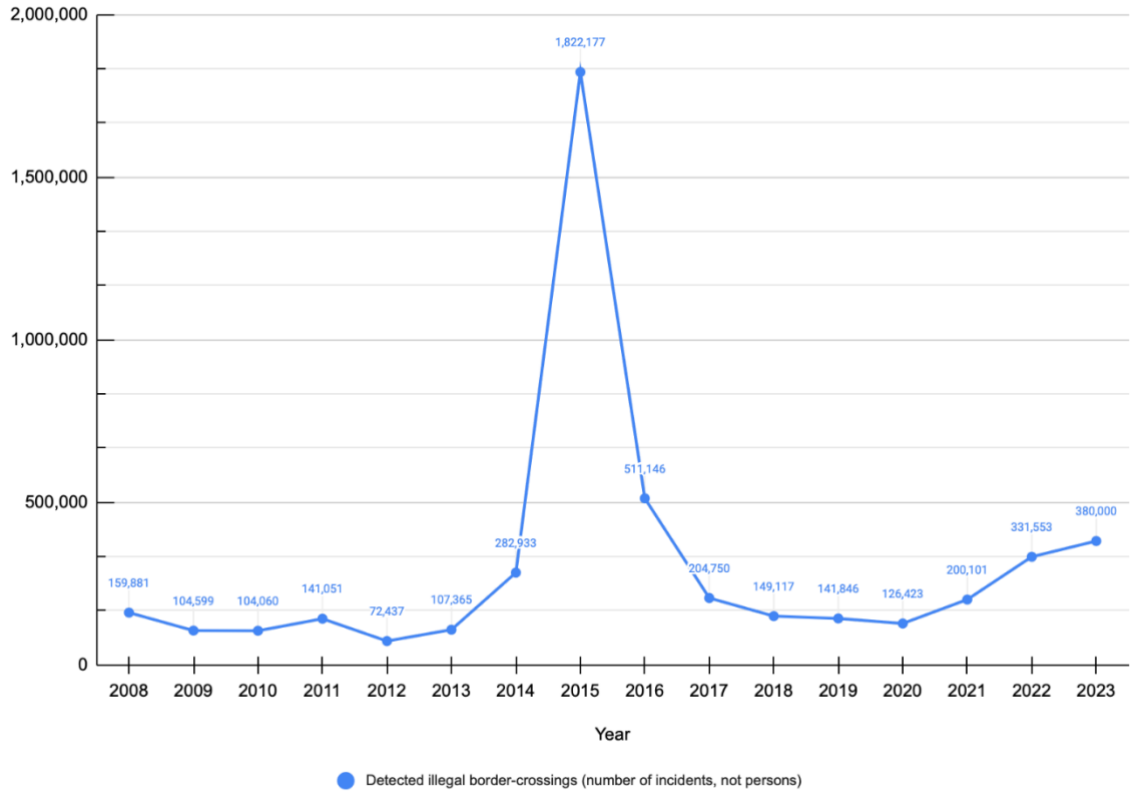


Figure 5. Detected illegal border crossings at the EU's external border, annual aggregated data from FRONTEX

Notes: Data have been collated from FRONTEX's *Annual Risk Analysis* reports. They refer to detections or incidents of illegal border-crossing rather than the number of persons, as the same person may cross the external border several times. However, there is currently no EU system in place capable of tracing each person's movements following an illegal border-crossing. Therefore, it is not possible to establish the precise number of persons who have illegally crossed the external border.

Table 9. Quality assessment of Frontex's data on detected illegal border crossings

Criteria	Quality	Explanation
Accessibility	High	Data from the *Annual Risk Analysis* is publicly available and electronically accessible without requiring permissions.
Documentation	Medium	The documentation outlines in the methodology used for the risk analysis, including the adoption of common indicators and the monitoring of these indicators. It describes the data collected through the FRAN (FRONTEX Risk Analysis Network), highlighting key indicators related to illegal border-crossing, refusals of entry, and other related statistics. However, the documentation does not provide comprehensive details about the specific methods used to collect and analyse the data. It mentions that data are compiled and analysed on a quarterly basis and that the data are used for management purposes. Still, it lacks information on the specific methodologies, data collection procedures, and quality assurance measures.
Validity and reliability	Medium	The data extracted for illegal border crossings exhibits challenges in terms of validity and reliability. Notably, variations in how incidents

		are defined and reported can lead to discrepancies in the data. It is also acknowledged in the reports that Member States' data processed by FRONTEX are not treated as official statistics and may occasionally vary from those officially published by national authorities, which raises concerns about the quality and reliability of the data. This inconsistency in data presentation raises doubts about the accuracy and completeness of the data, particularly when comparing statistics across different time frames and routes.

IOM: Missing Migrants Project (MMP)

The International Organization for Migration (IOM) launched MMP in 2013, following the tragic events near the Italian island of Lampedusa where 368 migrants died in the sinking of two boats (IOM, 2020). The MMP counts migrants all over the world who have died or disappeared at the external borders of states, or in the process of migration towards an international destination, regardless of their legal status.[25] It records only those migrants who die during their journey to a country different from their country of residence. Since 2016, MMP has published data for journeys through the Mediterranean[26] under the rubric of "attempted crossings into the EU" which, in addition to migrant deaths and disappearances, uses two additional indicators – "interceptions at sea" and "irregular arrivals in Europe" – which both pertain, at face value, to geographic flows.[27] IOM notes that the Mediterranean Sea is where irregular migration to Europe is most visible. The data are based on information from national authorities (Spanish Ministry of the Interior, Italian Ministry of Interior, Hellenic Coast Guard, Turkish Coast Guard, Libyan Coast Guard, Algerian Ministry of Interior, Tunisian National Guard, and Tunisian Ministry of Interior), NGOs and IOM staff in the field collected by IOM's Displacement Tracking Matrix (DTM) (IOM, n.d.-a). As shown in Figure 6, while data from MMP is not exactly comparable with that from FRONTEX presented previously, the same trend emerges – falling numbers from 2016, and steadily rising again since 2021. For both datasets, 2020 saw a decline which can be attributed to the COVID-19 pandemic.

IOM does not provide an explicit definition of "interceptions at sea" and "irregular arrivals in Europe." However, it is very clear about the scope of the data covered under "deaths and disappearances" (IOM, n.d.-b):

> Missing Migrants Project data include the deaths of migrants who die in transportation accidents, shipwrecks, violent attacks, or due to medical complications during their journeys. It also includes the number of corpses found at border crossings that are categorized as the bodies of migrants, on the basis of belongings and/or the characteristics of the death… The count excludes deaths that occur in immigration detention facilities or after deportation to a migrant's homeland, as well as deaths more loosely connected with migrants´ irregular status, such as those resulting from labour exploitation. Migrants who die or go missing after they are established in a new home are also not included in the data, so deaths in refugee camps or housing are

[25] On 06 March 2024, MMP released a report indicating that in ten years of data collection, 2023 is the "deadliest year on record" with nearly 8,600 recorded deaths globally.
[26] https://missingmigrants.iom.int/region/mediterranean
[27] In the US, a similar effort is carried out by the US Border Patrol through its *Southwest Border Missing Migrant Program.*

excluded. The deaths of internally displaced persons who die within their country of origin are also excluded.

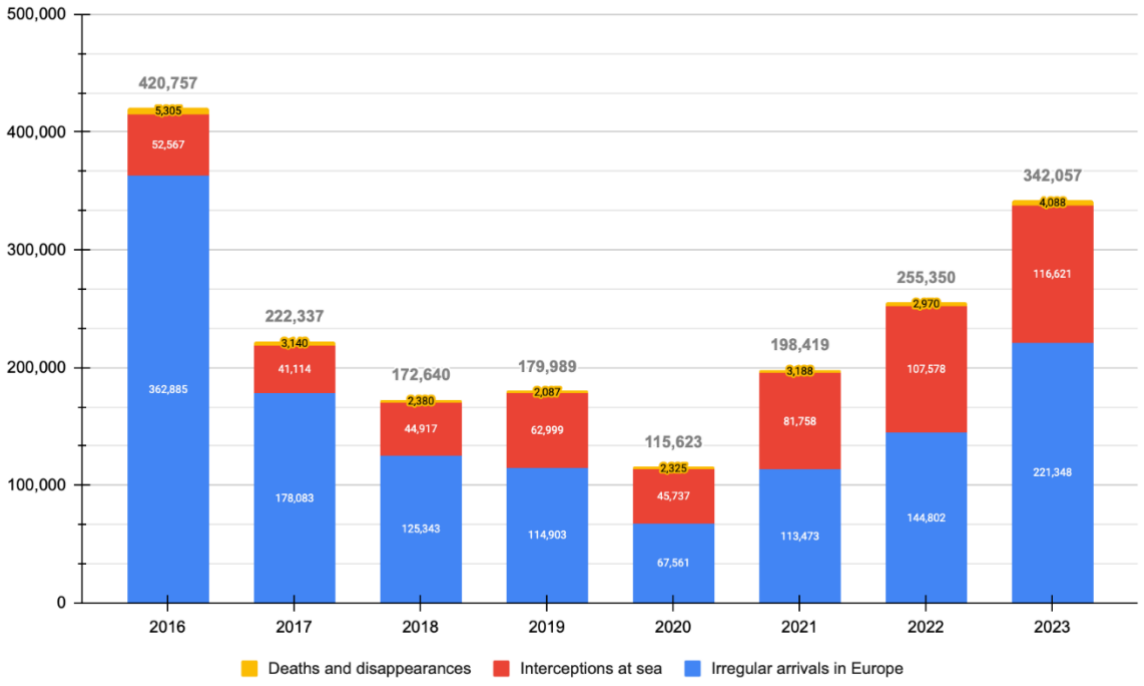


Figure 6. Attempted crossings into the EU, annual aggregated data from IOM's Missing Migrants Project

As noted in the Introduction, statistics on migrant fatalities are not directly relevant when accounting for the demographic balance between irregular in- and outflows in destination countries (Kraler & Reichel, 2011). However, statistics on migrant deaths and disappearances should not be overlooked because the majority of such events happen in the context of irregular migration (GMDAC, 2020). Data collected by MMP is presented in Figure 7. The highest numbers recorded were in 2016, and the lowest in 2019. Between 2016 and 2023, over 25,000 deaths and disappearances have been recorded. As noted by FRONTEX, the Central Mediterranean[28] is the most active migratory route into the EU (FRONTEX, 2024b), which likely explains the higher numbers of fatalities recorded along this route compared to others.

Our assessment of the quality of IOM data is presented in Table 10. The result is favourable while noting that the lack of data from before 2016[29], unfortunately, prevents a retrospective analysis of longer-term migration trends.

[28] The Central Mediterranean route is the overseas crossing from North Africa to Italy and, to a lesser degree, Malta. Those migrating on this route generally aim to reach Italian shores but leave from a variety of North African countries bordering the Mediterranean.

[29] IOM started collecting data in 2014, but the data collection method was changed, and comparable data has been available from 2016 onwards.

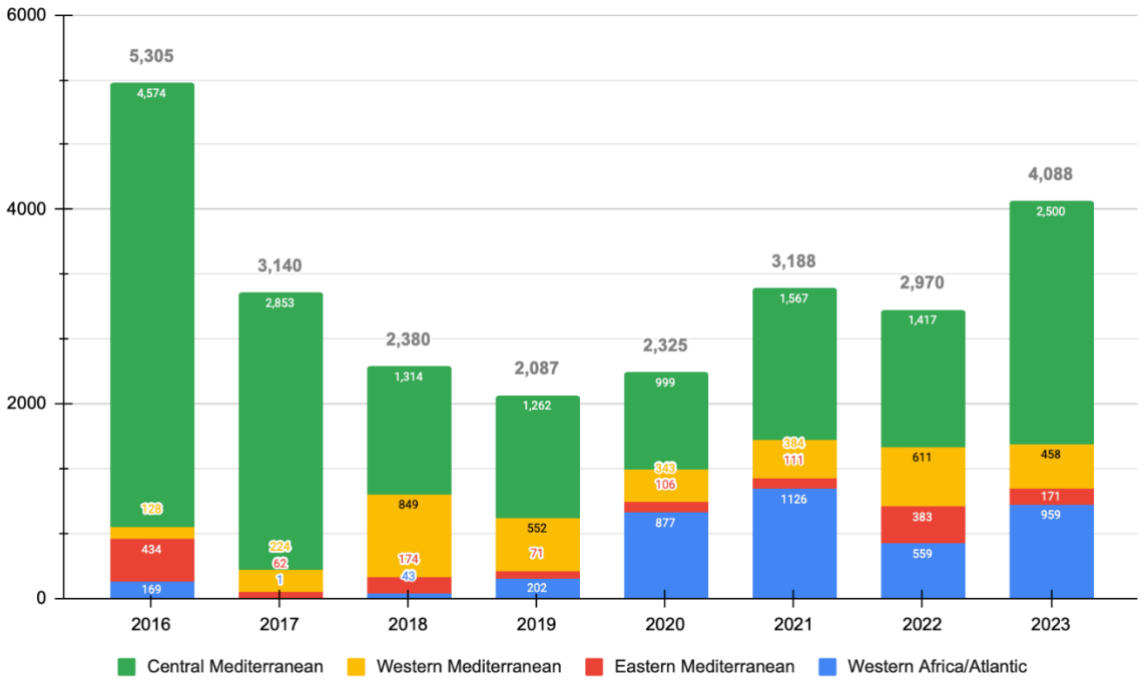


Figure 7. Recorded migrant deaths and disappearances by route, annual data from IOM's MMP

Table 10. Quality assessment of data collected in IOM's Missing Migrants Project

Criteria	Quality	Explanation
Accessibility	High	Data from IOM's Missing Migrants Project is publicly available and electronically accessible without requiring permissions.
Documentation	High	The Data Collection Guidelines is publicly available, updated in 2020, offer detailed information on the project's background, delineating what is encompassed and excluded in the data. The section on variables provides insight into the essential components required for input into the Missing Migrants Project Database. Clear instructions on submitting data, including a provided form and contact information, enhance accessibility. While the guidelines prioritize transparency by emphasising the necessity for improved data coverage and completeness, the presence of a comprehensive quality report is not explicitly addressed. Nevertheless, the guidelines provide a substantial level of detail about the project's data collection process, contributing to overall transparency and accessibility.
Validity and reliability	Medium	While the data from 2016 onwards is consistently disaggregated across diverse dimensions and offers valuable insights into migration patterns, it is imperative to acknowledge certain limitations. The reliability of the analysis is strengthened by the consistent disaggregation by year and specific dimensions, including total arrivals and transits, border type, and monthly breakdowns. However, a notable limitation arises from the absence of data before 2016, restricting the ability to conduct a more extensive

		retrospective analysis and comprehend long-term trends in migration. Furthermore, for migrant deaths and disappearances specifically, it is important to note that persons are counted regardless of their legal status.

IOM recognises inherent challenges with collecting data on migrant fatalities. Deaths often occur in remote areas chosen for their lack of legal routes, making bodies difficult to find and leading to underreporting; deaths at sea may result in many bodies never being recovered, compounded by the absence of passenger lists; survivors may fear reprisals from criminal actors or border guards, hindering reporting; and official reporting by states on deaths of non-nationals is scarce, with data often reliant on sporadic media coverage, leading to overrepresentation in regions with better media coverage and underrepresentation in areas with high volumes of irregular migration, despite the substantial dangers migrants face there. As such, ensuring full coverage and completeness of data cannot be guaranteed. Not only is the documented number of deaths and disappearances on these migration routes likely an undercount, but also that for most recorded cases, little to no information on the individuals who die is available.

These notwithstanding, the kind of statistics collected and presented in MMP illustrates the increasing role of other actors in collecting data and puts the spotlight on border management and border control practices (Pécoud, 2020) as well as the consequences of increased enforcement (Williams & Mountz, 2018). Recently, the office of the European Parliament's Ombudsman investigated the June 2023 deaths of 600 people in the Mediterranean. They found that while FRONTEX "acted in accordance with the legal rules and procedures," those rules contravene "the EU's commitment to saving migrant lives" (O'Reilly, 2024, para. 12). This underscores the urgent need for a re-evaluation of border control strategies to align with humanitarian principles, and the professed EU commitment to promoting fundamental rights, human dignity and safety (FRONTEX, 2024a).

From this exercise overall, we found that apart from Eurostat, there are only a few other sources for irregular migration data at the European level. While this may negatively impact the diversity and depth of insights available, it can have a positively impact on consistency in terms of data usage. FRONTEX collects a huge volume of irregular migration data, particularly geographic flows, and while it is not as transparent in terms of methodology, the resulting statistics are fairly consistent across time and space. On the other hand, IOM's MMP demonstrates a commendable approach to data collection, offering a comprehensive and nuanced understanding of irregular migration trends and transparency of methodology and data limitations.

In terms of data usage, FRONTEX data, as noted earlier, is notably widely used to highlight trends in irregular migration (flows in particular) across the European space. On the other hand, IOM's MMP data is often referenced by NGOs, academics and migrant advocates, in policy debates and discussions on the mortal impact of migration policies, particularly along the EU's external borders. Heller and Pécoud (2020, pp. 495-496) note however, that compared to civil society organisations, IOM's production of statistics on migrant fatalities is an attempt at depoliticisation as it precludes the "denunciation of governments' migration control strategies." As IOM positions itself as a neutral actor, it does "not demand a reorientation of immigration policies" and limits itself instead to "making recommendations

on how to improve data collection, combat smugglers and upholding basic legal norms and minimal humanitarian protection."

Curiously, a comparison of available data on geographic flow indicators from FRONTEX, IOM and Eurostat reveal an interesting picture. The data do not allow for a full temporal comparison, and these organisations use different indicators with different definitions, but as Figure 8 shows, there are significant variations in the numbers reported though they all pertain one way or another to border apprehensions. Border-related data is further complicated by the mixing of migration flows and the unclear distinction between people needing international protection (i.e., refugees, minors…) and other migrants. Additionally, the absence of an internationally agreed-upon definition for transit migration makes data comparability more complex.

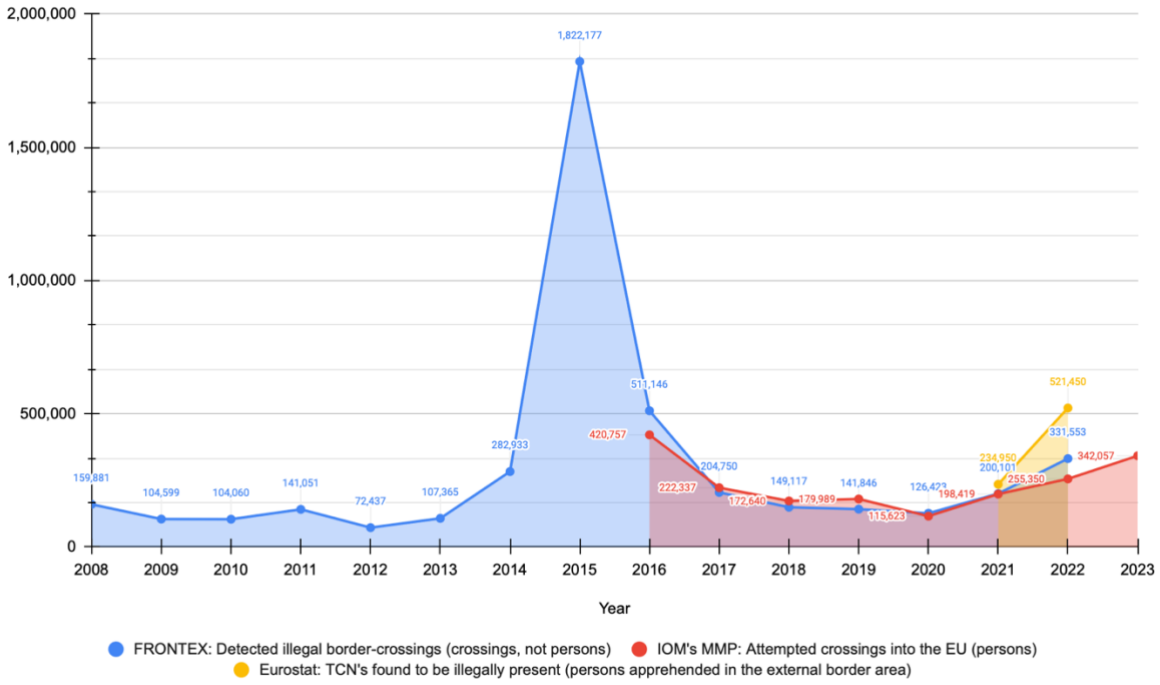


Figure 8. Comparison of selected geographic flow indicators from FRONTEX, IOM, and Eurostat

For users of migration statistics, this underscores the importance of maintaining a critical orientation when interpreting migration statistics – evaluating the sources and methodologies behind the data, considering the context in which the statistics were collected, and being mindful of potential biases or limitations. At the same time, it calls attention to the added value of cross-checking different data sources. While in this example each data source cannot be independently verified, triangulation can provide an indication of the internal validity of the available data.

3.3. COUNTRY LEVEL DATA

Overall data landscape

For all countries covered in MIrreM, government authorities and statistical agencies are the main sources of data relevant to irregular migration flows. In some cases, available data is supplemented by those collected in the data portals of international organisations like IOM, OECD, and UNHCR. Apart from Eurostat, reports from the European Migration Network (EMN) are also a good source of irregular migration data. Specialised and independent research institutes also collect and publish data on irregular flows including the Research Institute on Population and Social Policies (IRPSS), the Institute for International Political Studies (ISPI), and the ISMU Foundation Regional Observatory for Integration and Multi-Ethnicity (ORIM) in Italy; and the Migration Policy Institute (MPI), the Pew Research Center, and the Center for Migration Studies (CMS) in the US.

Available data collected by the national rapporteurs largely reflect Eurostat data. There is a noticeable emphasis on geographic flow indicators; there are some status-related flow figures, but little on demographic flows except for statistics on migrant deaths and disappearances.

In terms of geographic inflows, TCNs refused entry at the external borders, sometimes referred to as "rejection at the border" is the most widely used in the EU, also based on the data collected by the national rapporteurs. "Transit countries" Bosnia and Türkiye also use this indicator, as does the US ("inadmissibles"). Border apprehension figures, expressed in terms of "recorded detections" (UK), "interceptions" (Canada), and "illegal crossings of the state border" (Bosnia) are also reported for all countries covered in this report, albeit with different quality assessments. It is also interesting to note that some countries have publicly available and accessible data on the smuggling of migrants. This is the case for Austria, Belgium, Portugal and Türkiye. The datasets collected generally cover the geographic inflows of TCNs. There is one country, Finland, which has publicly accessible data on "negative decision of the registration of EU citizens" from 2015 onwards.

For status-related flows, two indicators are most frequently reported – negative asylum decisions and regularisations. The former is collected by Eurostat, and authorities in both Canada and the US do the same. There is no data received from "transit countries", but it should be noted that they are in fact, both a transit and a host country for people seeking asylum.[30] Not all countries have regularisation data as regularisation is not a policy implemented uniformly across all the countries covered in MIrreM.[31] In recent years, only Ireland, Italy, Poland, and Spain have implemented formal or "explicit forms of regularisation" that is, measures or programs specifically devised to offer legal pathways for migrants who do not possess the legal right to remain (Kraler et al., 2014).[32] There is

[30] Türkiye is for example, the world's leading refugee-hosting country.
[31] The issue of regularisation is addressed in MIrreM WP7.
[32] As PICUM reports, the most comprehensive study on regularisation programmes and mechanisms in the EU is the REGINE study (2009), which identified that 24 of the 27 EU Member States implemented regularisation programmes or mechanisms between 1996 and 2008, and some several times. An estimated total of 5.5 to 6 million people were regularised in that time. A total of 43 regularisation programmes were implemented in 17 EU Member States in those twelve years, involving 4.7 million

regularisation data also for France, Greece, and the Netherlands, but they are based on more "piecemeal," smaller-scale, or targeted regularisations based on humanitarian grounds, "exceptional reasons," or other "reasons worthy of consideration."[33]

Finally, for demographic flows, the data received from national rapporteurs is limited to migrant-related deaths in the US. The US Border Patrol has reported on deaths at the border since 2017 under its *Missing Migrant Program*. While birth and death registrations are legal requirements in many countries, records do not necessarily reflect migration status. The lack of data on the vital life-course events of migrants in irregular situations is discussed in more detail in the forthcoming *MIrreM Working Paper: understanding irregular migration – results of strategic case studies* (D5.4).

Data from the US

In the US, a significant portion of data on irregular migration flows to the US is produced by the U.S. Customs and Border Protection (CBP). The CBP collects detailed and extensive information under the umbrella of "Nationwide Encounters". This includes Title 8 Apprehensions, Office of Field Operations (OFO) Title 8 Inadmissibles (individuals placed into removal proceedings under Title 8 of the U.S. Code, where they may seek asylum or related relief), and Title 42 Expulsions (individuals expelled from the U.S. under Title 42 for public health reasons). The data is categorised for the Northern Land Border, Southwest Land Border, and Nationwide encounters, covering various modes of transportation such as air, land, and sea. Demographic breakdowns for USBP and OFO include Accompanied Minors (AM), Individuals in a Family Unit (FMUA), Single Adults, and Unaccompanied Children (UC) / Single Minors (Seelke et al., 2023; U.S. Customs and Border Protection (CBP), n.d.). Other enforcement statistics are compiled by the Department of Homeland Security (DHS) and the U.S. Immigration and Customs Enforcement (ICE).[34] In addition, several specialised research institutes (e.g. the Migration Policy Institute, the Pew Research Center, and the Center for Migration Studies) now provide estimates of irregular migration that allow comparisons and benchmarking of competing estimates. The estimates rely predominantly on different versions of the residual method.[35] However, while the estimates are overall of good quality given the extensive documentation provided regarding their methodology and the authoritative data on which they are based, they are for the most part, stock, rather than flow estimates.

applicants, of which at least 3.2 million were regularised. For details, please see: PICUM. (2022). *FAQ Regularisation and Access to a Secure Residence Status*. https://picum.org/wp-content/uploads/2023/08/FAQ-Regularisation-EN-2.pdf.

[33] We do not have the relevant datasets, but during the COVID-19 pandemic, some countries also implemented regularisation measures. Portugal for example, temporarily regularised the status of migrants and Italy implemented a regularisation scheme targeted for migrant workers in key sectors of the economy. Spain for its part, implemented various measures to enhance flexibility in the administrative processing of applications for regular status, aiming to prevent individuals from falling into irregularity.

[34] A comprehensive list of data sources with links is provided in Batalova, J., Shymonyak, A., & Mittelstadt, M. (2020). *Immigration Data Matters*. Migration Policy Institute (MPI). https://www.migrationpolicy.org/research/immigration-data-matters.

[35] The Migration Policy Institute has developed a methodology that uses Combined-Sample Multiple Imputation to assign legal status in the American Community Survey (ACS) and provides more granular estimates about the characteristics of the unauthorised migration population in the US.

The exception is the effort done by Robert Warren of the Center for Migration Studies to estimate the annual geographic outflow of irregular migrants in the US (Warren, 2021). Using a multi-step non-residual methodology based on the American Community Survey (ACS) data for non-citizens who arrived after 1981 and the ACS data on undocumented[36] residents in 2010, Warren estimated the number of undocumented migrants who left the US yearly from 2010 to 2019. This number that "left the population" in a year is computed as net change minus arrivals. Essentially, the term refers to the number that was in the population at the beginning of a year and was no longer in the population at the end of the year. The number that left the population includes different components for the foreign-born population and undocumented population. The raw data used in the estimates are not available, but in terms of documentation, reliability and methodology, the estimates appear to be of high quality.[37]

Relative to EU countries, the US appears to have a significantly more systematic and comprehensive approach to the collection of irregular migration data although the indicators and definitions used are not necessarily comparable.[38] The US is of course one country, as opposed to 27 in the EU with different distinct immigration and data collection systems. Efforts toward harmonisation are currently underway in the EU (e.g., concerning asylum policies), and substantial cooperation has been achieved mainly in the domain of border management. New systems such as the EES and ETIAS are in the pipeline, and while these are tools designed primarily for border and internal security, it is also anticipated that they will help generate better quality (irregular) migration data.

In summary, EU Member States submit EIL statistics to Eurostat, but more indicators of irregular flows are available at the country level, albeit with different qualities. The UK, USA, Canada, and the "transit countries" covered in MIrreM have their respective datasets which do not necessarily have a counterpart in EU-level data. The reader is advised to consult the MIrreM public database for details.

Overall quality assessment of country level data

Detailed assessments of every relevant dataset received from the rapporteurs based on MIrreM's quality assessment criteria are also accessible in the public database. The rapporteurs were also requested to provide a quality assessment of irregular migration data from each country more generally (covering both stocks and flows) and the results are summarised in the following Table.

[36] The original terminology used in the paper is reproduced here.
[37] Please consult the MIrreM public database for the detailed quality assessment.
[38] For example, in the US context, the term "unauthorized migrants" is used to refer to individuals without legal status, as well as those with temporary statuses, such as Deferred Action for Childhood Arrivals (DACA) and Temporary Protected Status (TPS) holders, and those who benefit from executive discretion and relief from deportation. Following this, one might argue that the EU would have much more comprehensive data if the comparison were to include all individuals with provisional statuses (TPS, humanitarian protection, asylum seekers, etc.), which, in the present case, are outside the scope of the MIrreM taxonomy of migrants with precarious status.

Table 11. Overall quality assessment of country level irregular migration data (stocks and flows) compiled from the country context documents prepared by rapporteurs

Country	Quality assessment
	EU countries
Austria	Estimates are available only to a very limited extent, and indicators are **difficult to assess** because categories are often combined (e.g., inland and border apprehensions) and data refer to cases instead of persons. Data sources often **lack more comprehensive information** that would allow for quality assessment, such as calculation methods, inclusion and exclusion criteria, information on double counting, etc.
Belgium	Data is generally accessible but there are **significant limitations** in terms of documentation, validity, reliability, and methodology. It appears that the data may be selective and contain internal contradictions, which raises doubts about its accuracy.
France	Data is also generally accessible, but they appear to have **significant limitations** in terms of documentation, validity, reliability, and methodology. There is limited information available regarding the data sources, collection methods, and overall data quality. Often. there is a single data source, which can be a limitation.
Finland	The quality of contemporary data varies. There is a **lack of systematic longitudinal data** on the number of migrants present, and the data on migration flows is only partially covered.
Germany	As an overall assessment, the use of the available administrative data on irregular migration for scientific purposes **requires a careful screening**.
Greece	Overall data on irregular migration flows is directly relevant to external borders control and is collected by the state authorities who are primary responsible to collect, process and disseminate the statistics. There is electronically **accessible information** on most of the indicators.
Ireland	The quality is **poor/inconsistent** and irregular migration data is provided in a piecemeal way. a substantial proportion of the data collected was provided through what can only be described as informal and/or ad hoc correspondence with Department of Justice officials. Further, a number of inconsistencies were identified as part of this data collection process.
Italy	Irregular migration in Italy **is not satisfactory**. Elaborations are accessible but rare and mainly unexplained; raw data is often not available and methodological choices are not explained.
Netherlands	Data is **generally of good quality**, but most of the data available pertain to a specific set of irregular migrants namely, failed asylum seekers, or those who come in contact with the authorities. This means that available data may not adequately capture other groups including irregularly staying family migrants and labour migrants.
Poland	Data are compiled by several institutions and are not accessible in one place but despite this data dispersion, the access, documentation, and quality of official data by state institutions is **of rather good quality**. The data are comparable from one year to another (very few breaks in time series), mostly defined and collected in the same way, the reports including documentation are based on the same templates.
Portugal	The data on irregular migration is **not available to the public** apart from the number of candidates that obtained an authorisation of residence on an annual basis on behalf of the exceptional regularisation regime. This data is objective but does not provide information over the timespan required by the Portuguese authorities to analyse the individual request for the exceptional regularisation.

Spain	The overall assessment of the quality of data on irregular migration in Spain is **positive**. The methods for estimating the irregular population use official, public and accessible data from administrative registers. The same apply for most of the different indicators on international protection, irregular entries and expulsions. However, the treatment of these data and the dissemination of information by the administration is **uneven**, as it focuses on communicating especially some facets of the phenomenon while paying less attention to others.
	Other countries
Bosnia and Herzegovina	The Ministry of Security (MoS) maintains an Information System on Migration (ISM) which is the primary source of data for international migration including irregular migration. The quality of administrative records integrated in the ISM are **good**, as the coverage is full.
Canada	There is a divergence in the definitional and conceptual understanding of "irregular migration" data in the Canadian context that diverges from European definitions. **The types of datasets laid out in accessible ways are a challenge to procure**. Migration (flows) and population (stocks) datasets, needed to estimate irregular migration, are limited to aggregated annual time series at national or provincial levels, with potential discrepancies in the timeliness and frequency of data collection.
Morocco Tunisia	Collecting data on irregular migration in Morocco and Tunisia, remains a **complex and sensitive** endeavour. Underreporting and data bias appeared to be significant concerns in the two countries, potentially distorting the accuracy of the collected information. There is limited transparency from government authorities producing data.
Türkiye	There is almost **no clarification on the methods and techniques** used by the organisations publishing irregular data. It is also difficult to say that different sources have a common terminology. In this respect, we do not have enough information to make any assessment of the quality of irregular migration data.
UK	In terms of flow indicators, the overall quality of irregular migration data is **high**. The number and types of indicators generated (primarily by the Home Office) has increased significantly and even over the last few years, as small boat arrivals have become a significant policy issue. There are now more detailed and more clearly flagged explanatory notes and methods, such as definitions, changes to methodology, limitations, etc. and access to data has improved.
USA	Both the availability and quality of irregular migration data and estimates **has been increasing** over the past decades in the United States. Several government bodies (e.g. Department of Homeland Security (DHS) and US Border Patrol) and specialised research institutes (e.g. the Migration Policy Institute, the Pew Research Center, and the Center for Migration Studies) now provide estimates of irregular migration that allow comparisons and benchmarking of competing estimates.

From this, we see that there are variations in the overall quality of irregular migration data across countries – some are of good quality and others have significant limitations. On this basis, emphasis on care and caution is again in order when using these data. This means acknowledging the potential discrepancies or inaccuracies in the data and taking steps to mitigate their impact when utilising them for analysis or decision-making purposes.

Data gaps at the country level

The gaps identified by the national rapporteurs generally align with those identified at the EU level. More specifically, these are:

- *Decentralised data collection points*

No single state institution estimates irregular flows and stocks because no single institution has a scope of competencies and objectives covering the total of irregular migration. Some institutions cover only border traffic, others deal with residence procedures, and others are concerned with labour market issues.

- *Limitations on intra-country data sharing and interoperability*

Related to the previous point, within countries, the exchange of data among authorities is limited or prohibited by data protection regulations. Different agencies and actors in migration governance often use different data collection and management systems. When data sharing is possible, different methods of relaying the data (e.g., manual, automated, etc.) often cause time lags and inaccuracies.

- *Limited availability of estimates*

While no country apart from the US has estimates on irregular flows, some countries have estimates on irregular stocks, but even then, the estimates typically cover a limited period using a specific selection of data from ministries or government agencies. Because of this, migration dynamics are unlikely fully captured in available estimates.

- *Lack of data on demographic flows*

While birth and death registrations are fundamental aspects of civil registration systems across the world, there is a noticeable lack of systematic data on births and deaths and irregularity. There are numerous studies on the qualitative characteristics of different irregular migration populations but statistics on births and deaths and other vital life course events are not available.

- *Lack of historical data*

Despite European regulations, some countries do not have complete data for all flow indicators which may be useful in the analysis of long-term trends. This is exemplified in the result of the percentage of missing values test discussed above.

- *Lack of detailed demographic breakdowns*

Many rapporteurs mentioned that publicly available data often lacked detailed demographic breakdowns, such as age, gender, or nationality, which could have provided a more nuanced understanding of irregular migration patterns. As above, this is reflected in the percentage of missing values test we conducted. Except in the US, population censuses do not include any information about irregular migrants.

- *Lack of data on overstayers*

Only the US has available data on overstaying collected and reported by the Department of Homeland Security. While Eurostat has since 2021, started collecting information on overstayers as part of its EIL statistics, this is limited to those apprehended by authorities.

In many EU countries, overstaying has been identified as one of the primary sources, if not the primary source, of irregularity. However, there is currently no adequate data to substantiate this claim.

- *No specific data on irregularity*

In Canada, apart from statistics on 'irregular border crossings,' there is no publicly available data that is disaggregated or presented in a form that directly speaks to persons 'out of status.'

- *Irregular collection of irregular migration data*

In "transit countries" covered in MIrreM, the lack of systematic and standardised collection of irregular migration data has been identified as a challenge. There is no periodic or regular collection of data resulting in stark chronological gaps, and different agencies in the same country (for example, Türkiye) appear to use different terminologies and methodologies.

Many of the data gaps identified, both at the EU and the country level are not specific to flows but to irregular migration more generally. And beyond data, there is a need to recognise that ultimately, the phenomenon of irregular migration is about persons. The general lack of data that provides information on living situations, risks, and resources has also been indicated as a gap that contributes to the invisibility of the group and prevents acknowledgement not only of their presence but also of their potential.

Country-level data and policymaking

National rapporteurs also provided information on how irregular migration data is generally used in policymaking. The findings are summarised in the following Table. Again, please consult the MIrreM public database for more details.

Table 12. Use of irregular migration data (stocks and flows) in policymaking compiled from the country context documents prepared by rapporteurs

Country	Use of irregular migration data in policymaking
	EU countries
Austria	Data is **used by various ministries** for the purpose of strategic planning and the development as well as evaluation of measures. In terms of political and societal discourse, the asylum system, rejected asylum seekers and irregular migration are often discussed together and interwoven. The mixing of these issues is problematic in many respects.
Belgium	Data is used for information and prevention campaigns, policy formulation, border control, asylum and refugee policy, resource allocation, international cooperation, and evaluation and adjustment of existing legislation and policies.
France	Data is used for immigration detention and enforcement selective immigration and point-based systems, Dublin procedure, and irregular migration management.
Finland	Data on irregular migrants is used **to address significant policy matters**, primarily focusing on national security and the provision of healthcare and social services to the extent deemed essential.
Germany	Overall, except of few key figures, the **available irregular migration data is not systematically used in policy making** but serves mainly purposes to coordinate and administer migration management and monitor the outcome and effectiveness of administrative decision on the issuing or waiving of residence status. The potential

	of available administrative data for an evaluation of policies and administrational performance is not used.
Greece	Data is used **to measure the impact of policies** and implemented measures, and to develop further policies in line with current priorities – national and European. Data on irregular flows is used mainly for the protection of the borders and further case management of those who arrive.
Ireland	Policymakers in Ireland appear to primarily rely on the work of CSOs to understand the scale, trends, and characteristics of unauthorised migration. Though the ESRI (the Irish National Contact Point for the EMN) publishes an annual migration and asylum report, there is **nothing** available from the Department of Justice **which suggests that these reports are relied on for the purpose of policy making**.
Italy	Unlike stereotypes and manipulations finalised to electoral gains in policy and public discourses, accurate data estimates on irregular migrations are **rarely considered strategic for policy making**. Policymakers show more attention towards other kind of data, especially GDPs, other economic statistics, or innovation trends, apart from electoral preferences and opinions.
Netherlands	Irregular migration data is used by the Parliament to serve as basis for further development of migration policy. Data is also routinely used by authorities to identify policy priorities and action points. During the validation workshop with different stakeholders however, it appeared that irregular migration data, while referred to when asked, is **not systematically used for policymaking**, but serves more as a 'benchmark' when discussing migration.
Poland	International migration is a highly politicised issue in Poland, often used by populists to lead xenophobic discourse and achieve particular objectives. Therefore, decisions undertaken at the highest country level **not necessarily are based** on the competencies of experts from state institutions and on statistical data.
Portugal	Portuguese policy makers employ the indicators of the stocks of irregular immigrants provided by the national authorities **to manage this social phenomenon**. The data also provides an idea about the resources required to manage the pool of requests for regularisation made by irregular immigrants.
Spain	The utilisation of irregular migration data in policymaking **varies**. Certain sources clarify that there exists a connection between the changes in the irregular population, economic trends, and regulatory modifications at the state level. However, governments do not always reveal how these factors have impacted the development of their policies. Meanwhile, local municipalities and CSOs utilise data to develop effective programs and to create recommendations and influence the development and modification of migration policies.
	Other countries
Bosnia and Herzegovina	Generally, use of data in policy making is **almost non-existent**, as many policy reviews have shown. Strategies are often developed without relevant data or by designing measures that do not address the policy needs at all. Many strategies developed are then not passed in the parliaments at all or only in the final year of its implementation. Lack of coordination between different institutions is also evident, at the same level of government and even more between different levels.
Canada	The use of "irregular" migration data in policy making remains **unclear**. Figures specific to persons without status or migrants crossing borders irregularly are a challenge to publicly access. Moreover, policies directed to persons without status, e.g., regularisation initiatives have only recently been considered by the federal government, is only in its infancy stages.
Morocco	

Tunisia	Generally, irregular migration data serves as a **referential foundation for policy making** in both Morocco and Tunisia. Firstly, they play a fundamental role in shaping immigration and border control policies. Secondly, data helps measuring and allocating the required resources to border security and law enforcement agencies, migration patterns and identifying hotspots. Third, it helps international humanitarian organisations to identify and allocate the nature and volume of assistance required (shelter, healthcare, and social services). Fourth, these data are instrumentalised in diplomatic negotiations and international cooperation agreements by Morocco and Tunisia with their EU counterparts and the governments of the migrant-exporting African countries.
Türkiye	Data is generally used for **political purposes** and to give a message to society. Especially in the last two years, it is clear that the government has devoted special attention to the issue of irregular migration. This can be considered to be closely related to the upcoming local elections. Information on irregular migration is provided to the public in order to convey the message that the necessary actions are being taken to combat irregular migration. Data is also used to carry out cooperation agreements especially with international organisations for capacity building and support programmes in the fight against irregular migration.
UK	Flow data are cited in a range of political settings, such as parliamentary questions and inquiries and policy statements. However, they **do not appear to be a part of, or come from, an evidence base** on which policies are made. Rather, actors on all sides of the debate use these data in different ways to bolster their own policy positions and, for those in power, justify the policies they implement.
USA	Policymakers in the US rely on various sources of data, including government agencies, research institutions, and non-governmental organisations, to understand the scale, trends, and characteristics of unauthorised migration. They are used further to assess the cost of federal programs and to estimate the economic impact of immigration policies. Estimates of irregular migration have a significant **impact on the political discourse** and are shaping electoral statements, making them a highly sensitive information.

In the literature, a distinction is often made between the substantive and the symbolic use of data in policymaking (Slaven & Boswell, 2019). Substantive policies are primarily for *steering* the object of intervention, while symbolic ones are used in *signalling* values and intent, usually directed at voters (Ataç & Schütze, 2020, p. 120, emphasis in the original). From the information just presented, what emerges is that the use of irregular migration data in policymaking is not easy to categorically establish. There is a clear link between the two, but whether policymaking is data-driven, or data is instead policy-driven is contingent on the prevailing political environment, and the extent of the interplay between data and policymaking is difficult to ascertain.[39] Thus, while it is possible to heuristically distinguish between the substantive and symbolic uses of irregular migration data, they may not be mutually exclusive in practice. Different stakeholders also vary in terms of their data needs and usage, with authorities using data for migration governance broadly speaking (external

[39] A recent effort in this direction is this publication by Pettrachin, A., & Hadj Abdou, L. (2024). Beyond evidence-based policymaking? Exploring knowledge formation and source effects in US migration policymaking. *Policy Sciences*, 1-26. https://doi.org/10.1007/s11077-024-09523-y It analyses how policymakers in key governmental agencies in the US working on irregular and asylum-related migration routinely and rationally use information produced outside of their own organisations. They conclude that these policymakers are not merely driven by political interests, though political and ideological factors determine what counts as 'evidence in the first place' (p. 1).

and internal policies), and others like municipalities and CSOs relying on data for the provision of services and planning of interventions to support the irregular migrant population. Also noteworthy is the use of data primarily in “transit countries”, to facilitate cooperation with international and humanitarian organisations. Ultimately and from a normative standpoint, identifying the use of data in policymaking requires understanding the context and the specific circumstances underpinning the production and usage of data and the design and implementation of the policy.

4. CONCLUSION

Fluctuations in the stocks of irregular migrants reflect net changes in flows. But while many countries have available stock estimates, there is persistence in the scarcity of available flow estimates as observed in the CLANDESTINO Project. The only methodology identified that can be used directly for estimating irregular migration flows in Europe is the extrapolation of border apprehensions with a simple multiplier, using an estimated ratio of border apprehensions to non-detected illegal entries (Jandl et al., 2008). While used particularly in the absence of alternatives, there are multiple issues with the multiplier method in terms of (1) the quality and availability of apprehension data, (2) the potential for overcounting because border apprehension data refer to cases or events rather than individuals, (3) the estimation of the "correct" multiplier, (4) externalities to increased detections or apprehensions, and (5) data volatility. There are also indirect approaches based on differential or residual methodology, but CLANDESTINO found that the reliability of resulting estimates could not be established. Overall, the conclusion made was that the methodologies for analysing irregular border crossings, visa overstays, and overall **irregular migration flows lag behind the study of irregular resident stocks** (Vogel et al., 2008). A summary of the flow trends identified therein is presented in the following Table. The data on flows collected then from 12 EU member states were found to be "even more problematic" than the data on stocks that eventually, contrary to initial plans, were not included in the database (CLANDESTINO Project, 2009a, pp. 13-14).

Table 13. Summary of flow trends identified in CLANDESTINO

Type of flow	Trends
Demographic	Quantitative importance is low or largely unknown but 'causing considerable human rights concerns.'
Geographic	Most visible flows and border guard apprehensions are commonly used as indicators. However, published data from countries 'have not achieved full comparability' and there is 'even less information' on outflows (e.g., departures from the EU, movements to another EU country)
Status-related	EU enlargement and regularisation programmes in Italy, Spain and Greece resulted in status-related outflows being far higher than inflows. Visa overstaying is the most relevant inflow indicator.

Source: CLANDESTINO Project (2009b, pp. 6-7)

In view of CLANDESTINO, this working paper looks into how irregular migration flows are measured and estimated in the EU and selected other countries. It aims in particular to provide insights into the quality of the available indicators and estimates. Furthermore, it

explores the use of these measures in migration governance and policymaking. The foregoing analysis points to the following **main conclusions**:

1) Compared to stocks, there are still almost no available estimates of irregular migration flows. However, in view of findings from the CLANDESTINO Project, there are now more irregular flow indicators, particularly for geographic flows, and to some extent, also asylum-related status flows. However, there is still a scarcity of data concerning demographic flows.
2) Eurostat data as a whole is better than (individual) country-level data in terms of accessibility, documentation, and validity and reliability. Among the countries covered in MIrreM, data from the US is arguably the most valid and reliable, which also includes some flow estimates. Data from EU countries and the UK are generally comprehensive, but they are difficult to compare at the country level. For "transit countries," data appear to be highly fragmented except for Bosnia and Herzegovina.
3) Zooming in on Eurostat data, asylum data seem to exhibit the highest level of validity and reliability. Despite reporting lags due to the unpredictable timing of first instance and final decisions, negative decisions on asylum applications serve as a meaningful status-related flow indicator.
4) However, there are still significant issues with regard to the overall quality of Eurostat data, not so much in relation to the FAIR Data Principles, but with validity and reliability. In terms of external validity, the data available often only describe an aspect of the phenomenon of irregular migration instead of being representative of the whole (e.g., asylum data only capture status-related flows). Among others, there are issues with double counting or missing data, particularly when disaggregating by age and sex, which pose a challenge to measurement precision. As for internal validity, it is difficult to independently assess since the data are generated by bureaucracies with limited oversight and there are not many opportunities to cross-validate the numbers with other information.
5) Policymakers use irregular migration data for purposes extending beyond the monitoring of irregular migratory flows. In addition to border management and risk assessment, data is also used for the development of new policies, as well as for situational awareness, contingency planning, and emergency preparedness. However, the ambiguity surrounding definitions and the lack of consensus on what precisely constitutes irregular migration even among migration scholars, practitioners, and other stakeholders, increases the risks associated with the use of irregular migration data for political purposes.
6) Good quality data is essential to migration governance. However, the interest in enhancing data collection on irregular migration and generating estimates must be carefully weighed against privacy considerations and societal interests. This balance is crucial so as not to impede trust on the part of irregular migrants and hinder the public service mission of providers or support groups, civil servants, and other street-level bureaucrats who regularly come into contact with them.

Post-CLANDESTINO, scholars observe that available migration data remain often "inaccurate, inconsistent and incomplete" as they are based on differing definitions (Bijak et al., 2019, p. 471). In addition to differing definitions and measures, there are persistent and interlinked gaps based on the drivers or reasons behind migration, geographic coverage,

demographic characteristics, and time lag in the availability of data (Ahmad-Yar & Bircan, 2021). International migration flows are particularly difficult to measure, and this is the case even with advancements in technology and data science (McAuliffe & Ruhs, 2017). Several international organisations including UN DESA and the OECD have been collecting and publishing international migration flows data but different definitions and data collection methods present challenges in harmonisation and comparability (Yildiz & Abel, 2021). As there is an inherent challenge in collecting data on clandestine or irregular processes, the difficulties are even more pronounced when capturing data on *irregular* migration flows (McAuliffe & Sawyer, 2021, p. 48, emphasis added).

While some datasets can be used as proxies to generate estimates, there are no official datasets that directly measure irregular migration and irregular migrants in the EU (Vespe et al., 2017). More specifically, it remains to be the case that there are no comprehensive and consistent statistics within the EU that accurately capture the dynamics of irregular migration flows (Misiuna & Pachocka, 2018). More recently, Angenendt et al. (2023) highlighted the great interest in using AI-based tools to predict irregular migration. However, technical and methodological limitations abound, and there remains a notable disparity between the projected efficacy of the new tools and their actual usefulness.

Based on the the analysis of EU level data and the data collected by national rapporteurs, irregular migration flows are measured through statistical indicators, particularly geographic and status-related flow indicators. Given the complexity of capturing migration flows compared to stocks, there are no estimates of irregular flows currently available for the countries covered in MIrreM, except for the US which, in comparison to the EU appears to have a significantly more systematic and comprehensive approach to the collection of irregular migration data. While there is hardly any data available for demographic flows, available EU level indicators for irregular flows are generally of good quality, particularly with respect to accessibility and documentation. A statistical analysis of a select number of indicators however, revealed some limitations in terms of validity and reliability. There are also known limitations with Eurostat data more broadly concerning double- or under-counting, geographical and temporal comparability (including time lags), and finally, interoperability across EU systems.

Some of the interview participants shared optimism that the launch of the EES and ETIAS can help fill some of the gaps in migration statistics (for example, overstayers) and address issues regarding interoperability while highlighting that these systems are put in place primarily to secure EU's borders, facilitating entry into the Schengen area, and creating a "seamless travel blueprint" for all incoming and outgoing persons. They also recognise that there are important ethical and human rights considerations to address when using big data and new technologies. The systems are continuously being tested and monitored by external evaluators, as well as internally, for example, by the Fundamental Rights Officer (FRO) of FRONTEX and the EU's Fundamental Rights Agency.

Zooming in on the quality of country level data, there are variations in the overall quality across countries – some are of good quality and others have significant limitations. They are attributed to among others, decentralised data collection points, lack of historical data, and lack of detailed demographic breakdown. In Canada, there is no data that directly speaks to

persons "out of status," and among "transit countries", data is generally scarce as there is no periodic or regular collection of data similar to the practice in the US or EU Member States.

In view of these, we recommend the following **main ways forward** to advance research on irregular migration flows and to prevent misuse of migration data:

- Continue improving data quality for (selected) flow indicators for example, by investing more resources into quality checks and making cohort data available (without compromising privacy considerations);
- Acknowledge that supplemental qualitative information is essential for the validation and triangulation of quantitative data; incorporating qualitative studies into the collection of migration data should be the norm; and finally,
- Consider using accessible informational resources, such as educational videos, to mitigate the misuse of migration data for political purposes; knowing the importance of a fact-based discourse can help in ensuring that statistics on migrant populations are not manipulated or misrepresented to serve political agendas.

Overall, there is recognition of the importance of data in policymaking while underscoring current limitations and potential pitfalls when such data is used without full transparency and accountability. While not a novel observation, it is important to underscore that the potential for misuse[40] cannot be underestimated – from the presentation of statistics to the utilisation of such statistics in political decisions and policymaking. Adequate data is essential for informing migration governance and public debates, particularly on topics that generate significant interest from all levels of society. Migration is one such topic, and as it is often polarising, all the more is good quality data – accurate, frequent and timely – of critical importance. Although migration policy can be influenced by diverse and at times conflicting interests, good quality data is imperative to inform all policy domains addressing different facets of migration and aiming for improved migration management (Santamaria & Vespe, 2018). Current available data on irregular migration flows may still not be adequate to serve as basis for a fully evidence-informed policymaking. While not easy to establish categorically, irregular migration data may be said to be used primarily to legitimise policies and to create a "policy narrative" that signals government efficiency and epistemic authority. This said, relying solely on irregular migration data without considering its limitations can result in flawed analyses and policy decisions. Therefore, it is crucial to approach the use of such data with care – acknowledging its limitations, triangulating information from multiple, including qualitative sources where possible, and interpreting findings with caution. Finally, efforts to improve the quality and consistency of irregular migration data collection and reporting should be prioritised to enhance understanding and response to this complex phenomenon, ensuring that data collection is purposeful and not undertaken merely for the sake of data accumulation. Striking a balance between collecting and using data and protecting individuals' privacy rights is essential, and so is navigating between addressing perceived security threats and responding to humanitarian principles.

[40] See for example, ECRE. (2022). *Asylum statistics and the need for protection in Europe: Updated Factsheet*. https://ecre.org/wp-content/uploads/2022/12/Asylum-statistics-and-the-need-for-protection-in-Europe-final.pdf. Also, Mouzourakis, M. (2014). *'Wrong number?' The Use and Misuse of Asylum Data in the European Union*. https://www.ceps.eu/ceps-publications/wrong-number-use-and-misuse-asylum-data-european-union/.

References

Ahmad-Yar, A. W., & Bircan, T. (2021). Anatomy of a misfit: International migration statistics. *Sustainability*, *13*(7), 4032. https://doi.org/10.3390/su13074032

Amelung, N., & Galis, V. (2023). Border control technologies: introduction. *Science as Culture*, *32*(3), 323-343. https://doi.org/10.1080/09505431.2023.2234932

Angenendt, S., Koch, A., & Tjaden, J. (2023). *Predicting irregular migration: High hopes, meagre results* (SWP Research Paper, Issue. Stiftung Wissenschaft und Politik (German Institute for International and Security Affairs). https://www.swp-berlin.org/publications/products/research_papers/2023RP11_PredictingIrregularMigration.pdf

Ataç, I., & Schütze, T. (2020). Crackdown or symbolism? An analysis of post-2015 policy responses towards rejected asylum seekers in Austria. In S. Spencer & A. Triandafyllidou (Eds.), *Migrants with Irregular Status in Europe: Evolving Conceptual and Policy Challenges* (pp. 117-137). Springer Open.

Batalova, J., Shymonyak, A., & Mittelstadt, M. (2020). *Immigration Data Matters*. Migration Policy Institute (MPI). https://www.migrationpolicy.org/research/immigration-data-matters

Bellanova, R., & Glouftsios, G. (2022). Controlling the Schengen Information System (SIS II): The infrastructural politics of fragility and maintenance. *Geopolitics*, *27*(1), 160-184. https://doi.org/10.1080/14650045.2020.1830765

Bigo, D. (2020). Interoperability: A political technology for the datafication of the field of EU internal security? In D. Bigo, T. Diez, E. Fanoulis, B. Rosamond, & Y. A. Stivachtis (Eds.), *The Routledge Handbook of Critical European Studies* (pp. 400-417). Routledge

Bijak, J., Disney, G., Findlay, A. M., Forster, J. J., Smith, P. W., & Wiśniowski, A. (2019). Assessing time series models for forecasting international migration: Lessons from the United Kingdom. *Journal of Forecasting*, *38*(5), 470-487. https://doi.org/10.1002/for.2576

Cantens, T. (2018). The Political Arithmetic of Borders: Towards an Enlightened Form of Criticism. *Global Trade and Customs Journal*, *13*(3). https://doi.org/10.54648/gtcj2018015

CLANDESTINO Project. (2009a). *CLANDESTINO Project Final Report*. https://cordis.europa.eu/docs/publications/1266/126625701-6_en.pdf

CLANDESTINO Project. (2009b). *Comparative Policy Brief - Size of Irregular Migration* https://irregular-migration.net/wp-content/uploads/2021/06/ComparativePolicyBrief_SizeOfIrregularMigration_Clandestino_Nov09_2.pdf

Collyer, M., & De Haas, H. (2012). Developing dynamic categorisations of transit migration. *Population, Space and Place*, *18*(4), 468-481. https://doi.org/10.1002/psp.635

Cummings, C., Pacitto, J., Lauro, D., & Foresti, M. (2015). *Why people move: understanding the drivers and trends of migration to Europe* (ODI Working Paper Issue 430). https://odi.cdn.ngo/media/documents/10485.pdf

Cyrus, N. (2023). *Ethical Benchmarking in Irregular Migration Research, MIrreM Working Paper No. 03.* https://doi.org/10.5281/zenodo.8389189

DeBono, D. (2016). Returning and deporting irregular migrants: Not a solution to the 'refugee crisis'. *Human Geography*, *9*(2), 101-112. https://doi.org/https://doi.org/10.1177/1942778616009002

Dong, Y., & Peng, C.-Y. J. (2013). *Principled missing data methods for researchers* (SpringerPlus, Issue. https://link.springer.com/article/10.1186/2193-1801-2-222

EASO. (2015). *Latest asylum trends – 2015 overview*. https://euaa.europa.eu/sites/default/files/public/LatestAsylumTrends20151.pdf

ECRE. (2022). *Asylum statistics and the need for protection in Europe: Updated Factsheet*. https://ecre.org/wp-content/uploads/2022/12/Asylum-statistics-and-the-need-for-protection-in-Europe-final.pdf

Espeland, W. N., & Stevens, M. L. (2008). A sociology of quantification. *European Journal of Sociology / Archives Européennes de Sociologie*, *49*(3), 401-436. https://doi.org/10.1017/S0003975609000150

EUAA. (n.d.). *4.14.2.2. Review, cessation and revocation of international protection*. European Union Agency for Asylum,. Retrieved 17 Feb. 2024 from https://euaa.europa.eu/asylum-report-2023/41422-review-cessation-and-revocation-international-protection

European Commission. (n.d.). *Migration and Home Affairs - Interoperability*. European Commission,. Retrieved 20 Feb. 2024 from https://home-affairs.ec.europa.eu/policies/schengen-borders-and-visa/interoperability_en

Eurostat. (2020). *Immigration law enforcement in the EU – figures for 2019* https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20200722-1

Eurostat. (n.d.-a). *Decisions on applications and resettlement (migr_asydec)*. https://ec.europa.eu/eurostat/cache/metadata/en/migr_asydec_esms.htm

Eurostat. (n.d.-b). *Decisions on applications and resettlement (migr_asydec) - reference metadata*. https://ec.europa.eu/eurostat/cache/metadata/en/migr_asydec_esms.htm#quality_mgmnt

Eurostat. (n.d.-c). *Enforcement of Immigration Legislation (migr_eil)*. Retrieved 04 Jan. 2024 from https://ec.europa.eu/eurostat/cache/metadata/en/migr_eil_esms.htm

Eurostat. (n.d.-d). *Enforcement of Immigration Legislation (migr_eil) - Reference metadata* https://ec.europa.eu/eurostat/cache/metadata/en/migr_eil_esms.htm#quality_mgmnt

Eurostat. (n.d.-e). *Enforcement of immigration legislation statistics introduced* https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Enforcement_of_immigration_legislation_statistics_introduced

Eurostat. (n.d.-f). *Statistics Explained Glossary: Asylum applicant*. Eurostat. Retrieved 20 Feb. 2024 from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Glossary:Asylum_applicant

Eurostat. (n.d.-g). *What is the European Statistical System?* Retrieved 04 Jan. 2024 from https://ec.europa.eu/eurostat/web/european-statistical-system/overview

FRONTEX. (2024a, 23 February). *Frontex hosts high-level visit by EU Commission President and Belgian Prime Minister* https://www.frontex.europa.eu/media-centre/news/news-release/frontex-hosts-high-level-visit-by-eu-commission-president-and-belgian-prime-minister-4JIe2L

FRONTEX. (2024b, 26 January). *Significant rise in irregular border crossings in 2023, highest since 2016* https://www.frontex.europa.eu/media-centre/news/news-release/significant-rise-in-irregular-border-crossings-in-2023-highest-since-2016-C0gGpm

Frowd, P. M. (2024). The 'Datafication' of Borders in Global Context: The Role of the International Organization for Migration. *Geopolitics*, 1-19. https://doi.org/10.1080/14650045.2024.2318580

GMDAC. (2020). *Calculating "death rates" in the context of migration journeys: Focus on the Central Mediterranean*. International Organization for Migration. https://publications.iom.int/es/system/files/pdf/mortality-rates.pdf

Heller, C., & Pécoud, A. (2020). Counting migrants' deaths at the border: From civil society counterstatistics to (inter) governmental recuperation. *American Behavioral Scientist*, *64*(4), 480-500. https://doi.org/10.1177/0002764219882

Hinck, R. S., Utterback, R., Kitsch, S. R., & Wenzel, S. (2023). Migration narratives in Northern Triangle, Mexican and US media from 1999 to 2019. *International Migration*, *61*(4), 320-336. https://doi.org/10.1111/imig.13110

İçduygu, A., & Yükseker, D. (2012). Rethinking transit migration in Turkey: reality and re-presentation in the creation of a migratory phenomenon. *Population, Space and Place*, *18*(4), 441-456. https://doi.org/10.1002/psp.633

IOM. (2020). *Missing Migrants - Tracking Deaths Along Migratory Routes Data Collection Guidelines*. https://missingmigrants.iom.int/sites/g/files/tmzbdl601/files/publication/file/MMP%2520data%2520collection%2520guidelines_EN.pdf

IOM. (n.d.-a). *Missing Migrants Project*. International Organization for Migration. https://missingmigrants.iom.int/

IOM. (n.d.-b). *Missing Migrants Project Methodology*. International Organization for Migration,. https://missingmigrants.iom.int/methodology

Jandl, M. (2008). *Methods for estimating stocks and flows of irregular migrants. Chapter 3 of CLANDESTINO Report on methodological issues*. https://migrant-integration.ec.europa.eu/library-document/clandestino-project-report-methodological-issues_en

Jandl, M., Vogel, D., & Iglicka, K. (2008). *Report on methodological issues* (Research Paper, CLANDESTINO Undocumented Migration: Counting the Uncountable, Issue. https://migrant-integration.ec.europa.eu/library-document/clandestino-project-report-methodological-issues_en

Kalkman, J. P. (2021). *Frontex: A literature review* (0020-7985). (International Migration, Issue. https://publons.com/publon/10.1111/imig.12729

Kraler, A., & Ahrens, J. (2023). *Conceptualising migrant irregularity for measurement purposes, MIrreM Working Paper 2/2023*. Zenodo. https://doi.org/10.5281/zenodo.7868237

Kraler, A., & Reichel, D. (2011). Measuring irregular migration and population flows–What available data can tell. *International Migration*, *49*(5), 97-128. https://doi.org/10.1111/j.1468-2435.2011.00699.x

Kraler, A., & Reichel, D. (2022). Migration Statistics. In P. Scholten (Ed.), *Introduction to Migration Studies - An Interactive Guide to the Literatures on Migration and Diversity* (pp. 439-462). Springer.

Kraler, A., Reichel, D., König, A., Baldwin-Edwards, M., & Şimşek, M. (2014). *Final Report – Feasibility Study on the Labour Market Trajectories of Regularised Immigrants within the European Union (REGANE I)*. https://www.icmpd.org/content/download/48327/file/Feasibility%20Study%20on%20the%20Labour%20Market%20Trajectories%20of%20Regularised%20Immigrants%20within%20the%20European%20Union%20(REGANE%20I).pdf

McAuliffe, M., & Ruhs, M. (2017). *World Migration Report 2018*. https://www.iom.int/sites/g/files/tmzbdl486/files/country/docs/china/r5_world_migration_report_2018_en.pdf

McAuliffe, M., & Sawyer, A. (2021). The roles and limitations of data science in understanding international migration flows and human mobility. In M. McAuliffe (Ed.), *Research Handbook on International Migration and Digital Technology* (pp. 42-57). Edward Elgar Publishing.

Misiuna, J., & Pachocka, M. (2018). *The EU and US approaches towards irregular immigration* (2083-8611). (Studia Ekonomiczne, Issue. https://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.cejsh-5451d017-b6d1-4329-b4f9-5cc7b596253c

Mouzourakis, M. (2014). *'Wrong number?' The Use and Misuse of Asylum Data in the European Union*. https://www.ceps.eu/ceps-publications/wrong-number-use-and-misuse-asylum-data-european-union/

O'Reilly, E. (2024, 28 February). Six hundred people drowning but no mayday call. Is this how Europe deters migrants? *The Guardian*. https://www.theguardian.com/commentisfree/2024/feb/28/600-people-drowning-eu-deters-migrants-adriana-tragedy

Pécoud, A. (2020). Death at the border: Revisiting the debate in light of the Euro-Mediterranean migration crisis. *American Behavioral Scientist*, *64*(4), 379-388. https://doi.org/10.1177/00027642198829

Perkowski, N., Stierl, M., & Burridge, A. (2023). The evolution of EUropean border governance through crisis: Frontex and the interplay of protracted and acute crisis narratives. *Environment and Planning D: Society and Space*, *41*(1), 110-129. https://doi.org/10.1177/026377582311524

Perret, S., & Aradau, C. (2023). Drawing data together: Inscriptions, asylum, and scripts of security. *Science, Technology, & Human Values*. https://doi.org/10.1177/0162243923120017

Pettrachin, A., & Hadj Abdou, L. (2024). Beyond evidence-based policymaking? Exploring knowledge formation and source effects in US migration policymaking. *Policy Sciences*, 1-26. https://doi.org/10.1007/s11077-024-09523-y

PICUM. (2022). *FAQ Regularisation and Access to a Secure Residence Status*. https://picum.org/wp-content/uploads/2023/08/FAQ-Regularisation-EN-2.pdf

Raineri, L., & Strazzari, F. (2021). Dissecting the EU response to th e 'migration crisis'. In R. Mac Ginty, S. Pogodda, & O. P. Richmond (Eds.), *The EU and crisis response* (pp. 201-226). Manchester University Press.

Raymer, J., Wiśniowski, A., Forster, J. J., Smith, P. W., & Bijak, J. (2013). Integrated modeling of European migration. *Journal of the American Statistical Association*, *108*(503), 801-819. https://doi.org/10.1080/01621459.2013.789435

Rodriguez Sanchez, A., & Tjaden, J. (2023). *Estimating irregular migration – a review of traditional and innovative methods. MIrreM Working Paper No.4.* . https://doi.org/10.5281/zenodo.8380854

Santamaria, C., & Vespe, M. (2018). *Towards an EU Policy on Migration Data*. P. O. o. t. E. Union. https://publications.jrc.ec.europa.eu/repository/handle/JRC112864

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*, 1-32. https://doi.org/10.1177/01979183231222169

Seelke, C. R., Meyer, P. J., & Senger, S. B. (2023). *U.S. Efforts to Manage Western Hemisphere Migration Flows*. https://crsreports.congress.gov/product/pdf/IF/IF12538

Singleton, A. (2020). *Legislative update for EU migration and asylum statistics – work in progress*. University of Bristol. https://migration.bristol.ac.uk/2020/07/28/legislative-update-for-eu-migration-and-asylum-statistics-work-in-progress/

Slaven, M., & Boswell, C. (2019). Why symbolise control? Irregular migration to the UK and symbolic policy-making in the 1960s. *Journal of Ethnic and Migration Studies*, *45*(9), 1477-1495. https://doi.org/10.1080/1369183X.2018.1459522

Slootjes, J., Kokkelmans, R., & Sohst, R. (2023). *Mapping Stakeholders' Needs and Usage of Irregular Migration Data (Version 1)*. https://doi.org/10.5281/zenodo.7589494

U.S. Customs and Border Protection (CBP). (n.d.). *Nationwide Encounters | U.S. Customs and Border Protection*. https://www.cbp.gov/newsroom/stats/nationwide-encounters

Ustek-Spilda, F. (2020). Statisticians as back-office policy-makers: Counting asylum-seekers and refugees in Europe. *Science, Technology, & Human Values*, *45*(2), 289-316. https://doi.org/10.1177/01622439198820

Vespe, M., Natale, F., & Pappalardo, L. (2017). Data sets on irregular migration and irregular migrants in the European Union. *MIgration Policy Practice*, *VII*(2). https://publications.iom.int/books/migration-policy-practice-vol-vii-number-2-april-september-2017

Vogel, D., Jandl, M., Kraler, A., & Vogel, D. (2008). *Report on methodological issues* (Report prepared for the research project CLANDESTINO Undocumented Migration: Counting the Uncountable. Data and Trends across Europe funded under the 6th Framework Programme of the European Union, Issue. https://migrant-integration.ec.europa.eu/library-document/clandestino-project-report-methodological-issues_en

Vollmer, B. A. (2011). Policy Discourses on Irregular Migration in the EU-'Number Games' and 'Political Games'. *European Journal of Migration and Law*, *13*(3), 317-339. https://doi.org/10.1163/157181611X587874

Wærp, E. (2024, 24 January). Frontex's Normalization of Crisis at the EU Borders. *NCCR On the Move*. https://nccr-onthemove.ch/blog/frontexs-normalization-of-crisis-at-the-eu-borders/

Warren, R. (2021). In 2019, the US undocumented population continued a decade-long decline and the foreign-born population neared zero growth. *Journal on Migration and Human Security*, *9*(1), 31-43. https://doi.org/10.1177/2331502421993

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T.,

Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Williams, K., & Mountz, A. (2018). Between enforcement and precarity: externalization and migrant deaths at sea. *International Migration*, *56*(5), 74-89. https://doi.org/10.1111/imig.12439

Wiśniowski, A. (2021). Migration forecasting using new technology and methods. In M. McAuliffe (Ed.), *Research Handbook on International Migration and Digital Technology* (pp. 376-392). Edward Elgar Publishing.

Yildiz, D., & Abel, G. (2021). Migration stocks and flows: data concepts, availability and comparability. In M. McAuliffe (Ed.), *Research handbook on international migration and digital technology* (pp. 29-41). Edward Elgar Publishing.

ANNEXES

Annex 1. The EU irregular migration data landscape

Within the EU, the enforcement of migration legislation covers two policy areas: the control of the EU's external borders and the management of unauthorised stays by non-EU citizens found in an EU Member State. Regulation (EC) No 862/2007[41] sets out the rules for collecting and compiling statistics on migration (emigration and immigration), international protection (asylum), regular and irregular migration and returns by EU and EFTA countries. On this basis, Member States are required to submit data on the number of non-EU citizens: (1) refused entry at the EU's external borders; (2) apprehended for being illegally present in the EU; and (3) removed from the EU as a result of their presence being unauthorised all of which are indicators of geographic flows. Referred to collectively as enforcement of migration statistics (EIL) (please refer to Annex 2 and Annex 3 for more information), this is compiled by Eurostat, the statistical office of the EU. To provide comparative statistics, it works in partnership with the 'National Statistical Institutes' (NSIs) and 'Other National Authorities' (ONAs) in each EU Member State, which altogether comprise the European Statistical System (ESS). Countries which are part of the European Free Trade Association (EFTA) also participate in the development, production, and dissemination of statistics (Eurostat, n.d.-g).[42]

A report by of the European Commission's Joint Research Centre (JRC) found that while the migration data landscape in the EU is better compared to many other world regions, there were gaps to address to increase data adequacy for policymaking purposes (Santamaria & Vespe, 2018, pp. 3-4). These gaps relate to the following:

- Gaps in existing data – with regard to timeliness of transmission and receipt, quality, and level of disaggregation available
- Dissemination issues – unclear methodologies and underlying assumptions, and reduced visibility of key data
- Data not collected – topics for which data is not routinely collected for example, information on children and intra-EU movements of TCNs

[41] Regulation (EC) 862/2007 of the European Parliament and of the Council of 11 July 2007 on Community statistics on migration and international protection and repealing Council Regulation (EEC) No 311/76 on the compilation of statistics on foreign workers.

[42] Generally, data are published in the Eurostat database approximately within 3 working days after their reception, if they do not contain any errors. In some cases (large revisions, further validation, technical reasons, etc.) data may be disseminated later than 3 working days. Data are revised on a **continuous basis** according to the most recently updated data provided by the countries.

- Useful data that is currently inaccessible – existing data that are not accessible to EU policymakers due to legal or proprietary reasons

In 2020, while recognising that irregular migration remains a phenomenon difficult to quantify, an amendment was introduced in Regulation (EU) 2020/851[43] to include additional statistics to be collected by Eurostat and add rules for introducing pilot studies to test the feasibility of new data collections or disaggregations within the scope of Regulation (EC) No 862/2007.[44] Thus, from 2021, EIL statistics collected by Eurostat also include the following:

- Statistics on TCNs found to be illegally present by 'place of apprehension' and by the 'grounds of apprehension'
- Quarterly rather than annual statistics on returns, and mandatory breakdowns by type of return, type of assistance received and destination country
- Statistics on unaccompanied minors ordered to leave and on unaccompanied minors who returned following an order to leave.

In addition to EIL statistics, Eurostat is also collecting data on asylum migration such as:

- Number of asylum applicants
- Decisions on applications and resettlement
- Outgoing and incoming Dublin requests and transfers

In terms of irregular migration status-related flows, what is relevant from asylum statistics would be first instance or final **negative decisions** as well as **decisions withdrawing status** granted at first instance or as final decision. Both indicators capture **status-related flows** (i.e., inflows). The collection of asylum statistics on the EU-level has also evolved over time with the most significant developments as follows (Eurostat, n.d.-a):

- Period January 1985 – December 2007: data collected on the voluntary basis
- Period January 2008 – December 2014: data by Regulation (EC) No 862/2007
- Period January 2014 – December 2020: implementation of the new 2013 Asylum Guidelines
- Period from January 2021: implementation of the new 2021 Asylum Guidelines, data collected on reason for the decision withdrawing status (revocation, ending, refusal to renew, unknown)

[43] Regulation (EU) 2020/851 of the European Parliament and of the Council of 18 June 2020 amending Regulation (EC) No 862/2007 on Community statistics on migration and international protection.
[44] For more detailed analysis, please see: Singleton, A. (2020). *Legislative update for EU migration and asylum statistics – work in progress*. University of Bristol. https://migration.bristol.ac.uk/2020/07/28/legislative-update-for-eu-migration-and-asylum-statistics-work-in-progress/.

Annex 2. Main EIL variables used by Eurostat

Main EIL variables

Variable	Sub-category	Breakdown	Frequency, Length of time series, Availability
Mandatory data Collection			
Third-country nationals refused entry at the external borders		Citizenship, Type of border, Grounds for refusal	Annual (2008 - ...); M+3
Third-country nationals found to be illegally present		Age, Sex, Citizenship, Grounds for apprehension, Place of apprehension (*)	Annual (2008 - ...); M+3
Third-country nationals ordered to leave	Unaccompanied minors	Age, Sex, Citizenship	Annual (2008 - 2020); M+3 Quarterly (2021 - ...); M+2
Third-country nationals returned following an order to leave		Age, Sex, Citizenship, Destination	Annual (2008 - 2020); M+3
Third-country nationals returned following an order to leave	Unaccompanied minors	Age, Sex, Citizenship, Type of return, Country of destination	Quarterly (2021 - ...); M+2
Voluntary data collection			
Third-country nationals who have left the territory by type of return and citizenship		Citizenship, Type of return	Annual (2008 - 2020); M+3
Third-country nationals who have left the territory by type of assistance received		Citizenship, Type of assistance received	Annual (2008 - 2020); M+3
Third-country nationals who have left the territory to a third country by type of agreement procedure		Citizenship, Type of agreement procedure	Annual (2008 - 2020); M+3
Third-country nationals who have left the territory to a third country by destination country		Citizenship, Country of destination	Annual (2008-...); M+3

(*) Breakdown by grounds for and place of apprehension will be collected for the first time in 2022 for reference year 2021.

eurostat

Source: Eurostat (n.d.-c)

Annex 3. Classification and breakdown used in EIL statistics

Presentation of the classification and breakdown used in EIL statistics

Classification/Breakdown	
Age	Under 14 years-old From 14 to 17 years-old From 18 to 34 years-old 35 years-old or over Unknown
Sex	Female; Male; Unknown
Citizenship	Classification of citizenship based on the ISO-3166 including stateless and unknown nationality
Type of border	Land border Air border Sea border
Ground for refusal	No valid travel document(s) False/counterfeit/forged travel document No valid visa or residence permit False/counterfeit/forged visa or residence permit No appropriate documentation justifying the purpose and conditions of stay Already stayed for 90 days in the preceding 180-day period on the territory of the Member States of the EU No sufficient means of subsistence for the period and form of stay, or any means to return to the country of origin or transit Person for whom an alert has been issued for the purposes of refusing entry Person considered to be a threat to public policy, internal security, public health or the international relations of the Member States of the EU
Grounds for apprehension	Illegal entry Overstay Other reason (including unknown reason)
Place of apprehension	External border area o At border crossing point (Sea + Land + Air border) (e.g. upon exit) o Between border crossing points at Sea (e.g. upon entry or exit or unknown direction) o Between border crossing points at Land (e.g. upon entry or exit or unknown direction) Inland area
Type of return and assistance	Assisted voluntary returns Non-assisted voluntary returns Assisted forced returns
Country of destination (returns)	Another country from 'EU + EFTA zone' Third country o Country of citizenship o Transit country (only related to returns to third country) o Other third country Unknown
Type of return	Voluntary return Enforced return Other concluded return
Type of assistance received	Assisted return Non-assisted return Unknown
Type of agreement procedure	Returned under EU readmission agreements (EURA) Returned under other readmission agreement(s) Returned without existing a readmission agreement Unknown

eurostat

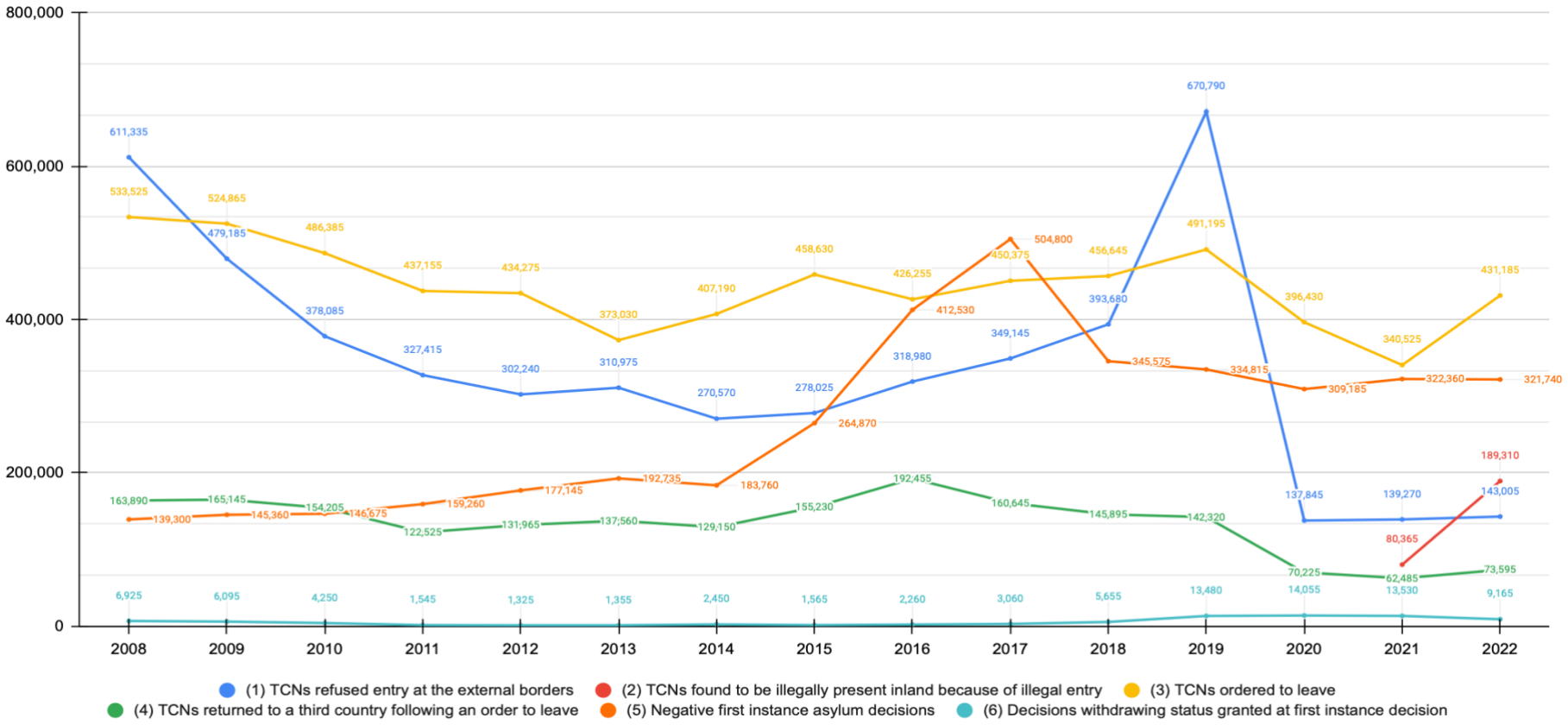
Source: Eurostat (n.d.-c)

Annex 4. Definition of selected irregular flow indicators

Indicator	Eurostat definition and notes on data
1) **TCNs refused entry at the external border**	TCNs formally refused permission to enter the territory of a Member State. The external border is defined as in the Schengen Borders Code. For countries which are not in the Schengen area, the external border is the same as the international border.
2) **TCNs found to be illegally present inland because of illegal entry**	TCNs who are detected by Member States' authorities and have been determined to be illegally present under national laws relating to immigration. This category relates to persons who have been found to have entered illegally for example, by avoiding immigration controls or by employing a fraudulent document. Only persons who are apprehended or otherwise come to the attention of national immigration authorities are recorded in these statistics. These are not intended to be a measure of the total number of persons who are present in the country on an unauthorised basis.
3) **TCNs ordered to leave**	TCNs found to be illegally present who are subject to an administrative or judicial decision or act stating that their stay is illegal and imposing an obligation to leave the territory of the Member State. These statistics do not include persons who are transferred from one Member State to another under the mechanism established by the Dublin Regulation.
4) **TCNs returned to a third country**	TCNs who have in fact left the territory of the Member State, following an administrative or judicial decision or act stating that their stay is illegal and imposing an obligation to leave the territory. On a voluntary basis Member States provide Eurostat with a subcategory which relates to third country nationals returned to a third country only (either the country of origin or another third country). Persons who left the territory within the year may have been subject to an obligation to leave in a previous year. As such, the number of persons who actually left the territory may be greater than those who were subject to an obligation to leave in the same year.
5) **Rejected applicants for international protection**	Persons covered by first instance decisions rejecting applications for international protection, such as decisions considering applications as inadmissible or as unfounded and decisions under priority and accelerated procedures, taken by administrative or judicial bodies during the reference period.
6) **Withdrawals of first instance asylum decisions**	Persons covered by first instance decisions withdrawing refugee status, subsidiary protection status, temporary protection or authorisation to stay for humanitarian reasons under national law concerning international protection status, taken by administrative or judicial bodies during the reference period. Reasons for withdrawals include revocation, ending, or refusal to renew.

Source: (Eurostat, n.d.-b, n.d.-e)

Annex 5. Trends of selected irregular flow indicators, total annual aggregated data from Eurostat for EU27 (2008-2022)



Notes: Data from Croatia is available from 2013 onwards. Please refer to Figure 1 to see the trends for the same indicators with data from the UK (2008-2019) included.

Annex 6. FRONTEX and the production of irregular migration data

Through a series of revisions in 2007, 2011, 2016, and 2019, FRONTEX underwent a gradual shift from its initial coordinating role to adopting a more enforcement-oriented approach (D'Auria, 2023; Sachseder et al., 2022). This transformation was accelerated by the migration crisis from 2014 to 2016. Following the enactment of EU Regulation (EU) 2016/1624, FRONTEX underwent a name change to the "European Border and Coast Guard Agency", encompassing both the Agency itself and national authorities responsible for border management. A subsequent development (Regulation (EU) 2019/1896) included the establishment of the European Border and Coast Guard Standing Corps, representing the EU's inaugural uniformed law enforcement service empowered to assist Member States in safeguarding external borders. This initiative implied an unprecedented personnel influx, predicting up to 10,000 operational staff available to the Agency by 2027. Simultaneously, FRONTEX's budget experienced substantial growth, surging from EUR 333 million in 2019 to EUR 364 million in 2020 and further to EUR 754 million in 2022. This financial trajectory starkly contrasts with the EUR 6 million budget in 2005 (D'Auria, 2023).

However, the credibility of FRONTEX's activities has recently come under intense scrutiny. The 2021 FRONTEX Scrutiny Working Group report, while not finding conclusive evidence of direct pushbacks or collective expulsions, raised alarms about fundamental rights violations. Delays in recruiting fundamental rights monitors and FRONTEX's lack of responsiveness to allegations further fuelled the debate. The subsequent October 2022 European Anti-Fraud Office (OLAF) report overturned earlier findings and confirmed FRONTEX's involvement in illegal pushbacks. This included the revelation of staff disregarding fundamental rights obligations. FRONTEX's involvement in enabling the interception of migrant boats by the Libyan Coast Guard, as accused by Human Rights Watch and Border Forensics in December 2022, adds another layer of complexity to the agency's potential complicity in actions that may lead to systematic abuse when migrants are returned to Libya (D'Auria, 2023).

Concomitantly, there was an important shift in FRONTEX's data collection and analysis methodologies. Historically, FRONTEX had functioned as a coordinating agency with a predominant focus on risk analysis, employing qualitative methods and prioritising expert judgments. However, the 2016 Regulation introduced the element of FRONTEX's Vulnerability Assessments, for which the agency started adopting a more quantitative approach. The methodology employed for vulnerability assessments contributes to emphasise quantified threat levels based on numerical indicators, such as the number of illegal border crossings or entries with fraudulent documents detected. The choice to adopt a more quantitative methodology in vulnerability assessments was not merely a technical shift but also implied political considerations. The perceived objectivity and concreteness of numerical claims were deemed advantageous for earning trust and acceptance from policymakers, to enhance the legitimacy of FRONTEX's operations (Fjørtoft, 2022).

ABOUT THE MAIN AUTHORS

Lalaine Siruno is a researcher at the Institute for Societal Policy and Innovation Research (INSPIRE), Maastricht University; and the Governance and Public Policy of Innovation Unit, United Nations University-MERIT, the Netherlands.

Arjen Leerkes is a Professorial Fellow at INSPIRE, Maastricht University; an Associate Professor of Sociology at the Erasmus School of Social and Behavioural Sciences, Erasmus University Rotterdam; and a research fellow at the Research and Documentation Centre (WODC) of the Dutch Ministry of Justice and Security.

COPYRIGHT CLAUSE

THE MIRREM CONSORTIUM

University for Continuing Education Krems (Coordinator)
European University Institute
University of Osnabrück
University of Maastricht
University of Turku
Complutense University Madrid
Hellenic Foundation for European and Foreign Policy (ELIAMEP)
University of Milan
University of Potsdam
Platform for International Cooperation on Undocumented Migration (PICUM)
International Centre for Migration Policy Development (ICMPD)
Migration Policy Institute Europe (MPI-E)
University of Warsaw
Vrije Universiteit Brussel
Instituto Universitário de Lisboa

Associated Partners:

Toronto Metropolitan University
University of Leicester
University of Oxford