# Chapter 3

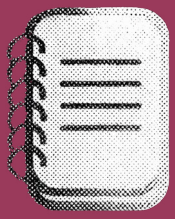
## Ethics and data on irregular migration

*Jill Ahrens*

To cite: Ahrens, J. (2025). Ethics and data on irregular migration. In D. Kierans and A. Kraler (eds), H*andbook on Irregular Migration Data. Concepts, Methods and Practices.* Krems: University of Krems Press. https://doi.org/10.48341/g31s-vq79-03

*Keywords:* Research ethics, minimise harm, , transparency, accountability, data protection, irregular migration, informed consent, undocumented migrant children, MIMAP survey, data protection sensitive sampling, Pseudonymisation, record linkage, Austrian Micro-Data Centre

# Ethics and data on irregular migration



## Key points

- Embed ethics and rights-based approaches into migration research: Go beyond compliance with the European Union's General Data Protection Regulation (GDPR) and the Artificial Intelligence Act, as well as other technical guidelines to minimise harms such as surveillance, discrimination and misuse of data concerning vulnerable groups
- Ensure transparency and accountability: Acknowledge uncertainty, avoid uncritical use of categories, and prevent data misuse in shaping restrictive policies.
- Build trust and inclusive governance: Strengthen safeguards, clear communication and migrant engagement to support ethical and effective data collection and use.

## Introduction

Ethical and data protection principles need to be integral to collecting, analysing and sharing data on irregular migration. This research area raises distinct ethics challenges due to migrants' vulnerability to migration enforcement, surveillance, and social exclusion. In politically charged contexts such as irregular migration, ethics requires going beyond mere compliance with laws and guidelines; proactive research ethics and integrity ensure responsible, substantive transparency, and accountable data practices while protecting participants' rights and broader social interests.

Here, transparency means not just providing access to information, but actively explaining methodological limitations, potential errors, assumptions, and the ways that data can – and cannot – be interpreted. At their core, research ethics aim to safeguard individuals through principles such as *autonomy, beneficence* (or 'do no harm' and maximise benefits), and *justice*. When working with irregular migration data, ensuring autonomy means ensuring negotiated and informed consent that takes fears about abuse and deportation seriously; beneficence requires careful assessment of the risks such as profiling or stigmatisation; and working towards justice demands recognising and mitigating power imbalances between researchers, policymakers, and migrant communities.

Research integrity builds on these ethical foundations through four principles which *The European Code of Conduct for Research Integrity*

(ALLEA, 2023, p.4) defines as:

- Reliability
- Honesty
- Respect
- Accountability.

Embedding these principles in research practice goes beyond the individual's responsibility, to include institutions and scholarly communities, and thus fostering a culture that prevents misconduct, and upholds public trust and promotes reflexive and risk-aware decision-making.

Professional standards in politically charged fields such as irregular migration include ensuring data quality, designing proportionate and privacy-conscious data collection, communicating findings with care, acknowledging others' contributions, and anticipating real-world consequences of research. In this context, transparent communication requires researchers to explicitly acknowledge uncertainty, document assumptions, and explain how findings should and should not be interpreted, rather than relying on vague claims of openness. Such an approach treats ethics not as a bureaucratic hurdle or checklist, but as an ongoing, reflexive practice shaping every stage of the data lifecycle. This is essential given the heightened risks of misuse, discrimination, and rights violations faced by irregular migrants, and the responsibility to avoid reinforcing existing inequalities through research (see Box 3.1, for the example of irregular migrant children).

**Box 3.1: Making undocumented migrant children visible: A balanced approach to data collection, analysis and use**

*Marzia Rango, Naomi Lindt, Sebastian Palmas and Danzhen You*

Collecting, producing and disseminating data and statistics on children who migrate without proper documentation or authorisation requires careful consideration. The lack of reliable data on migration that can be disaggregated by dimensions including age, sex and migratory status often renders this population statistically 'invisible', complicating efforts to uphold their most basic rights. However, if the generation of this evidence is not grounded in a child-sensitive, rights-based approach, undocumented migrant children can be potentially exposed to further rights violations, such as detention, deportation, family separation and human trafficking.

The well-being of undocumented migrant children is often undermined by their lack of legal status, particularly if they are unaccompanied or separated. The need to shed light on their deprivations and the risks they face, while also identifying and mitigating potential risks of harm that result from data collection, production and use must be thoughtfully balanced and informed by best practices (Sherr et al., 2025).

As enshrined in the UN Convention on the Rights of the Child, the best interests of children must be prioritised in all data work. To guarantee that the process of evidence generation for children is truly ethical, some core principles need to be adhered to at all times: Benefit, "Do no harm," non-discrimination, respect, justice or fairness, integrity and accountability (Rahman and Keseru, 2021). In practice, this means designing and adopting an approach centred on children's rights, which ensures that children's views are heard and their dignity respected, all while maintaining strict confidentiality and data protection protocols. Data collectors must carefully consider which data are needed to adequately represent an undocumented migrant child's circumstances and how the data will be collected, stored and used. Building trust and providing a safe environment for children to share their experiences are also crucial (Graham et al., 2013).

In an attempt to operationalise these principles, UNICEF, in collaboration with The GovLab at New York University, launched the Responsible Data for Children (RD4C) initiative.[1] This framework provides a comprehensive set of principles to guide data handling throughout its entire lifecycle –

from collection and storage to analysis and use:

- Participatory: Involving and informing children, their caregivers, and communities in the data process.
- People-centric: Prioritizing the needs and expectations of children, their caregivers, and their communities.
- Prevention of harms: Assessing and mitigating risks at every stage of the data lifecycle.
- Professionally accountable: Establishing institutional processes and roles to ensure responsible data practices are implemented.
- Purpose-driven: Ensuring data is collected with a clear objective that benefits children.
- Protective of children's rights: Upholding the rights of the child throughout the data process.

Additional resources relevant to the ethical collection and production of data and on children in vulnerable situations – such as undocumented migrant children – include the UNHCR-UNICEF Guidance Note on Responsible Disaggregation of Data on Refugee Children (UNICEF and UNHCR, 2023), UNICEF e-course on Ethics in Evidence Generation,[2] the Compendium on Ethical Research Involving Children (Graham et al., 2013) and the report Researching Sensitive Topics Involving Children (Sherr et al., 2025). A series of reports also address the ethical dimension of the use of new technologies and novel data sources for evidence generation for children (Berman and Albright, 2017; Berman et al., 2018a; Berman et al., 2018b; Rahman and Keseru, 2021).

The International Data Alliance for Children on the Move (IDAC) was launched in 2020 as a direct response to the need for better data on children on the move, particularly those who are the most vulnerable. More about IDAC, its mandate, events and resources are available at dataforchildrenonthemove.org.

**References**

Berman, G. & Albright, K. (2017). Rights and Ethics in a Big Data World.UNICEF Office of Research – Innocenti. https://doi.org/10.18356/7d33bc21-en

Berman, G., de la Rosa, S. & Accone, T. (2018a). Ethical Considerations When Using Geospatial Technologies for Evidence Generation. UNICEF Office of Research – Innocenti. https://doi.org/10.18356/60c0e27b-en

Berman, G., Powell, J. & Garcia Herranz, M. (2018b). Ethical Considerations in Using Social Media for Evidence Generation. UNICEF Office of Research – Innocenti. https://doi.org/10.18356/5bf16c88-en

Graham, A., Powell, M., Taylor, N., Anderson, D. & Fitzgerald, R. (2013). Ethical Research Involving Children. Florence: UNICEF Office of Research - Innocenti. https://www.unicef.org/innocenti/reports/ethical-research-involving-children

Rahman, Z. & Keseru, J. (2021) Predictive Analytics for Children: An assessment of ethical considerations, risks and benefits. UNICEF Office of Research – Innocenti. https://www.unicef.org/innocenti/media/5161/file/UNICEF-Predictive-Analytics-Working-Paper-2021.pdf

1 See https://rd4c.org/.

2 See https://agora.unicef.org/course/info.php?id=33813.

Sherr, L., Mathews, S., Hussein, M., Carter, K.L., & Bianchera, E. (2025). Researching Sensitive Topics Involving Children: Managing Risk, Distress and Trauma in Research Design and Data Collection. UNICEF Office of Research – Innocenti.
https://www.unicef.org/innocenti/reports/researching-sensitive-topics-involving-children

UNICEF and UNHCR. (2023). Responsible Disaggregation of Data on Refugee Children. New York and Copenhagen.
https://data.unicef.org/resources/responsible-disaggregation-of-data-on-refugee-children/

## Ethics risks in data collection and use

### Intrusive practices and group privacy

Collecting data about people in irregular situations carries significant risks of harm, especially through intrusive or disproportionate data processing. Under the General Data Protection Regulation (GDPR), personal data includes any information that can identify a person, requiring careful control over collection, storage, use, sharing, and deletion (Regulation (EU) 2016/679). Strict enforcement of GDPR compliance means that: personal data will not be shared with third parties, informed consent is mandatory, and any sharing within the project (e.g. for associated researchers or external colleagues) will follow carefully regulated agreements.

Yet, compliance with formal data protection law is only the starting point. Ethical research must also consider 'group privacy' (Floridi et al., 2018). Even anonymised and aggregated big data can enable profiling, reinforcing existing surveillance and discrimination. For migrants in an irregular situation, who are already subject to heightened scrutiny, combining or linking datasets can expose group-level patterns (e.g. concentrations in certain locations or demographic profiles) that risk further stigmatisation or enforcement action and that are also not covered in emergent regulations aimed at AI use (e.g. the EU AI act).

To mitigate these risks, it is advisable to adopt a 'dynamic approach to anonymisation' (Reed-Berendt et al., 2022). Rather than treating anonymisation as a one-off technical step, this approach recognises that identifiability can change over time or through the linking of datasets. Researchers must therefore remain vigilant, proactively assessing and reducing the risk of harmful inferences that can be made about vulnerable groups. This demands careful design of data access policies, technical safeguards, and ethical review processes, ensuring that individual and collective rights are protected at all stages.

### Data sources and uncritical categories

A further ethics risk is that existing migration data sources and infrastructures have become 'invisible' or are taken for granted. Taylor and Meissner (2024) encourage researcher to uncover "a new form of metadata", namely that of data infrastructures, so as to understand who designed them, with what interests and with what assumptions about migration. Ethical practice requires resisting the role of passive data consumers, and interrogating the powers, politics and purposes built into the data systems that frame (ir)regular migration.

We should also avoid uncritically reproducing politically charged categories. Research on irregular migration often relies on legal-administrative labels that obscure lived experiences and intersectional inequalities. Such labels risk treating 'irregular' status as a dominant or 'master status' that overshadows other factors such as gender, ethnicity, racialisation, class, etc. This reification of legal categories can have real-world consequences, including legitimising restrictive policies and contributing to public fears or moral panics. As Bakewell (2008) warns for refugee studies, there is a danger of "co-producing" the problem we claim to study by adopting policy actors' assumptions uncritically.

Critical reflection on categorisation is therefore essential. Calls to 'de-migranticise' migration research (Dahinden, 2016) or adopt 'methodological denationalism' (Anderson, 2019) highlight the need to challenge taken-for-granted national or legal frameworks. Scheel and Tazzioli (2022) extend this critique by arguing that migrants are not a fixed group, but are continually shaped by the policies and practices of bordering and 'migrantisation'. Mohan et al. (2023) similarly urge researchers to reframe irregularity, foregrounding how migrant status is produced through institutional and political processes. This means treating migrant status as one variable among many, paying close attention to the processes of 'irregularisation' and the intersectional experiences of migrants. Conceptual attentiveness is not merely an academic concern but an ethical responsibility to avoid reinforcing the very inequalities and exclusions that research seeks to shed light on (see also chapter 2).

## Misuse of findings

Data and findings are susceptible to misinterpretation or misuse that may inadvertently justify restrictive policies or surveillance strategies. Estimates of irregular migrant stocks and flows, or evaluations of regularisation schemes can shape public debates and policy decisions – sometimes in harmful ways. Findings can be misinterpreted, intentionally distorted, or weaponised to justify restrictive policies, surveillance technologies, or immigration enforcement measures that undermine migrants' rights.

Researchers have an ethical duty to anticipate these risks, to mitigate them and to reflect continuously on what constitutes responsible use of data (Cyrus 2023; Hendow et al. 2024). Reflexive research requires scrutiny not just of categories or analytical choices, but of the ways data might be applied in policy and public discourse. Estimates produced for analytical purposes may inadvertently aid the development of surveillance tools or influence how status determination procedures are designed in ways that limit social inclusion.

To mitigate these risks, researchers should adopt transparent and interpretively responsible communication, documenting uncertainty and methodological assumptions and providing metadata on data reliability and to always clarify what the data can and cannot show. When sharing estimates of irregular migrant stocks, it is more appropriate to publish ranges rather than point estimates alone, explicitly explaining underlying assumptions and limitations, and avoiding the impression of false precision. Ethical communication also involves being sensitive to language and framing, recognising that labels can stigmatise, and data can be appropriated to serve various purposes.

## Topic bias and data gaps

Existing data on irregular migration often overlooks complex trajectories, status loss, and duration of stay in an irregular situation, creating systematic biases that must be transparently acknowledged and addressed. Bias arises not only from analytical choices but also from the fragmented and selective nature of available datasets, which are shaped by institutional priorities.

Hendow et al. (2024) argue that enthusiasm for new data sources must be tempered: data are not a panacea, and policymakers need clarity on what new data can and cannot reveal. They highlight persisting data 'blind spots' (e.g. patterns of overstaying, secondary movements, unverifiable returns, etc.), which hamper efforts to produce comprehensive EU-level estimates on the irregular migrant population. They call for data collection to be proportionate to its aims and in line with EU law, and more efforts to harmonise flow indicators and reduce double-counting. This needs to be supported by regular data exchanges and related efforts to anonymise data to address legitimate concerns over data protection and privacy.

Several studies emphasise how data collection and data use can affect the (in)visibility of irregular migrants. Jasso et al. (2008) combined administrative and survey microdata to show that administrative sources understate prior irregular experiences (e.g. entry without inspection, overstaying, and unauthorised work, etc.) and revealed important differences across origin countries, migrant categories, and within the wider population in an irregular situation (cf. chapter 2). Meanwhile, Descamps (2024), uses retrospective biographical survey data from the Trajectoires et Origines 2 survey to identify and qualify measurement biases (i.e. social desirability, recall errors, and non-proactivity) in migration status trajectory reporting, concluding that

such biases are relatively minor. She argues that migration status should more often be included in surveys, because this would enrich theoretical understandings of migrants' experiences and inform policy development. However, this assumes that migrants know their status and would report it openly and accurately.

This highlights the importance of reflexivity in interpretation, recognising whose experiences are represented and whose are overlooked. Such critical awareness should guide transparent communication to policymakers in order to prevent decisions based on incomplete or skewed evidence, which may further marginalise already vulnerable populations.

## Safeguarding rights and responsible data use

### Trust and mistrust

Trust is a foundational element in the collection and use of migration data. Descamps and Boswell (2018) show how institutional mistrust (e.g. fuelled by rivalries, lack of transparency, conflicting incentives, etc.) undermines coordination and data sharing. Mistrust between agencies can lead to fragmented systems, duplicated efforts, and ultimately weaker evidence for policymaking.

At the same time, migrants themselves may deeply distrust data collection efforts. Fear of surveillance, deportation, or misuse of personal information reduces willingness to participate or share accurate data (Kraler et al., 2015). This affects not only research quality but also the credibility of policy responses. However, when trust is established through robust safeguards and ethical practice, data collection and use can serve positive purposes. Responsible data use can inform the design of social inclusion programmes, improve service provision, and support policies that protect migrant's rights.

Researchers need to recognise that trust cannot be demanded but must be earned through ethical practice, including respecting autonomy, ensuring confidentiality, negotiating consent to participate, and demonstrating commitment to protecting research participants from harm. These principles must guide both data collection and the wider institutional relationships on which migration data systems depend (see Box 3.2, for an example).

**Box 3.2: Addressing ethical challenges in surveying irregular migrants – The MIMAP survey on the im-/mobility of rejected asylum seekers**

*Randy Stache*

When no sampling frame exists (e.g. when studying irregular migrants unknown to the authorities) or when particularly sensitive topics are being explored, conventional survey methods quickly reach their limits. Irregular migrants are hard-to-reach and hard-to-survey: The group is blurry and elusive (hard to identify, highly mobile with mistrust against authorities and researchers). The group also is socially and legally marginalised, vulnerable and typically lacks prior engagement with empirical research. Many are familiar with interviews only in the context of authorities, such as police or asylum proceedings. These conditions raise ethical challenges, including data protection, informed consent, and the positionality of researchers. In consequence, innovative and adaptive methodological approaches are needed.

One example for such an approach is the MIMAP project ("Feasibility study on the im-/mobility of rejected asylum seekers"). Conducted between 2022 and 2025 by the Research Centre of the Federal Office for Migration and Refugees in Germany, a part of the project focused on irregular migrants from Anglophone West Africa who had undergone an asylum procedure in Germany. It employed an innovative mixed-methods design, combining quantitative survey research with in-depth ethnographic fieldwork. Ten rejected asylum seekers were repeatedly interviewed and accompanied in their everyday lives. This ethnographic engagement facilitated trust-building and enabled the identification of key community individuals who acted as gatekeepers for the quantitative study. The survey applied Respondent-Driven Sampling (RDS), implemented via a custom-designed mobile application. The app hosted the survey, ensured full anonymity by collecting no personal data, and enabled participants to digitally refer the survey to up to three peers. Participants received a digital €10 shopping voucher both for completing the survey and for each successful referral.

To explore the sensitive issue of mobility aspirations (staying, returning, or migrating onward) the survey incorporated a factorial survey. Participants evaluated four hypothetical profiles of individuals with a 'tolerated' status, whose characteristics (e.g., length of stay: 1, 4, or 10 years) were experimentally varied. Respondents were asked to recommend whether each fictional individual should stay in Germany, return to the country of origin, or migrate to another country. The experimental variation enabled the identification of factors that shape im-/mobility aspirations. In line with the contextualizing qualitative interviews, the quantitative findings show that employment status, conditions in the country of origin, and the location of own children strongly influence (im) mobility aspirations. In contrast, migration enforcement policies such as deportation pressures and return assistance play minor roles (Stache et al., 2025).

Combining qualitative interviews and ethnographic trust-building with a respondent-driven sampling featuring an anonymous, app-based survey and a survey experiment, enabled the systematic investigation of sensitive topics among a highly inaccessible population – while maintaining ethical rigor and contextual depth.

**References:**

Peitz, L. and Stache, R. and Johnson, L. (2024) How to Survey Hard-to-Reach Populations: A Practical Guide to App-Based Respondent-Driven Sampling. Robert Schuman Centre for Advanced Studies Research Paper No. 2024/23, http://dx.doi.org/10.2139/ssrn.4901241

Stache, R., Johnson, L., Peitz, L., & Carwehl, A.-K. (2025). Bleibe- und Rückkehrabsichten von Geduldeten: Erkenntnisse aus einem Umfrageexperiment mit westafrikanischen Schutzsuchenden. (BAMF-Kurzanalyse, 2-2025). Nürnberg: BAMF. https://doi.org/10.48570/bamf.fz.ka.02/2025.d.2025.bleibrueckabsicht.1.0

## Transparency, contestability and responsibility

Quantifying irregular migration can lend findings a veneer of objectivity and authority that masks their contingent, uncertain nature. Numbers often carry persuasive power in policy debates, but when poorly communicated or misinterpreted estimates can mislead decision-makers or the public.

Ethical responsibility demands that researchers clearly communicate the limits and assumptions of

their methods. Transparency here is substantive: it requires explaining potential sources of error, methodological assumptions, and the ways findings can and cannot be interpreted. For algorithmic methods, transparency can help other experts (and it is important to acknowledge this facet) to contest the assumptions and biases embedded into computational analysis. By doing so, policymaker and researchers help ensure that data supports informed, balanced policy decisions rather than fuelling sensationalism or punitive responses.

## Considerations for data linkage and anonymisation strategies

Data protection law, especially the GDPR, imposes clear limits on how personal data may be collected, used, and shared. While these rules are crucial for protecting individual rights, they can also pose practical challenges for research, particularly in linking datasets across sources or countries.

It is necessary to respond to these challenges through careful anonymisation strategies. Pseudonymisation of individuals' identities is a standard practice, with participants given choices about the level of disclosure they are comfortable with. Researchers can use coded protocols for interviews, workshops, and surveys to minimise identifiability. Anonymisation should not be treated as a one-off exercise but as an ongoing obligation to protect participants' rights as data is processed, analysed, and shared. This also involves putting in place technical safeguards, for example: access controls that limit who can view or process data; and secure environments supported by encryption (see for an innovative example of pseudonymisation by design, Box 3.3).

**Box 3.3: Linkage of administrative data in a data protection sensitive way – The case of Austria**
*Albert Kraler*

On the national level, a wide range of statistical indicators on irregular migration are available from different administrative databases, including those on migration enforcement (apprehensions, return orders, rejections at the border, migrant smuggling, etc.), asylum databases, and residence permit databases. Despite some inherent limitations associated to their administrative purpose, the anchoring of measurement concepts in operational and legal categories and their specific scope linked to domain specific regulatory frameworks, administrative databases provide a rich source for scientific analysis. This is particularly true when they contain historical data and allow examining migrants' trajectories (chapter 7) or when they allow linkage of different databases (record linkage). In both cases, questions about data protection arise. For example, in compliance with the privacy regulations databases generally foresee a certain timeframe after which personal data needs to be deleted, if no longer necessary for the particular administrative purpose they are meant to serve. Sometimes, specific events will lead to the deletion of records from registers. For example acquisition of citizenship will result in the deletion of that person's records from residence permit registers). Similarly, record linkage can be restricted by law, as is the question of who has access to different types of data.

The case of Austria is a good example of database linkage and the preservation of historical records are possible in a data protection compliant way. In Austria, the pseudonymisation of register data for statistical purposes is achieved through the use of (encrypted) sector specific personal identifiers (verschlüsselte Bereichsspezifische Personenkennzahl Amtliche Statistik – bPK-AS). The bPK-AS is generated by the Stammzahlenregisterbehörde (Central Register Authority). It is a cryptographically derived identifier derived from the personal identifier used in a specific domain (for example social security, or the population register code) and a code for the domain.[3] It is unique to each individual

3 The principle of encryption used for the generation of the sector specific identifiers is described (in German) here : https://www.bundeskanzleramt.gv.at/agenda/digitalisierung/stammzahlenregisterbehoerde/bereichsspezifische-personenkennzeichen/beschreibung.html.
The encryption procedure is based on Central Register Authority Ordinance (Stammzahlenregisterbehördenverordnung) 2022, see https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=20011934.

and serves as a key to match data from different registers with each other. Crucially, the bPK-AS is not reversible, meaning it cannot be traced back to the original personal identification number (Statistics Austria 2024). Statistics Austria uses these anonymised personal identifiers to link data from various sources – such as social insurance records, employment data, and education registers – through deterministic linkage and without revealing personal identities. While Statistics Austria gets updates from administrative databases in real time, it uses anonymized statistical mirror databases for statistical purposes (Fuchs et al. 2024). All register data is stored in a historicised way, allowing longitudinal analysis.

Since 2022, all statistical databases based on data collected by Statistics Austria itself (through surveys and other statistical reporting systems) as well as a wide range of administrative databases from different public bodies are assembled in the "Austria Micro Data Centre" (AMDC).[4] By mid-2026, all public administrative database – with the exception of security related databases – should be made available by the AMDC. In addition, researchers can link their own datasets to the AMDC by obtaining a sector specific identifier from the Central Register Authority for their own dataset, which in turn enables Statistics Austria to include this dataset in the AMDC, making it linkable to all datasets contained in the AMDC. A precondition for including a dataset in the AMDC is that researchers collect personal information (notably name, date of birth, place of residence) to enable pseudonymisation by the Central Register Authority. The AMDC is open for researchers in accredited institutions, which need to meet a number of criteria for accreditation (such as scientific purpose of the organisation, research quality, independence).

While immigration and migration enforcement related databases are not (yet) linked to the AMDC and therefore cannot be used to analyse legal status trajectories, the design of the system nevertheless can serve as a model for balancing data utility and privacy protection.

**References**

Fuchs, R., Göllner, T., Hartmann, S. & Thomas, T. (2024). "Fostering Excellent Research by the Austrian Micro Data Center (AMDC)" Jahrbücher für Nationalökonomie und Statistik, 244(4), 433-445. https://doi.org/10.1515/jbnst-2023-0043

Statistik Austria (2024): Standard-Dokumentation Metainformationen (Definitionen, Erläuterungen, Methoden, Qualität) zu den Registerbasierten Erwerbsverläufen. https://www.statistik.at/fileadmin/shared/QM/Standarddokumentationen/B_1/std_b_erv.pdf

Importantly, 'special categories' of personal data, such as 'race', ethnic origin or political opinions, carry heightened risks. The MIrreM project applies the 'data minimisation principle' by deliberately limiting data collection to what is strictly necessary, while ensuring individuals are fully informed of their rights and protections.

When applied carefully, these practices allow data to be used constructively: for example, to understand migration patterns, design inclusive services and improve resource allocation without compromising individual privacy.

4 https://www.statistik.at/en/services/tools/services/center-for-science/austrian-micro-data-center-amdc

### Secondary use of data

MIrreM also uses existing datasets to estimate irregular migrant populations. Even if these are anonymised or aggregated, ethical issues remain. Researchers and policymakers must consider the conditions under which data were originally collected, and if this included informed consent, voluntariness and transparency, and how linking datasets may create new risks or reinforce surveillance logics.

To address this, researchers need to commit to clear documentation of data sources, ethical review of any secondary use, and a critical assessment of how data linking may affect the rights and perceptions of the populations concerned. Policymakers must be wary of normalising data practices that reinforce securitisation narratives, where migrants are framed primarily as risks to be managed rather than individuals with rights.

At the same time, responsible linking and analysing of secondary data can yield valuable insights for planning services, understanding the characteristics of migrant populations and evaluating the effectiveness of policies. This requires a careful balance between administrative utility and respect for fundamental human rights.

### Inclusive governance and legal safeguards

The use of data about irregular migration should complement, not replace, engagement with migrants themselves or with civil society organisations that work with them directly. Policymakers should strive for inclusive governance in migration data systems, ensuring that policy proposals reflect diverse perspectives and do not solely rely on technocratic or quantitative assessments.

Policymakers must also consider the need for updated legal frameworks to regulate the use of linked or repurposed datasets, especially when applied to groups that may lack formal protections. This includes reviewing data protection laws and institutional safeguards to ensure that they cover the specific vulnerabilities associated with an irregular residence status.

When governance frameworks are inclusive and transparent, data can be used proactively to identify gaps in protection, target resources effectively and support interventions that benefit migrants and wider communities.

## Conclusion

Research ethics in the context of irregular migration cannot be reduced to a checklist. Compliance with legal frameworks such as GDPR and the EU AI Act is necessary, but only as a baseline. What is required instead is an ongoing reflexive approach about the risks, responsibilities, and power relations involved at every stage – from research design and data collection to analysis and communication. For those involved in data collection and processing, such as researchers, statisticians, public sector officials and those working in migrant support organisations, this means embedding ethics awareness in all activities, recognising the rights and dignity of those whose lives are studied, and promoting transparency and accountability in the production and use of migration data. By treating ethics as an integral, continuous process, researchers can help ensure that their work contributes to more just, humane, and evidence-informed migration policy.

# References

Anderson, B. (2019). New directions in migration studies: Towards methodological de nationalism. Comparative Migration Studies, 7(1), 1-13. https://doi.org/10.1186/s40878-019-0140-8

ALLEA. (2023). The European Code of Conduct for Research Integrity (Revised ed.). Berlin: All European Academies. https://doi.org/10.26356/ECOC

Bakewell, O. (2008). Research beyond the categories: The importance of policy irrelevant research into forced migration. Journal of Refugee Studies, 21(4), 432–453. https://doi.org/10.1093/jrs/fen042

Cyrus, N. (2023). Ethical Benchmarking for the Measurement of Irregular Migration. In MIrreM Policy Brief No.1. Krems: University for Continuing Education Krems (Danube University Krems). https://doi.org/10.5281/zenodo.10042022

Dahinden, J. (2016). A plea for the 'de-migranticization' of research on migration and integration. Ethnic and racial studies, 39(13), 2207-2225. https://doi.org/10.1080/01419870.2015.1124129

Descamps, J. (2024). Can We See Their ID? Measuring Immigrants' Legal Trajectory: Lessons From a French Survey. International Migration Review, https://doi.org/10.1177/01979183241295995

European Union (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. Official Journal of the European Union, L, 2024(1689), 1–127. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union, L119, 1–88. https://eur-lex.europa.eu/eli/reg/2016/679/oj

Floridi, L., Taddeo, M., & Turilli, M. (2018). Group privacy: A defence and an interpretation. In L. Taylor, L. Floridi, & B. van der Sloot (Eds.), Group privacy: New challenges of data technologies (pp. 83–100). Cham, Switzerland: Springer.

Hendow, M., Wagner, M., Ahrens, J., Cherti, M., Kierans, D., Kraler, A., Leerkes, A., Leon, L., Rodríguez Sánchez, A., Siruno, L., Tjaden, J., & Vargas-Silva, C. (2024). How fit is the available data on irregular migration for policymaking?. In MIrreM Policy Brief No. 3. Krems: University for Continuing Education Krems (Danube University Krems). https://doi.org/10.5281/zenodo.13757685

Kraler, A., Reichel, D., & Entzinger, H. (2015). Migration statistics in Europe: A core component of governance and population research. Integrating immigrants in Europe: Research-policy dialogues, 39-58. https://doi.org/10.1007/978-3-319-16256-0_3

Jasso, G., Massey, D. S., Rosenzweig, M. R., & Smith, J. P. (2008). From illegal to legal: Estimating previous illegal experience among new legal immigrants to the United States. International Migration Review, 42(4), 803–843. https://doi.org/10.1111/j.1747-7379.2008.00148.x

Mohan, S. S., Mountz, A., Romero, M., & Visan, A. (2023). How to research 'irregular' migration: approaches and perspectives from the field. In Research Handbook on Irregular Migration (pp. 36-48). Edward Elgar Publishing. https://doi.org/10.4337/9781800377509

Reed-Berendt, R., Dove, E. S., Pareek, M., & Group, U.-R. S. C. (2022). The Ethical Implications of Big Data Research in Public Health: "Big Data Ethics by Design" in the UK-REACH Study. Ethics & Human Research, 44(1), 2–17. https://doi.org/10.1002/eahr.500111

Scheel, S., & Tazzioli, M. (2022). Who is a migrant? Abandoning the nation-state point of view in the study of migration. Migration Politics, 1(1), 002. https://doi.org/10.21468/MigPol.1.1.002

Taylor, L., & Meissner, F. (2024). Migration statistics in times of large-scale mobility data: ethical concerns and concerns with ethics. In W. L. Allen, & C. Vargas-Silva (Eds.), Handbook of research methods in migration (2nd edition, pp. 280-295). Edward Elgar Publishing. https://doi.org/10.4337/9781800378032.00031