ethmig survey data

# SURVEYS TO ETHNIC AND MIGRANT MINORITIES ACROSS EUROPE

## Identifying Knowledge Strengths and Gaps Using Survey Metadata

REPORT 2 – WG1 & WG2


**LAURA MORALES** Sciences Po, CEE & LIEPP, CNRS

**AMI SAJI** Sciences Po, CEE, CNRS

**DIMITRI PRANDNER** Paris Lodron University of Salzburg & Johannes Kepler University of Linz

**JOHANNES BERGH** Institute for Social Research, Oslo

**ANIKÓ BERNAT** TARKI Social Research Institute

**MÓNICA MÉNDEZ LAGO** Centro de Investigaciones Sociológicas



cost EUROPEAN COOPERATION IN SCIENCE & TECHNOLOGY

Funded by the Horizon 2020 Framework Programme of the European Union

SciencesPo CENTRE FOR EUROPEAN STUDIES AND COMPARATIVE POLITICS

SSHOC social sciences & humanities open cloud

cnrs

## ABSTRACT

This report, the second in the series of reports produced by COST Action 16111 Ethmigsurveydata, presents the results from the analysis of the survey metadata compiled through the Ethnic and Migrant Minorities (EMM) Survey Registry launched by the COST Action and with the support of the H2020 project Social Sciences and Humanities Open Cloud (SSHOC). The current version of the report focuses on six countries for which the metadata is complete and controlled for quality: Croatia, Norway, Switzerland, Romania, Germany and Turkey. The report begins with an introduction about the survey metadata compilation undertaken by Ethmigsurveydata, followed by a detailed overview of the methodological approaches used to develop the EMM Survey Registry. The report is then structured into five additional sections that present the findings for the EMM surveys in these six countries in relation to their geographical and target group coverage, the definition and measurement of EMMs, the topics covered by the surveys, their technical characteristics, and their accessibility and reusability. The final section offers conclusions and recommendations for funders, data producers and data analysis.



**KEYWORDS:**

Survey research, ethnic minorities, migrant minorities, FAIR principles, metadata, Europe.

**SPECIFICATION OF THE VERSION:**

Version 1 including data for Croatia, Germany, Norway, Romania, Switzerland, and Turkey.

**DATE OF PUBLICATION:**

April 2020, Sciences Po, Paris

**CITATION FOR THE REPORT:**

Morales, Laura; Saji, Ami; Prandner, Dimitri; Bergh, Johannes; Bernat, Anikó & Méndez Lago, Mónica. (2020, April 30). Surveys to Ethnic and Migrant Minorities across Europe. Identifying Knowledge Strengths and Gaps Using Survey Metadata (Version 1). Zenodo.

http://doi.org/10.5281/zenodo.3839677

# TABLE OF CONTENTS

# ACKNOWLEDGEMENTS

The production of this report has been possible thanks to the financial support of COST Action 16111 Ethmigsurveydata, funded by the Horizon2020 Framework Programme of the European Union. The work on the development of the Ethnic and Migrant Minorities Survey Registry has also been possible thanks to funding from the H2020 project Social Sciences and Humanities Open Cloud (SSHOC) funded through Grant Agreement 823782, in the context of WP9 'Data Communities'. We are very thankful for this funding but we note that the funders bear no responsibility for the data or our interpretation presented in this report.

This report and the data on which it is based would not have been possible without the collective effort and collaboration of a large number of researchers that have contributed their time and ideas to COST Action 16111 Ethmigsurveydata. They are too numerous to be listed here. We would like, nevertheless, to thank all of the members of Work Group 1 and Work Group 2 and, especially the researchers that compiled the metadata for the countries included in this report and those that participated in the quality control process: Margareta Gregurovis (Croatia), Simona Kuti (Croatia), Jonas Elis (Germany), Lukas Fiege (Germany), Manuel Diaz Garcia (Germany), Sabrina Jasmin Mayer (Germany), Dennis Spies (Germany), Dina Heider Hov (Norway), Monica Șerban (Romania), Paula Tufiș (Romania), Natalia-Cornelia Malancu (Switzerland), Burak Agalday (Turkey), Sedef Turper Alışık (Turkey), Feyza Köseoğlu Darılmaz (Turkey), Resmiye Demet Kuru (Turkey), Z. Maviş Yildirim (Turkey), Ludhiya Johnson (quality control), Hugo Marcos-Marne (quality control), Foteini Panagiotopoulou (quality control), Carles Pamies (quality control), Alina Thiemann (quality control), Claire Vincent-Mory (quality control), and Anny Yip-Ching Yu (data preparation).

# 1 EXECUTIVE SUMMARY

This report presents an overview of existing quantitative surveys that target ethnic and migrant minorities (EMMs) across Europe and neighbouring countries. It uses survey-level metadata (i.e. data about the data) that has been rigorously and systematically compiled by members of the COST Action "International Ethnic and Immigrant Minorities' Survey Data Network" (COST Action 16111 Ethmigsurveydata). The primary motivation of Ethmigsurveydata to undertake this metadata compilation work has been to improve how quantitative surveys on EMMs' integration is accessed, shared, and (re)used by researchers and policymakers; in other words, making the data FAIR (Findable, Accessible, Interoperable and Reusable).

The current version of the report focuses on six countries for which the metadata is complete and controlled for quality: Croatia, Norway, Switzerland, Romania, Germany and Turkey. This report is the first presentation of the metadata that has been compiled by Ethmigsurveydata. Updated versions of this report and additional publications are forthcoming; they will cover more than the six countries included in this version. The metadata used in this report is also being made available through a publicly available online tool developed by Ethmigsurveydata (with support of the H2020 project SSHOC), the EMM Survey Registry.

This report begins with an introduction about the survey metadata compilation undertaken by Ethmigsurveydata (section 2), followed by a detailed overview of the methodological approaches used to develop the EMM Survey Registry (section 3). The report is then structured into six additional sections and the main findings for each section are presented below:

## Section 4 - Geographical and target group coverage of the EMM surveys in Europe:

1. Half of the surveys were conducted at the national level and the other half at the subnational level. At the country level, this balanced divide also occurred in Switzerland, Germany, and Croatia.
2. Subnational surveys conducted in Germany and Croatia covered different geographical levels. However, those conducted in Switzerland, Norway, and Turkey were predominantly focused on cities and urban centres.

3. Across the countries, there was heterogeneity in terms of when the surveys were conducted (i.e. highly uneven start/end dates with no clear identifiable trend). Nevertheless, the highest number of surveys were conducted in 2015 – the year of the summer of migration. Most of these 2015 surveys were conducted in Germany (primarily at the national level) and Turkey (primarily at the subnational level).

4. Just one in five surveys (n=31) were part of an international or cross-country collaboration. Germany was notably a common participant in international survey programmes.

5. Of the 161 surveys, more than twice as many were targeted at migrants than ethnic minorities. In fact, it was only in Romania and Croatia where ethic minority-targeted surveys (many of those conducted with the Roma population in Romania and to both Roma and other ethnic minorities in Croatia) were "well represented".

## Section 5 - The definition and measurement of EMMs in surveys across Europe

1. 63 different EMM categories were identified and they can be classified into five distinct types: by origin, by religion, by legal or administrative status, by ethnicity, and by socio-economic status.

2. Surveys often defined their EMM population using generic categories (e.g. migrants, refugees), though some specific categories referring to certain countries of origin or specific ethnic minorities were quite common (e.g. Roma, Turkish).

3. Certain EMMs were much more studied in some countries than in others. For example, Turkish-origin migrants were the most studied EMM group in Germany and Switzerland.

4. Across the six countries and within each country, a majority of the surveys had strictly adult respondents (8 years old and older or 15 years old and older) and an overwhelming majority surveyed both male and female respondents.

5. When operationalising the EMM target population, the most frequently used criterion referred to factual issues such as the current citizenship/nationality of the respondent (59%), the respondent's country of birth (32%), and/or the country of birth of the respondent's parents/grandparents (29%). Nevertheless, subjective criteria such as self-identification as part of an ethnic/migrant group was used to operationalise the EMM target population in roughly a fifth of the 161 surveys.

6. At the country level, nationality and country of birth of the respondent were the most widely used criteria to operationalise the EMM target population, with the exception of Romania. In the nine surveys conducted in Romania, some sort of ethnic self-identification was used, while a respondent's nationality/country of birth was used less frequently primarily because the ethnic minority population is native born.

7. When surveys included EMM-specific questions, they were primarily those that were

"hard" and factual; specifically, around half of the surveys applied questions on the current nationality of the respondent, on the mother tongue/language and/or on the country of birth of the respondent. However, for surveys targeting ethnic minorities (e.g. Roma), "soft" EMM-specific questions, such as self-identification, were much more common.

8. Over 40% of the surveys included a majority population as a subgroup. However, these surveys tended not to be general population surveys, and instead those that were targeted specifically at an EMM population.

## Section 6 - Topics covered by EMM surveys in Europe

1. A wide range of topics were covered across the 161 surveys. Specifically, 29 different topics were named, plus some additional "other" topics, ranging from child care to military service as well personal wellbeing.

2. The topics covered by the surveys have changed over time. Since 2000, political inclusion, participation, and social/political attitudes has become a less frequently studied topic. Noticeable topic shifts were also observed in the post-2014 period. After 2014, economy related topics (e.g. labor market integration, poverty related issues) were covered more frequently, and there was increased interest in topics such as human capital, skills, and educational attainment. Moreover, after 2015, there was an increase in the number of surveys dealing with forced migration and asylum seekers.

## Section 7 - Technical characteristics of EMM surveys in Europe

1. Almost half (n=80) of the surveys were single cross-section. The remaining were either repeated cross-section (n=40) or longitudinal (n=41).

2. 45% of the surveys were identified as being representative of the target (sub) population. Surveys in Switzerland and Norway in particular had a high proportion of surveys identified as "representative".

3. Random sampling/selection was the most common sample design across the 161 surveys. This trend also held true for surveys identified in Norway (100%), Romania (89%), and Germany (52%).

4. Over half of the surveys had at least one partitioned subgroup as part of its survey. The presence of a partitioned subgroup appeared to be equally prevalent at the national and subnational level.

5. A variety of sampling frames were used/identified. The most common ones across the six surveys, within each territorial scope (national vs. subnational), and within each country were (generally): population registers/registries, census, and telephone list/directory/records. One true anomaly was Norway where the electoral roll was the most common sampling frame.

6. Individuals and households were the two most frequently used sampling units at the aggregate, territorial scope, and country level.

7. We principally targeted national surveys with more than 300 respondents and subnational level surveys with more than 150 respondents. Most of the surveys met the relevant threshold.

8. Response rates for surveys were highly varied; Germany in particular had a wide range, as it included some surveys with very low response rates, and some with unusually high rates of response.

9. Response rate calculations were missing for a sizeable number of surveys. Only surveys in Norway had this information reported for all of their surveys.

10. A substantial share of the surveys were done via face-to-face. Telephone, web- and postal surveys were also frequently used. Except for the relatively high frequency of face-to-face interviews, there was no particular pattern in terms of data collection mode.

11. A relatively small number of surveys provided information about personal interviews that had been conducted with the EMM respondents. Across the six countries, professional interviewers were more common than non-professional ones, and a majority of interviewers knew an EMM spoken language.

12. About half of the surveys provided a questionnaire in an EMM-spoken language.

## Section 8 - Accessibility and reusability of EMM surveys in Europe

1. When it comes to simple availability of the referenced survey data, approximately 40% is either publicly available (via data archives) or available upon request from the involved researcher. However, for a third of the surveys the data is unavailable and for a quarter of the surveys the data availability is unknown.

2. Strong national differences were observed when it came to data availability. In Croatia, Romania, and Turkey an overwhelming majority of the data were unavailable or their availability was unknown. By contrast, data for Norwegian and Swiss surveys were all available and two thirds of the data sets for German surveys were available, either publicly or via request.

3. Data availability seems to have been correlated with the data archiving practices of the country. For instance, all the surveys conducted in Norway (n=9) were archived and the data for all of these surveys have been made available.

4. Availability inevitably had an impact on how findable and citable surveys (including their data and documentation) were. As surveys identified in Norway and Switzerland were "more available", they were more likely to have citable documentation and data; they were also more likely to have provided a link or DOI for their data and documentation.

5. To facilitate access to and reuse of surveys, offering documentation in English is critical.

While all nine surveys identified in Switzerland had English language documentation, not a single survey in Turkey did. Moreover, for the other four countries, the results were more mixed, where the availability of English language documentation varied from survey to survey.

## Section 9 - Conclusions and recommendations for future action

The final section offers a number of tentative recommendations subject to the completion of the analysis of the remaining countries:

1. Future surveys should improve the knowledge base on the integration of EMMs in towns and rural areas, as these are considerably less covered by the existing sets of surveys.

2. Cross-national programmes of surveys targeting EMM populations should be fostered to allow for a better comparative knowledge base.

3. Surveys that aim at covering the entire migrant or ethnic minority population are required for a more comprehensive understanding of the integration of EMM populations.

4. Countries with intense survey research exclusively on migrant minorities should consider expanding the focus to ethnic minorities as well, and vice versa.

5. Funders may want to promote multi-purpose surveys to EMM populations that contribute to the knowledge needs of academics, policy makers and civil society organisations through more integrated approaches to survey design that do not create knowledge silos.

6. Governments — at all levels, EU, national, regional, local — may want to consider including at least one periodical survey (repeated cross-section or longitudinal) specifically targeting the EMM population within their jurisdiction as part of their official statistics portfolio.

7. Surveys that focus on EMMs that are also religious minorities are not very common and, hence, future funding efforts could be directed at studying whether these groups face specific challenges in terms of equal opportunities for integration and inclusion.

8. A future balanced focus on the EMM youth and children would be particularly useful, as the current knowledge base on these age groups is very limited across most countries.

9. When designing the pre-screening and questionnaire instruments, survey data producers should consider a wider range of options to define ethnic and migrant minorities. Survey instruments that allow for a richer set of variables gauging migrant background and ethnic identities will result in more nuanced analyses about orientations, experiences and outcomes.

10. Surveys targeting specifically the EMM population should always consider the opportunity and benefits of including a subgroup of the majority population as a point of reference. Typically, surveys to EMMs are designed to inform about experiences and challenges of integration but it is very difficult to gauge these without a point of reference to the non-minority population.

11. Surveys to EMM populations do not seem to be sufficiently capturing certain topics that would seem a priori critical for the proper understanding of the integration of these populations.

In particular, surveys directed at migrant populations should always include sufficient information about the reasons and drivers for migration, as well as about the legal status and administrative situation.

12. The study of important topics — such as the situation of unaccompanied migrant minors and human smuggling and trafficking — should be promoted in future survey research as the knowledge base is very limited.

13. A number of countries should consider investing in surveys that can introduce a temporal element with repeated cross-sectional and longitudinal designs, as this adds value to survey data collection and enhances the knowledge base.

14. Survey research on EMM populations relies too much on samples that are not statistically representative of the underlying population and survey data producers should, in the future, learn from the best practices in the field to adapt to local circumstances while retaining the beneficial properties of probability sampling.

15. In the future, survey data producers of studies on EMM populations need to make a greater effort at providing detailed documentation and transparency in relation to the technical characteristics of the surveys, including (but not restricted to) sample design, questionnaire administration, fieldwork operations and response rates.

16. Survey data producers should bear in mind that the translation of the questionnaires into the mother tongue of the minorities studied is not only good practice but also highly advisable to ensure acceptable response rates that will reduce the bias in the outcomes obtained from the survey.

17. International and national efforts are required to ensure that a larger number of surveys can be fully documented, archived and made available and citable for future re-use. National data archives as well as CESSDA should play proactive roles in fostering the future preservation of the existing surveys focusing on EMM populations.

18. Research funders, data producers and data users should become more aware about the collective benefits of making EMM survey data FAIR and embed the FAIR principles into their ordinary research practices.

# 2 INTRODUCTION

The purpose of this report is to present an overview of survey data that targets ethnic and immigrant minorities across Europe and neighbouring countries. The current version of the report (v1) focuses on six countries for which the data is complete: Croatia, Norway, Switzerland, Romania, Germany and Turkey. This metadata (data about data) is the result of the efforts of the members of the COST Action "International Ethnic and Immigrant Minorities' Survey Data Network" (COST Action 16111 Ethmigsurveydata). The main aim of Ethmigsurveydata is to gather information about survey data where the target groups are ethnic and immigrant minorities. Detailed information about each survey has been collected in a rigorous and systematic way, which enables analysis and comparison of metadata between surveys and between countries. This report is the first presentation of the metadata from Ethmigsurveydata. Updated versions of this report and additional publications are forthcoming that will cover more than the six countries in this version. The metadata is also being made available at the online EMM Survey Registry.

The work of Ethmigsurveydata is limited to quantitative surveys of ethnic and immigrant minorities. The data in this report and in the survey data registry exclusively contains quantitative survey data. Other types of data that are important and of use to researchers, such as qualitative data, census data, register data, data on organizations, institutions et cetera, is not covered by the work of the COST Action, and are therefore not a part of this report.

The motivation for the work of the COST Action in collecting metadata was a belief by a number of its participants that data was being collected in individual countries that was not being shared with researchers in other countries, nor was there as much systematic comparison of results across countries as ideally there should be. Several notable comparative studies utilize survey data of minorities to compare across countries1. There are also numerous individual-country studies. It still seemed at the outset of the COST Action's work that there would be an unused potential in single-country survey data that targeted immigrant- and ethnic minorities. That suspicion was justified when the metadata collection revealed numerous in more than 30 European countries. The purpose of the metadata collection and of this report is to encourage research in existing data and to engender learning of methodology, best practices and the like from these previous data collection efforts. The latter point is important, because researchers are faced with some

1 See e.g. Bilodeau, A. (ed. 2016), *Just Ordinary Citizens. Towards a Comparative Portrait of the Political Immigrant.* Toronto, Canada: Toronto University Press.
Bird, K., T. Saalfeld and A. Wüst (eds. 2010): The Political Representation of Immigrants and Minorities: Voters, Parties and Parliaments in *Liberal Democracies*, Routledge Publishers.
Morales, L., Giugni, M. (Eds. 2011): Social Capital, Political Participation and Migration in Europe. Making Multicultural Democracy Work? Palgrave Macmillan.

distinct challenges when surveying minority populations. Four distinct challenges stand out, with respect to 1) sampling, 2) response/non-response 3) mode of interviewing and 4) trust in researchers or data-collectors.

## Challenge #1: SAMPLING

For some researchers in some countries, drawing an adequate (gross) sample of a minority population is the largest methodological hurdle. If there is no information about the relevant minority status in national population registers or in census data, which can be used to draw samples, researchers or data-collectors have to use other methods to research their targeted population. The three most common methods to do that is, first, to *target one or several specific geographic locations*. By conducting interviews in geographic areas where there is a high density of the minority population, one can sometimes reach an adequate sample. The second approach is to use *screening interviews*. The first question for the respondent is whether he or she is a member of the relevant minority group. If not, the interview ends right away; full interviews are only done with people in the targeted minority group. Third, the *snowball* technique requires that one starts with a small selected group of the relevant population, for instance the members of an ethnic minority organization. The interviewers then asks for contact information of others in the same minority group, and the process repeats itself until one has an adequate sample. Finally, a combination of two or perhaps all three of these techniques is frequently used.

The problem with these three techniques is that none of them produces a fully random or representative sample of the entire minority population. These approaches are also quite costly and time-consuming. In some instances, like in the Nordic countries, population register data that includes information about some minority statuses can be used by researchers to draw samples. In Europe, these types of data do not include information about race or ethnicity, but sometimes immigrant-related information is available, such as country of birth. If that information can be used to identify the target population, researchers have an excellent and cost-effective tool for drawing samples of those groups.

## Challenge #2: RESPONSE/NON-RESPONSE

The second challenge faced by all survey researches is that of generally declining response rates. While there is some variation from one country to another or from one minority group to another, the general pattern seems to be that response rates are even lower among minority populations than among majority populations. The main problem is usually lack of contact with

the respondents. This could be due to inadequate or erroneous contact information (addresses or telephone numbers), or it could be because the contacted individuals choose not to respond. The key problem with low response rates, as is well known by most researchers in this field, is that the people who do respond to the survey may not be representative of those that do not. Any efforts that can be made to raise the response rate of a survey will therefore also raise the credibility and probably also the quality of the data.

## Challenge #3:
## MODE OF INTERVIEWING

The four principal methods for conducting survey interviews are: 1) face-to-face, 2) telephone, 3) paper fill-in questionnaire and, 4) web-based questionnaires. The methods are ordered by the costs involved, with face-to-face being the most expensive, and web-based being the least expensive. Face-to-face interviews are often seen as the gold standard of survey research, and is still used by for instance a number of national election study programs. There is probably no better way of conducting a fairly long interview that requires the sustained attention of the interviewee. Apart from that, there is no universal agreement on what a preferred mode of interviewing is. It depends on the length, the purpose, the targeted population of the survey, and certainly on budgetary restraints. Another important issue when targeting minority populations whose native language differs from the official language(s) of a country is whether questionnaires should be administered in the official language(s), in the native languages of the targeted population, or both.

## Challenge #4:
## TRUST

A final challenge when surveying minority populations may be one of trust. Some minorities are the subject of frequent surveys or other types of research efforts. There is a joke in Norway's Sami community that the average Sami family consists of a husband, a wife, two children and an anthropologist... Being the frequent target of researchers could render some sceptical of the value of an additional survey. If previous research is seen as having a stigmatizing effect on the minority community, people may be especially reluctant to take part.

There is great variation between minority communities in this respect, but in some cases, the relationship between a minority and majority community is contentious. Researchers and others conducting surveys of minorities should be sensitive to and aware of these types of trust-issues

In this context, the EMM Survey Registry intends to provide a free and open data tool that will make EMM Surveys across Europe FAIR - Findable, Accessible, Interoperable and Reusable - by

providing detailed information (metadata) about their aims, coverage, characteristics, focus and availability for public or restricted use. The EMM Survey Registry offers a single and specialized point of access to the information of all the surveys undertaken across Europe since January 2020 to samples (or sub-samples) of EMM populations that are large enough to allow analysis of the patterns of integration of these populations. In other words, the EMM Survey Registry facilitates finding (the F in FAIR) quantitative survey data that allows studying the integration of EMM populations. Additionally, to the extent that some of the surveys are already deposited and archived in some repository or data archive, we also promote their accessibility and re-use (the A and R in FAIR) by a wider range of potential users.

Hence, the main purpose of the EMM Survey Registry is to extend the visibility and the future re-use of EMM survey data across Europe with an aim to foster wider-ranging and more comprehensive comparative analyses of the integration of EMM populations. To this end, the researchers collaborating within COST Action 16111 Ethmigsurveydata have collected extremely comprehensive metadata on 11 elements describing each of the surveys identified:

1. General identification information about the survey;
2. Information about the inclusion of the survey in a larger study;
3. Ethnic and migrant minority (EMM) target population;
4. Sampling method;
5. Sample size for the overall survey;
6. Sample sizes for any subgroups in which the survey is partitioned;
7. Data collection information;
8. Availability to research community;
9. Data producers, owners, distributors and citations;
10. Additional information;
11. Information on this compilation of metadata;

For each of these elements, several pieces of information are compiled that allow gaining a proper and detailed understanding of how the survey was designed and conducted, as well as of the types of data it has produced. The aim of the data compilation process for each of the countries covered in the EMM Survey Registry and in this report has been to be as comprehensive as possible in the detection of surveys that have been produced in each of the countries included in this scoping exercise. Nevertheless, because many of the EMM surveys remain somewhat hidden because of the low rates of data preservation in central data repositories and archives, it is unavoidable to have missed some.

Beyond the main purpose of rendering EMM survey data visible to the various research and user communities in the field of ethnic and migrant minority studies, the extremely rich metadata that has been collected by the researchers collaborating within COST Action 16111 Ethmigsurveydata also allows us to obtain a very clear picture of the strengths and weaknesses in knowledge accumulation in this field of research. This report will, thus, not only record the main features and purposes of the surveys that exist on EMM integration, but will also allow us to point out to the knowledge gaps stemming from the last 20 years of quantitative survey research on EMM integration. In so doing, this report aims at contributing to shape future research agendas and funding investments, by highlighting areas where our cumulative knowledge is reasonably sound and others where it is highly deficient and in need of more (or better) data.

Version 1 of the report includes metadata only for six countries: Croatia, Germany, Norway, Romania, Switzerland and Turkey. The report is structured as follows. The next section (section 3) describes the methodology that was used to design the EMM Survey Registry: how we delimited the surveys of interest, the process we employed to search and find information about the surveys, what metadata variables we have collected, and which processes of data quality and technical adaptations we have followed to standardise the metadata across countries. Section 4 follows with a detailed description of the coverage of the surveys found in terms of the geographical or territorial scope, the time coverage, the types of EMM population covered, and the main purpose of the survey. Next, section 5 focuses in depth on how EMM populations are studied, defined, identified, categorised and measured in these surveys; allowing to gauge the heterogeneity in the study of EMM populations and their integration across Europe. After that, section 6 looks in detail at the topics that are covered by EMM surveys and how these vary across countries, across geographical scope (e.g. national versus subnational surveys) and over time. In turn, section 7 examines closely the technical characteristics of the surveys found and documented, describing their representativeness, sample design, response rates and data collection modes. Section 8 then follows with an analysis of the accessibility and reusability of the survey data that we have detected across Europe, offering an overview of types, levels and ease of access to the survey data as well as to the documentation and questionnaires. Finally, section 9 presents our conclusions and a number of recommendations for future action by research funders, social science data archives and data producers.

# 3 METHODOLOGICAL APPROACH TO THE CREATION OF THE EMM SURVEY REGISTRY

## A. DELIMITING THE UNIVERSE OF SURVEYS OF INTEREST

The EMM Survey Registry is the result of the collective effort of dozens of researchers brought together thanks to COST Action 16111 Ethmigsurveydata. Between June 2017 and December 2018, collective deliberations among the members of work groups 1 and 2 (WG1 & WG2) proceeded in various stages. First, the team discussed how to delimit the universe of surveys that we aimed at identifying and documenting with metadata. These discussions took place between June and December 2017 and ended with the decision to limit the scope of the search to quantitative sample-based survey data that focused either exclusively or within general population studies on one or several dimensions of the integration of EMM populations. This meant excluding from the registry all administrative and official statistics that were not coming from sample-based structured questionnaire-produced surveys, as well as excluding qualitative studies. These decisions were based, on the one hand, on pragmatic considerations of workload and capacity to reach all these different types of data, but also, on the other hand, on the fact that other organisations are compiling metadata hubs on administrative records and official statistics - such as the Data Catalogue of the Knowledge Centre on Migration and Demography of the European Commission - or on qualitative studies - such as the CrossMigration Database. Perhaps the more controversial of the delimiting decisions we made was the exclusion of the official census in each country, given that in some countries the official census is no longer (only) a full traditional decennial census of the population but a combination of traditional census data and/or of register data with very large sample-based surveys (e.g. in France, Israel, the Netherlands and the United States). Eventually, we agreed to exclude official census data precisely because of the variation in how they were designed and produced across the countries covered by the EMM Survey Registry and because we felt that these sources were sufficiently well known and findable to not require inclusion in our registry.

Other decisions that we made were relating to the sample size thresholds for inclusion of a survey in the EMM Survey registry. Given that we faced a wide range of situations in terms of the extent to which countries were data-rich or data-poor, we privileged an approach that gave each national team the freedom to choose between two options and to properly document their choice with the delivery of the metadata. Hence, we gave the following instructions to each country team:

(i) For nation-wide surveys specifically addressed to ethnic and/or migrant minority (EMM) respondents, national teams can choose between:

- Covering all nation-wide surveys, no matter the sample size, or
- Covering all nation-wide surveys with an achieved sample size of 300 or larg er (before weighing).

(ii) For sub-national surveys specifically addressed to ethnic and/or migrant minority (EMM) respondents, national teams can choose between:

- Covering all sub-national surveys, no matter the sample size, or
- Covering all sub-national surveys with an achieved sample size of 150 or larger.

(iii) For nation-wide or subnational general population surveys with substantial sub-samples of ethnic and/or migrant minority (EMM) respondents, national teams need to exert their discretion and include these surveys if, at least, one of the following conditions of relevance are met:

- The sub-sample for the EMM respondents includes 400 or more individuals (for countries with relatively small minority populations) or 800 or more individuals (for countries with relatively large minority populations);
- The topics of the survey are particularly well suited for the study of the integration of EMMs in that country and the survey includes a sizeable number or EMM respondents (even if smaller than 400);
- There are no (or very few) alternative sources of survey data on EMM's integration in that country and the survey includes a sizeable number or EMM respondents (even if smaller than 400).

## B. HOMOGENIZING THE SEARCH PROCESS TO ENSURE COMPARABLE COVERAGE ACROSS COUNTRIES

The second stage of the methodological discussions and decision-making by the COST Action members focused on the elaboration of a detailed document with guidelines on how to search for and gather the metadata of the EMM surveys. This phase took place in iterative waves between December 2017 and December 2018. The resulting Guidelines document outlines key aspects relating to the time coverage of the search (since 1st January 2000), the types of surveys that should be included and excluded from the search, how to deal with multi-location surveys, and very importantly the search strategy to follow.

In relation to the latter, the search process was homogenized across all countries by providing the national teams with very precise instructions on how to search for surveys. Teams were instructed to follow these recommendations:

- Undertake a systematic search in the existing social science data archives in their country (see list of keywords below);
- Send a clear and well-targeted email to the email lists in their country of all associations/societies of: Demography, Economics, Geography, Ethnic/Migration Studies, Sociology, Survey Research, and Political Science;
- Contact all the main research groups undertaking research on ethnic/migrant minorities;
- Contact all the think tanks, NGOs and local/regional/national government departments that are known for undertaking research on ethnic/migrant minorities;
- Search in the PhD dissertation repositories (see list of keywords below);
- Search in Google, limiting the search to the given country and the time frame discussed above (see list of keywords below);
- Ensure that you check the appendices with the list of relevant surveys and studies contained in the EASO report on "A review of empirical surveys of asylum related migrants" produced by the University of Siena, which you can find here: https://www.easo.europa.eu/sites/default/files/easo-review-surveys-1-2.pdf.

In order to ensure that such searches were as similar as possible across all countries, a minimal dictionary of keywords for the search was provided in the Guidelines:

- Survey AND ethnic
- Survey AND migrant
- Survey AND immigrant
- Survey AND minority
- Survey AND Roma (replace Roma with other words used in your language to refer to the Roma minority: e.g. Travellers, Inti, Gypsy, etc.)
- Survey AND refugee
- Survey AND asylum seeker

These search guidelines have yielded very positive results, with more than 800 surveys having been uncovered across Europe through this scoping exercise. In the current version of this report, we only include data from six countries and 161 surveys. More data will be added in future versions of the report once the metadata for other countries is checked for quality and processed for statistical analysis

## C. METADATA SCHEMA AND METADATA COMPILATION

To consistently and effectively capture rich and meaningful survey-level metadata that can be leveraged by a wide range of user communities (most notably through our EMM Survey Registry), WG1 and WG2 developed a metadata schema through an iterative process. For each iteration of the metadata schema, WG1 and WG2 rigorously deliberated and tested it with the larger Ethmigsurveydata membership.

The final version of the metadata schema, which exists as an Excel-based template, defines the set of metadata variables that each survey record will be documenting responses for. It captures pertinent information such as: the key features of the survey; the target population of the survey; the sampling methods used; the sample sizes for the survey as a whole and for any partitioned subgroups; the method of data collection used; the availability of the survey data, questionnaire, and technical documentation; and details and specifications about data ownership and distribution. As the metadata schema was also developed to help make EMM survey data more FAIR (Findable, Accessible, Interoperable, Reusable), the metadata variables were strategically selected and defined to correspond to elements of DDI-Codebook, an international and widely-used social science data documentation standard.

Both national and subnational surveys use the same metadata schema, with the exception of additional metadata variables included for the subnational surveys. As such, the metadata schema for the national surveys includes 215 metadata variables, whereas for the subnational surveys it includes 221 metadata variables.

Given that the metadata schema is composed of over 200 metadata variables, it has been structured and organised into ten sections. Each section represents a set of metadata variables that collectively provide a specific type of information about a survey. Table 1 provides a detailed overview of each section, including the type of information it captures.

National teams representing each of the 35 COST Action participating countries were the parties responsible for filling out the Excel-based template for their respective country. This template was deliberately designed so that, at the end of the metadata compilation process, each national team would produce a single Excel file (i.e. a metadata template) that documents the survey-level metadata for every survey (national and subnational) they had uncovered through their scoping and search process (as described in section B). Specifically, the Excel template was set up with five tabs, where each tab would be used for a specific purpose:

**Instructions:**

National teams were instructed to read and review this tab carefully before using the metadata template. This tab describes how the metadata template is set up, as well as how the national teams should be using the available resources (e.g. the guidelines document for the metadata compilation work) to correctly document the survey-level metadata for the identified surveys. National teams did not document any information on this tab.

**Annotations:**

National teams were required to document information about themselves, as the individuals filling out the metadata template, and the overall experience of searching for and compiling the survey-level metadata. This tab was intended to provide a country level overview of the metadata compilation work and process. The information captured on this tab was transformed into an additional section (section 11, Information on this compilation of metadata) for the EMM Survey Registry.

**National surveys:**

National teams used this tab to document the survey-level metadata for all the national level surveys they had identified. The metadata schema was displayed in the first column of the tab, where each metadata variable (including any headings and subheadings) was placed in its own row. All other columns (those appearing to the right) would be used to document the survey-level metadata for each identified survey. In other words, with the exception of the first column, one column would represent the survey-level metadata for one survey only.

**Subnational surveys:**

This tab was set up in the same way as the one for national surveys. The only exception being that this tab was used to document the survey-level metadata for all the subnational surveys identified by the national team.

**Codes:**

This tab was not used by the national teams. It documents all the codes that had been set up and used for the closed response metadata variables requiring a drop-down menu in the corresponding cells.

## Table 1. Metadata schema by section

| Section | Information covered in the section |
|---|---|
| **1. General identification information about the survey** | Full survey name (in English and in the native language), survey acronym, territorial scope of the survey (national vs. subnational), representativeness of the target population of the survey, type of survey (e.g. cross-section, longitudinal), survey start/end dates, main topics of the survey, main purpose of the survey, coverage of the survey's target population in terms of age, coverage of the survey's target population in terms of the sex of the respondents |
| **2. Information about the inclusion of the survey in a larger study** | Full name (in English and in the native language) of the larger study that the survey belongs to; acronym of the larger study; names of other countries/regions/cities (in English) that are also part of the larger study; for larger studies that are repeated cross-section/longitudinal, date of the first survey, frequency of the waves/panels, and wave number of the survey (since each wave is coded as its own survey record); for larger studies that are part of an international survey programme, date when the survey became a part of it and frequency of the waves since the survey joined it; for larger studies that have pooled samples, number of surveys pooled, other surveys with which the survey has been pooled, and indication as to whether or not emigrants from more than one country have been pooled; details about any qualitative studies that have been linked to the survey |
| **3. EMM target population** | Description of the survey's EMM target population, classification of the survey's EMM target population into type, ways in which the EMM target population has been operationalized (by the survey producer), migrant/minority-related questions that have been included in the survey, size of the EMM target population from which the EMM sample was likely to have been drawn, indication of whether a majority subgroup was included in the survey, indication of whether the survey was designed as a general population survey |
| **4. Sampling method** | Sampling strategy used for the survey (closed and open responses), sample design of the survey, sampling frame of the survey, sampling units of the survey |
| **5. Sample size for the overall survey** | Gross/issued sample for the survey as a whole; net/achieved sample for the survey as a whole; overall response rate of the survey, including the re-sponse rate calculation used; issues/challenges with the sample identified by the survey producer; details of the survey weights (if used) |
| **6. Sample sizes for any subgroups in which the survey is partitioned** | For each partitioned subgroup of the survey (at least up to five subgroups), gross/issued sample, net/achieved sample, response rate, calculation used for the response rate, issues/challenges with the sample identified by the survey producer |
| **7. Data collection information** | Name of the individual or entity that undertook the survey fieldwork; data collection mode used; for surveys that included personal interviews, specifics on who conducted the interviews and indication of whether the interviewer spoke any of the EMM target population languages and if so, which ones; indication of whether the questionnaire was offered in the languages spoken by the EMM target population and if so, which ones; average duration/length of the interviews; number of questions included in the survey |
| **8. Availability to re-search community** | Availability of the survey dataset, technical documentation, and questionnaire; data archive/repository IDs and DOIs for the survey dataset, technical documentation, and questionnaire; languages in which the survey dataset, technical documentation, and questionnaire are available; data documentation standards used for the survey's technical documentation; details about any access/use restrictions for the survey's dataset |
| **9. Data producers, owners, distributors and citations** | Names of the survey producer, owner, and distributor; contact details for any queries or requests about the survey; recommended citations for the survey dataset, technical documentation, questionnaire, and other related documents |
| **10. Additional information** | Overall quality rating of the survey (for internal use only), comments about the quality rating of the survey, sources of information used to compile the survey's metadata, any other comments about the survey and the metadata compiled |

NOTE: For more detailed information about each section and the information it captures, please consult pages 8-22 of the guidelines document used for the metadata compilation work: FINAL CA16111 ETHMIGSURVEYDATA WG1 WG2 Metadata Template

## D. VERSION CONTROL OF THE SURVEY-LEVEL METADATA

As the metadata schema exists as an Excel-based template, each of the 35 COST Action participating countries (represented by national teams) produced an Excel file (i.e. a metadata template) that documents the survey-level metadata that has been compiled for each of the identified surveys. To ensure that the metadata templates have been properly coded to respect the aforementioned metadata schema, they have been regularly reviewed or version controlled by the WG1 and WG2 leaders, as well as the central team of Ethmigsurveydata, which is based at Sciences Po. All the metadata templates were required to undergo the full version control process, which consists of the following steps (see Image 1 for a visual representation).

The national team of the respective COST Action participating country codes five surveys onto the metadata template and submits it for an initial quality control (i.e. **first stage quality control**).

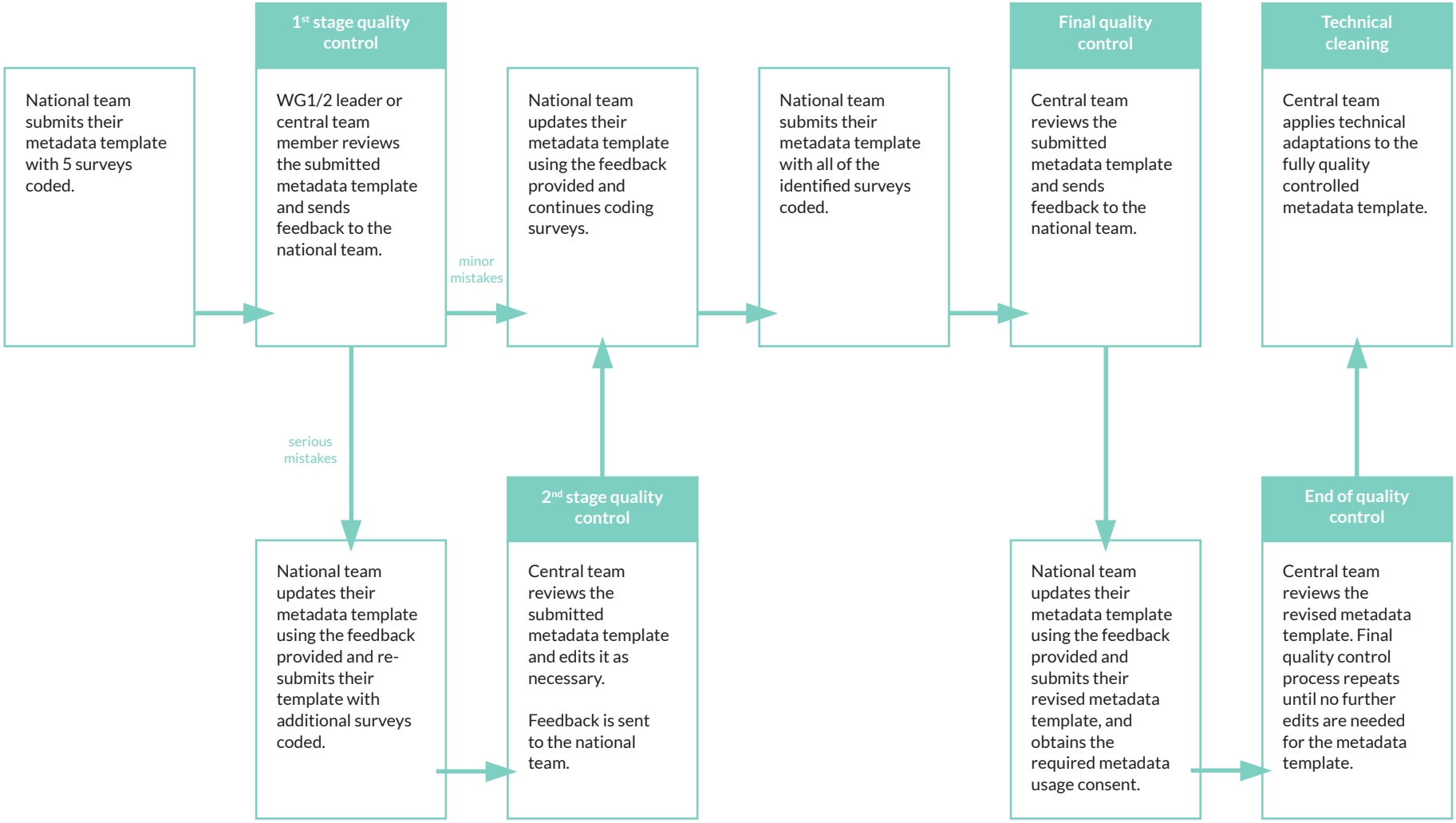
A WG1 or WG2 leader or a central team member completes the first stage quality control by reviewing the submitted metadata template; specifically this means checking that the metadata template has been filled out or used correctly and that the information captured for each survey is coherent and logical. After the review is completed, detailed feedback (via email) is provided to the national team. The target turnaround time is within two to three weeks of the template being submitted for review.

The national team updates their submitted metadata template using the feedback provided. If their template had serious mistakes (which is communicated to them by the reviewer), the national team codes additional surveys onto the metadata template and submit it for a second round of quality control (i.e. **second stage quality control**).

The national team finishes coding all of the surveys found and submits the final version of their metadata template for review (i.e. **final quality control**).

A central team member conducts the final quality control of the submitted final metadata template by carefully reviewing that the metadata template; similarly to the first stage quality control, the reviewer verifies that the metadata template has been filled out and used correctly and that the each survey record has coherent and logical information documented. For any straightforward edits/modifications that are spotted during this review process, the central team member handles them themselves; however, for any edits/modifications that require the national teams' input and clarification, the central team member contacts them directly to request their assistance. The central team member also liaises with the national team to identify surveys on the metadata template that have been deposited to a data archives/repositories/registries/catalogues in order to obtain consent (for proprietary and copyright reasons) to store and reuse the survey-level metadata that they may have produced. Once the metadata template has completed the final quality control process, including the steps where the national team is involved, the metadata template can progress to the technical cleaning stage.

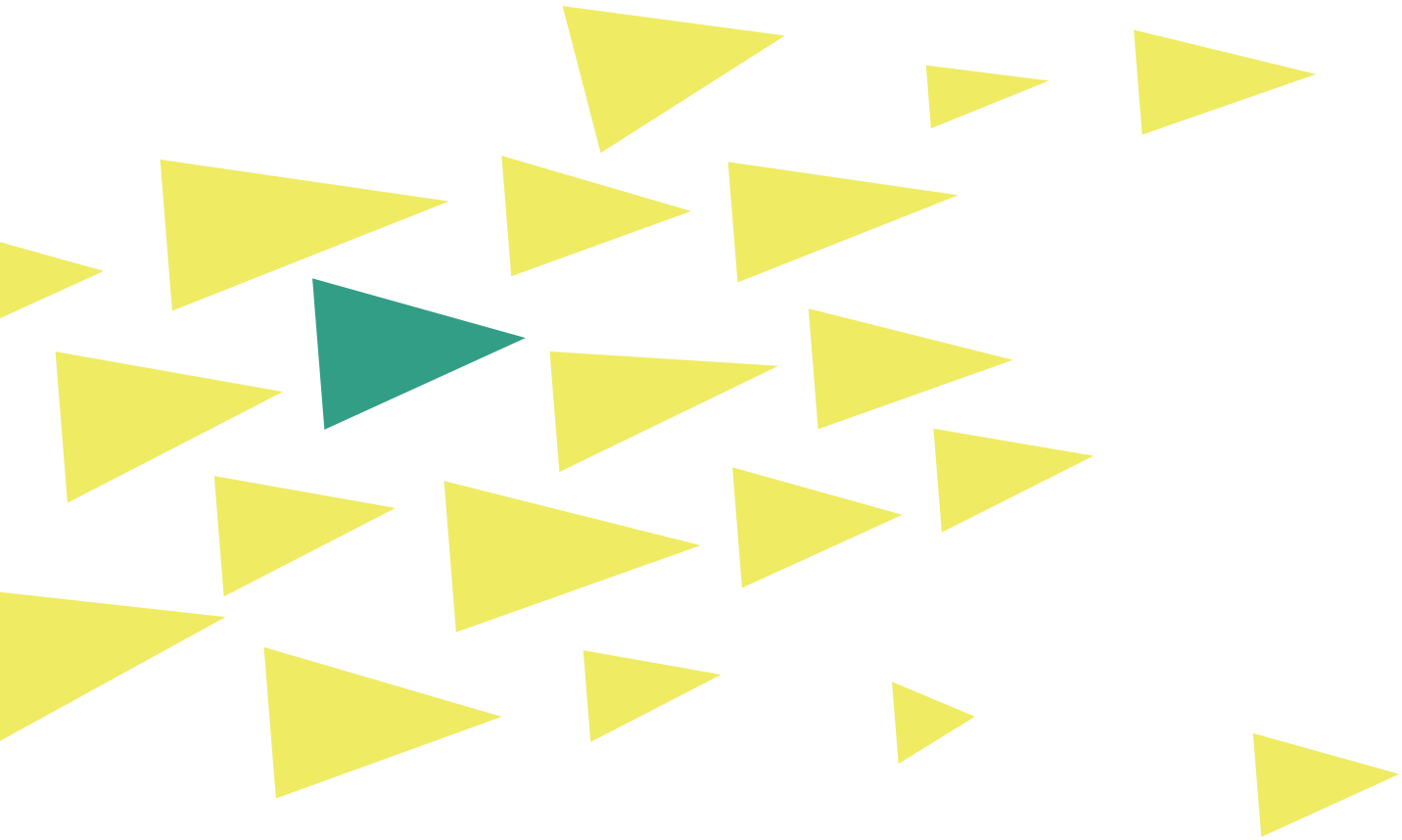
**Image 1: Version control process for the metadata templates**



The final step of the version control process is the **technical cleaning stage**. A central team member prepares the fully quality controlled metadata template for data analysis and for uploading onto the EMM Survey Registry by performing necessary technical adaptation steps (i.e. re-formatting the template, post-coding/processing certain metadata variables, and performing any other data cleaning steps in order to make the template fully readable and usable for both data analysis and the EMM Survey Registry).

A detailed guidelines document was also produced to describe and define the version control process that was utilised for the metadata templates. The central team was responsible for guiding each of the 35 COST Action participating countries through each of the above steps. If any country was delayed or had difficulties progressing through the version control process, the central team would follow up and liaise with the national team of the country-in-question.

The metadata for the six countries included in this version of the report have, hence, undergone exhaustive quality control processes to ensure - as much as is possible - the coherence and reliability of the information provided.



# 4

# GEOGRAPHICAL AND TARGET GROUP COVERAGE OF THE EMM SURVEYS IN EUROPE

When researching the impact of surveys as well as how they can be utilized to describe a given national research landscape and the information that can be derived from them it is necessary to get a detailed look at the scope and type of research that is done. Additionally, it is of benefit to find out if research is done sporadically or at regular intervals, or if it is organised in smaller research clusters or integrated in an international research landscape. Thus, this section of the report will cover the territorial scope of the surveys and their research agenda.

As established in the previous sections the current data set includes six countries that show a very different distribution when it comes to registered surveys. Those differences can be attributed to various factors. In some instances, the differences may be due to country size, e.g. Germany has approximately ten times the population of Switzerland, in others it may be due to political and structural differences, for example between Northern and Eastern European countries. Furthermore, there may be problems to uncover the necessary information in specific countries, as data management procedures may make it hard to find out if there is even any research in a given area. Making the field of ethnic and migrant minorities (*EMM*) not only hard to do research in, but also one hard to do research on.

However, the existing material makes it possible to highlight differences in the prevalence of identifiable ethnic and migration minority related studies in different countries. The first of such differences are the different priorities, when it comes to national and subnational studies. Altogether half (50%) of the projects have a national scope and the other half (50%) have a subnational reach. When comparing across countries, this divide applies for Switzerland, Germany and Croatia as well. In those countries a nearly equal mix of national and subnational surveys have been identified. In contrast, the registered data for Norway and Romania skews towards national programs, while subnational surveys are dominant only in Turkey (Table 2). However, as stated before, this result should be interpreted carefully, as it is based on reports of datasets and projects that could be explored by the project, thus it may provide a selective picture regarding the survey landscape.

**Table 2. Territorial scope (percentages)**

| Territorial scope of survey | Country | | | | | | |
|---|---|---|---|---|---|---|---|
| | Switzerland (CH) | Germany (DE) | Croatia (HR) | Norway (NO) | Romania (RO) | Turkey (TR) | Total |
| National | 55.6% | 47.9% | 48.6% | 88.9% | 100.0% | 25.0% | **50.3%** |
| Subnational | 44.4% | 52.1% | 51.4% | 11.1% | 0.0% | 75.0% | **49.7%** |
| Total | 9 | 73 | 37 | 9 | 9 | 24 | 161 |

Furthermore, besides the significant differences between the shares of national and subnational surveys across countries, there are also differences in terms of the territorial scope of the surveys at subnational level. While the subnational level surveys of Germany and Croatia cover several geographical levels, these were predominantly focusing on cities and urban centres in Switzerland, Norway and Turkey (Table 3).

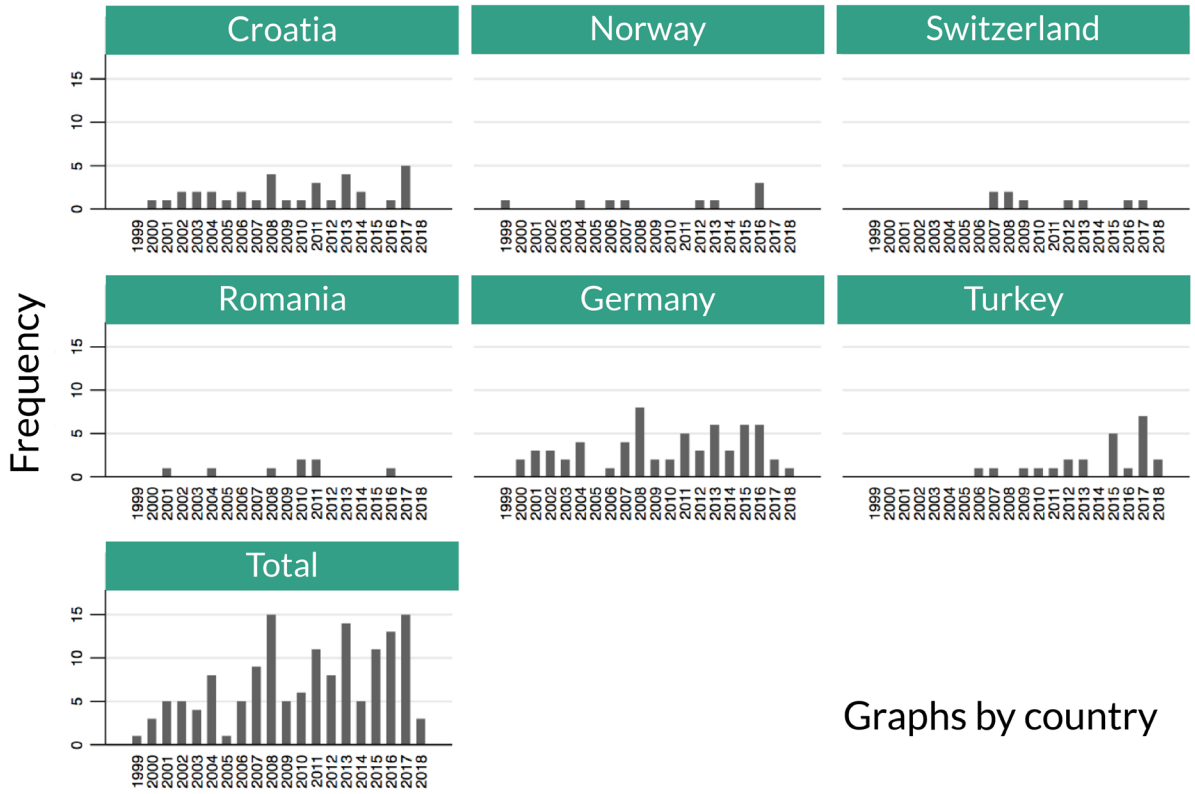
**Table 3. Subnational surveys (percentages)**

| Territorial scope of subnational surveys | Country | | | | | | |
|---|---|---|---|---|---|---|---|
| | Switzerland (CH) | Germany (DE) | Croatia (HR) | Norway (NO) | Romania (RO) | Turkey (TR) | Total |
| Predominantly urban / Cities (densely populated areas) | 100.0% | 44.7% | 15.8% | 100.0% | | 88.9% | **51.3%** |
| Intermediate / Towns and suburbs (intermediate density areas | 0.0% | 15.8% | 36.8% | 0.0% | | 0.0% | **16.3%** |
| Predominantly rural / Rural areas (thinly populated areas) | 0.0% | 0.0% | 21.1% | 0.0% | | 0.0% | **5.0%** |
| Mix (more than one subnational area type) | 0.0% | 39.5% | 26.3% | 0.0% | | 11.1% | **27.5%** |
| Total | 4 | 38 | 19 | 1 | 0 | 18 | 80 |

Most of the registered surveys were carried out only in one country at once. Only 1 in 5 surveys (31 of the 161 studies) are part of an international collaboration. More than half of these cross-country surveys were identified in Germany, which corresponds to 1 in 4 German surveys (18 surveys joined an international program out of the 73 German surveys identified). Besides Germany, Romania and Switzerland also appear to be active in international EMM survey collaborations (6 out of 9, and 5 out of 9 respectively), with more than half of the EMM surveys identified in these countries being inserted in such cross-national collaborations. The absence of international surveys in Norway and Turkey might be related to the fact that these are non-EU countries and cross-country survey projects are often undertaken through EU projects or initiatives in Europe, and the same might apply to the newest EU member state, Croatia, which had also less chances to join EU-wide surveys prior to its accession. Furthermore, there are no surveys that are part of an international program since 2015 for these six countries.

When we focus on the timing of the surveys using as a reference the year when they were completed, the distribution suggests that the number of surveys undertaken is slightly increasing over time, but not enormously. Across the six countries studied the overtime trend is highly uneven and no clear trend can be identified. Additionally, 14 surveys could not be dated. The highest number of surveys could be identified in 2008 and 2017, with a considerable number of surveys undertaken between 2015 and 2017 – following the so called "refugee crisis". Most of those are in two highly affected countries, Germany and Turkey. However, the German studies of 2015-2017 are dominantly focusing on the national level, while all the Turkish studies were carried out at subnational level, which is in line with actual territorial impact of the migration flow in these countries (see Figure 1).

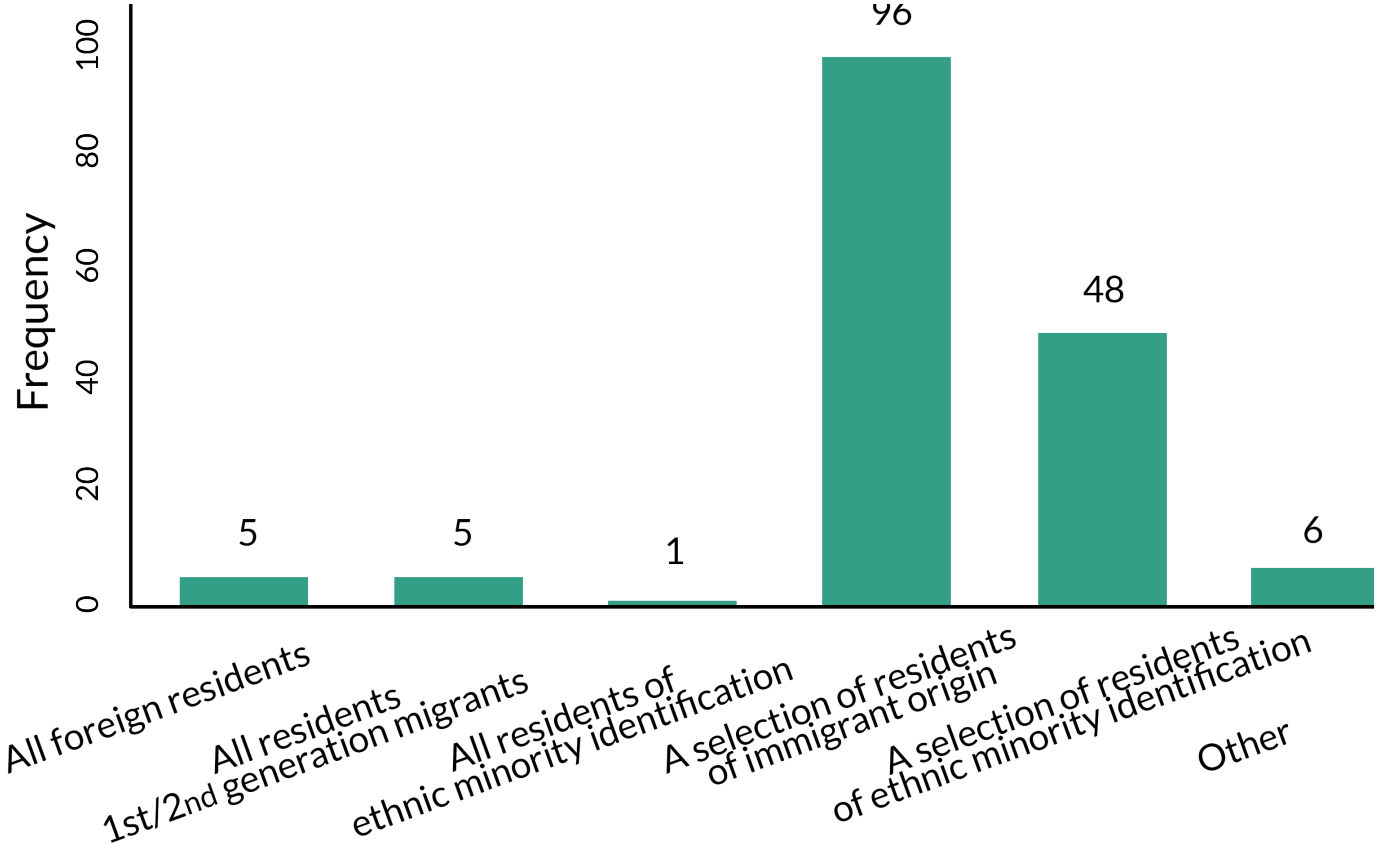
**Figure 1. Registered Surveys (n, per country/year)**



If we turn our attention to the coverage of the target population groups (Figure 2), the EMM Survey Registry detected two times more surveys to migrant minorities than to ethnic minorities for these six countries. Most EMM surveys (60%) identified in the six countries covered the target group by focusing on a selection of residents of foreign or immigrant origin or ancestry in the city, or region, or country. This means that instead of covering the overall population of residents of immigrant or foreign origin or ancestry, they were focusing on a reduced set of groups (for example, focusing on specific origins or focusing only on those who were already citizens). The opposite approach of covering all foreign residents or all 1st or 2nd generation migrants was only adopted by 6% of the surveys (3% and 3% respectively). The other widely used approach (30%) aiming at covering ethnic minorities as the target group also focused on a selection of residents of ethnic minority identification in the city, or region, or country. Only less than 1% of the surveys targeted all residents with an ethnic minority identification. Other approaches to the coverage of target groups are used rarely: only 4% of the surveys targeted the EMM population through other approaches, for example by focusing on returning migrants.

The Swiss and Norwegian surveys are exclusively covering the migrant minority population (primarily through focusing on a selection of residents of foreign or immigrant origin or ancestry), and this was the dominant approach also in the German (95%) and Turkish (62%) surveys. By contrast, all the Romanian surveys targeted ethnic minorities, and the same applies to the majority (76%) of the Croatian surveys.

**Figure 2. Coverage of the various types of target population (n=161)**



Finally, Table 4 allows us to assess the specific purpose for which the surveys were produced. In most countries, surveys targeting the EMM population are first and foremost undertaken to respond to research (primarily, academic) needs. Nevertheless, a majority of them are also designed to respond to public policy design needs, and this is actually the first purpose that the surveys undertaken in Romania pursue. Generating knowledge for NGOs or non-profit organisations is also a common goal, particularly in Turkey. Only in Norway are these surveys fulfilling a major role in the generation of government statistics, but they occasionally take on this role also in Germany and Turkey. In sum, public (i.e. academic research, policy or governmental statistics) surveys are dominant in Germany and Switzerland and hold an exclusive position in Norway. By contrast, these countries provide no or just very few examples of surveys to the EMM population addressing the research and information needs of the non-profit sector. Although numerous NGO-driven surveys to EMM populations were identified in Turkey, Romania and Croatia, these are the least dominant types of surveys in terms of research purpose also in these countries.

**Table 4. Type of the survey by research purpose (multiple answers)**

| | Country | | | | | | |
|---|---|---|---|---|---|---|---|
| | Croatia | Norway | Switzerland | Romania | Germany | Turkey | Total |
| **Research/academic** | 70.0% | 100.0% | 100.0% | 56.0% | 77.0% | 83.0% | **78.0%** |
| **Public policy** | 51.0% | 78.0% | 0.0% | 100.0% | 44.0% | 62.0% | **51.0%** |
| **Government statistics** | 0.0% | 89.0% | 0.0% | 0.0% | 29.0% | 17.0% | **20.0%** |
| **NGO/non-profit organisations** | 27.0% | 0.0% | 0.0% | 33.0% | 5.0% | 42.0% | **17.0%** |
| **Commercial** | 0.0% | 0.0% | 0.0% | 0.0% | 1.0% | 0.0% | **1.0%** |
| **Other** | 0.0% | 0.0% | 0.0% | 0.0% | 1.0% | 0.0% | **1.0%** |
| **Total N** | **37** | **9** | **9** | **9** | **73** | **24** | **161** |

# 5

# THE DEFINITION AND MEASUREMENT OF EMMS IN SURVEYS ACROSS EUROPE

The definition and empirical measurement of the groups of individuals that may be said to belong to ethnic and migrant minority groups is a matter of intense contestation, both within academic fields and social and political arenas. Within the academic debate in the social sciences, a wide range of approaches to the definition of ethnicity have been proposed and there is no consensus around a preferred one[1]. There is, equally, a wide range of ways to define who is a migrant or of migrant background, and they are equally contested[2]. In fact, research has shown that defining, naming and measuring minorities - whether ethnic or of migrant origin -- can also have a mixed range of stigmatising as well as empowering effects[3]. As a consequence, there is no single approach to the definition and measurement of EMMs as target respondents in surveys across Europe, and this section will focus on describing the wide and varied range of ways in which this is done in practice.

The survey metadata that the EMM Survey Registry collects in this regard is very rich. A first variable (3.1) asks for an open-ended free description of the EMM target population for the survey, where the groups that the survey intended to study can be described - for example, "1st and 2nd generation immigrants who are Norwegian citizens; Foreign nationals with voting

1 See, for example, Barthes, F. (1969). *Ethnic groups and boundaries*, Boston: Little Brown; Gans, H. (1979) "Symbolic ethnicity: the future of ethnic groups and cultures in America", *Ethnic and Racial Studies*, vol. 2: 1-20; Isajiw, W.W. (1993). "Definitions and dimensions of ethnicity: A theoretical framework", in *Statistics Canada and U.S. Bureau of the Census, Challenges of measuring and ethnic world: Science, politics and reality. Proceedings of the joint Canada-United States Conference on the measurement of ethnicity, April 1-3, 1992*. U.S. Government Printing Office: Washington, D.C., pp. 407-427. Smith, K. (2002). "Some Critical Observations on the Use of the Concept of 'Ethnicity' in Modood et al., Ethnic Minorities in Britain", *Sociology*, vol. 36(2): 399-417. Modood, T., Berthoud, R., & Nazroo, J. (2002). "Race, racism and ethnicity: A response to Ken Smith", *Sociology*, vol. 36(2): 419-427.

2 See, e.g., Anderson, B., & Blinder, S. (2011). "Who counts as a migrant? Definitions and their consequences." Briefing, The Migration Observatory at the University of Oxford. Anderson, B. (2017). "Towards a new politics of migration?", *Ethnic and Racial Studies*, vol. 40(9): 1527-1537. Crawley. H., & Skleparis, D. (2018). "Refugees, migrants, neither, both: categorical fetishism and the politics of bounding in Europe's 'migration crisis", *Journal of Ethnic and Migration Studies*, vol. 44(1): 48-64.

3 See, among others, Lucassen, L. (1991). "The power of definition. Stigmatisation, minoritisation and ethnicity illustrated by the history of gypsies in the Netherlands", *Netherlands Journal of Social Sciences*, vol. 27: 80-91. Simon, P. (2008). "The Choice of Ignorance. The Debate on Ethnic and Racial Statistics in France", *French Politics, Culture & Society*, vol. 26(1): 7-31.

rights". A second variable (3.1a) extracts categories from the previous open-ended responses into standardised terms that allow for comparisons across surveys and across countries - in our example, this would be "First generation migrants, Second generation migrants, Foreign citizens, Foreign nationals with voting rights, Naturalized citizens". A third variable (3.2, already described in section 4 of this report) asks for a classification of the EMM target population into seven categories that distinguish between types of target groups (foreigners, migrant-origin population, ethnic minority population) and whether the whole or a subset of the relevant population was targeted - in our Norwegian example, this would be only a selection of the foreign/immigrant origin population, as it is only applicable to those with voting rights. A fourth variable (3.3) details which characteristics of the respondents, their ancestors or the geographical area were taken into account in the sample design or questionnaire screening instruments to operationalize the target population as a specific EMM group or set of groups - for example, the country of birth of the respondents and/or the parents, the citizenship at birth of the respondents and/or the parents, the classification by a third agent or by proxy, etc. A fifth variable (3.4) additionally asks which of these characteristics of the respondents, their ancestors or the area were specifically measured through the questionnaire instrument, so that they can be part of the survey data analyses - unlike aspects that are only used for sampling design. Finally, another question reports whether the survey includes a subgroup of the majority population in the sample.

In this section we describe in detail the rich and heterogeneous picture that this array of variables portrays of the way EMMs are defined, categorised and measured in surveys across Europe through the analysis of each of these metadata variables in turns.

## A. CATEGORISATION OF EMM GROUPS THROUGH OPEN-ENDED QUESTIONS

Across the 161 surveys currently included in the EMM Survey registry, we have identified 63 different categories of groups through variable 3.1 (Table 5). The most frequent among these tend to be generic categories, such as 'migrants', 'ethnic minorities', 'refugees' or 'foreigners', but we see that some specific categories referring to certain countries of origin or specific ethnic minorities are quite common, such as 'Turkish', 'Roma', 'Serbian', 'Italian' or Syrian.

The specific origin or ethnic minorities are, however, quite patterned across the six countries studied (Table 6), as certain EMMs are much more studied in some countries than in others. For example, the Turkish-origin migrants are the most studied EMM group in Germany and Switzerland, whereas the Roma minority is among the most studied in Croatia and Romania, and Syrian refugees dominate the survey landscape in Turkey. By contrast, more generic categorisations are the most common in Norwegian surveys, which tend to focus less on very specific migrant or ethnic minorities and are widening the scope to larger EMM groups such as first and second generation migrants, foreign citizens or naturalized citizens.

## Table 5. Categories used for the target population (variable 3.1), all surveys

| | Total | |
|---|---|---|
| | No. | % |
| Turkish | 47 | 29.2 |
| Migrants | 44 | 27.3 |
| Ethnic minorities | 23 | 14.3 |
| Refugees | 23 | 14.3 |
| Roma | 21 | 13 |
| Serbian | 21 | 13 |
| Foreigners | 19 | 11.8 |
| Italian | 19 | 11.8 |
| Syrian | 18 | 11.2 |
| Former Yugoslavian | 15 | 9.3 |
| Greek | 13 | 8.1 |
| Russian | 13 | 8.1 |
| Spanish | 12 | 7.5 |
| Citizens | 10 | 6.2 |
| Second generation migrants | 10 | 6.2 |
| First generation migrants | 8 | 5 |
| Jewish | 7 | 4.3 |
| Hungarian | 6 | 3.7 |
| Kurdish | 6 | 3.7 |
| Naturalized | 6 | 3.7 |
| Pakistani | 6 | 3.7 |
| African | 5 | 3.1 |
| Croat | 5 | 3.1 |
| German | 5 | 3.1 |
| Muslim | 5 | 3.1 |
| Polish | 5 | 3.1 |
| Albanian | 4 | 2.5 |
| Bulgarian | 4 | 2.5 |
| Repatriates | 4 | 2.5 |
| Arab | 3 | 1.9 |
| Bosnian | 3 | 1.9 |
| Czech | 3 | 1.9 |
| Kosovar | 3 | 1.9 |
| Migrant youth | 3 | 1.9 |
| Moroccan | 3 | 1.9 |
| Armenian | 2 | 1.2 |
| Asylum seekers | 2 | 1.2 |
| Christian | 2 | 1.2 |
| Ethnic German | 2 | 1.2 |
| Majority population | 2 | 1.2 |
| Migrant parents | 2 | 1.2 |
| Migrants with voting rights | 2 | 1.2 |
| Non-western migrants | 2 | 1.2 |
| Residents of multi-ethnic are-as | 2 | 1.2 |
| Zaza | 2 | 1.2 |
| Adult migrants | 1 | 0.6 |
| Afghans | 1 | 0.6 |
| Alevis | 1 | 0.6 |
| Azerbaijani | 1 | 0.6 |
| Bayash | 1 | 0.6 |
| Family reunification | 1 | 0.6 |
| Former USSR | 1 | 0.6 |
| High skilled migrants | 1 | 0.6 |
| Iraqi | 1 | 0.6 |
| Kazakh | 1 | 0.6 |
| Maghrebi | 1 | 0.6 |
| Migrant spouses | 1 | 0.6 |
| Montenegrin | 1 | 0.6 |
| Slovak | 1 | 0.6 |
| Slovenian | 1 | 0.6 |
| Students | 1 | 0.6 |
| Sub-Saharan African | 1 | 0.6 |
| Religious minorities | 1 | 0.6 |
| **Total** | **161** | **100.0** |

**Table 6. Top-5 categories of target population (variable 3.1), by country**

| | 1st | | 2nd | | 3rd | | 4th | | 5th | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Category | N | Category | N | Category | N | Category | N | Category | N | N |
| **Croatia** | Ethnic minorities | 18 | Serbian | 17 | Roma | 7 | Albanian | 4 | Italian / Refugees | 3 | **37** |
| **Norway** | 2nd generation migrants | 5 | 1st generation migrants | 4 | Foreigners | 4 | Naturalized citizens | 3 | Citizens | 3 | **9** |
| **Switzerland** | Turkish | 5 | Former Yugoslavian | 4 | Citizens | 3 | Italian | 3 | Kosovar | 3 | **9** |
| **Romania** | Roma | 9 | | | | | | | | | **9** |
| **Germany** | Turkish | 40 | Migrants | 34 | Greek | 13 | Italian | 13 | Russian | 13 | **73** |
| **Turkey** | Refugees | 14 | Syrian | 14 | Kurdish | 6 | Migrants | 5 | Ethnic minorities | 5 | **24** |
| **Total** | **Turkish** | **47** | **Migrants** | **44** | **Ethnic minorities** | **23** | **Refugees** | **23** | **Roma** | **21** | **161** |

This leads us to examine the various ways of categorising EMMs around five major types: by origin (country or world region, e.g. Greek, African, etc.), by religion (e.g. Muslim, Jewish, etc.), by legal or administrative status (e.g. refugee, naturalized citizens, etc.), by ethnicity (e.g. Sub-Saharan African, Roma, Arab, etc.), or by socio-economic status (e.g. highly-skilled migrants, adult migrants, migrant spouses, etc.). These types have been generated by us through the aggregation of the original open-ended categories recorded in variable 3.1. Table 7 shows the distribution by country and suggests that, overall, categorisations by origin and by legal or administrative status are the most common across most countries but with some exceptions in the Eastern European countries, where categorisations by ethnicity are the most frequent.

**Table 7. Types of categorisation of target population (recoded from 3.1), by country**

| | | Croatia | Norway | Switzerland | Romania | Germany | Turkey | Total |
|---|---|---|---|---|---|---|---|---|
| Origin | % | 64.9 | 33.3 | 88.9 | 0.0 | 56.2 | 75 | **58.4** |
| | No. | 24 | 3 | 8 | 0 | 41 | 18 | **94** |
| Religion | % | 0.0 | 0.0 | 22.2 | 0.0 | 12.3 | 8.3 | **8.1** |
| | No. | 0 | 0 | 2 | 0 | 9 | 2 | **13** |
| Legal /ad-ministrative status | % | 16.2 | 55.6 | 44.4 | 0.0 | 76.7 | 79.2 | **55.9** |
| | No. | 6 | 5 | 4 | 0 | 56 | 19 | **90** |
| Ethnicity | % | 73 | 0.0 | 0.0 | 100 | 9.6 | 29.2 | **31.1** |
| | No. | 27 | 0 | 0 | 9 | 7 | 7 | **50** |
| Socio--economic | % | 8.1 | 55.6 | 44.4 | 0.0 | 9.6 | 0.0 | **11.8** |
| | No. | 3 | 5 | 4 | 0 | 7 | 0 | **19** |
| Total | % | **100** | **100** | **100** | **100** | **100** | **100** | **100** |
| | No. | **37** | **9** | **9** | **9** | **73** | **24** | **161** |

## B. COVERAGE OF THE TARGET POPULATION IN TERMS OF AGE AND SEX

For the 161 surveys identified in the six countries, the majority (60%) had strictly adult (18 years old and older OR 15 years old and older) respondents, 23% included a combination of minor and adult respondents, 11% had youth (between 13 to 25 years old) respondents, and 6% did not identify or provide information about the age of the respondents. This rank order observed across the six countries was also observed within each country as shown in Table 8 below.

**Table 8. Coverage of the target population in terms of age, by country (frequencies)**

| | Coverage of the target population in terms of age (frequencies) | | | | |
|---|---|---|---|---|---|
| Country | Adult (18+ or 15+) | Combination (minors and adults) | Youth (13-25 years old only) | Not identified / provided | Total |
| **Croatia** | 23 | 10 | 4 | 0 | **37** |
| **Germany** | 42 | 16 | 11 | 4 | **73** |
| **Norway** | 5 | 3 | 0 | 1 | **9** |
| **Romania** | 8 | 8 | 0 | 1 | **9** |
| **Switzerland** | 9 | 0 | 0 | 0 | **18** |
| **Turkey** | 9 | 9 | 3 | 4 | **24** |
| **Total** | **96** | **37** | **18** | **10** | **161** |

The age coverage of the survey can also be analysed in terms of the EMM categories identified in section A above (i.e. origin, religion, legal / administrative status, ethnicity, socio-economic status). For all three age coverage classifications, the EMM target population was most commonly identified by origin or legal / administrative status (Table 9). Interestingly none of the youth only surveys used religion to identify the EMM target population.

**Table 9. Coverage of the target population in terms of age, by EMM categories (frequencies and percentages)**

| EMM categories | Coverage of the target population in terms of age | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Youth (13-25 years old only) | | Adult (18+ or 15+) | | Combination (minors and adults) | | Not identified / provided | | Total | |
| | No. | % | No. | % | No. | % | No. | % | No. | % |
| **Origin** | | | | | | | | | | |
| **Yes** | 9 | 50.0 | 51 | 53.1 | 27 | 73.0 | 7 | 70.0 | **94** | **58.4** |
| **No** | 9 | 50.0 | 45 | 46.9 | 10 | 27.0 | 3 | 30.0 | **67** | **41.6** |
| **Total** | **18** | **100.0** | **96** | **100.0** | **37** | **100.0** | **10** | **100.0** | **161** | **100.0** |
| **Religion** | | | | | | | | | | |
| **Yes** | 0 | 0.0 | 6 | 6.2 | 6 | 16.2 | 1 | 10.0 | **13** | **8.1** |
| **No** | 18 | 100.0 | 90 | 93.8 | 31 | 83.8 | 9 | 90.0 | **148** | **91.9** |
| **Total** | **18** | **100.0** | **96** | **100.0** | **37** | **100.0** | **10** | **100.0** | **161** | **100.0** |
| **Legal / administrative status** | | | | | | | | | | |
| **Yes** | 11 | 61.1 | 53 | 55.2 | 19 | 51.4 | 7 | 70.0 | **90** | **55.9** |
| **No** | 7 | 38.9 | 43 | 44.8 | 18 | 48.6 | 3 | 30.0 | **71** | **44.1** |
| **Total** | **18** | **100.0** | **96** | **100.0** | **37** | **100.0** | **10** | **100.0** | **161** | **100.0** |
| **Ethnicity** | | | | | | | | | | |
| **Yes** | 1 | 5.6 | 29 | 30.2 | 14 | 37.8 | 6 | 60.0 | **50** | **31.1** |
| **No** | 17 | 94.4 | 67 | 69.8 | 23 | 62.2 | 4 | 40.0 | **111** | **68.9** |
| **Total** | **18** | **100.0** | **96** | **100.0** | **37** | **100.0** | **10** | **100.0** | **161** | **100.0** |
| **Socio-economic status** | | | | | | | | | | |
| **Yes** | 2 | 11.1 | 10 | 10.4 | 6 | 16.2 | 1 | 10.0 | **19** | **11.8** |
| **No** | 16 | 88.9 | 86 | 89.6 | 31 | 83.8 | 9 | 90.0 | **142** | **88.2** |
| **Total** | **18** | **100.0** | **96** | **100.0** | **37** | **100.0** | **10** | **100.0** | **161** | **100.0** |

As for the sex of the respondents, 94% (n=152) of the 161 surveys included both female and male respondents. Only one survey conducted in Turkey had only female respondents and only two surveys conducted in Croatia had only male respondents. Five surveys — three in Germany, one in Norway, and one Turkey — did not identify or provide information about the sex of the respondents of the target population.

In terms of the EMM categories from section A, the EMM target population was most commonly identified by origin or legal / administrative status for surveys to both men and women and to women only. For surveys conducted with strictly male respondents, the EMM target population was only identified in terms of origin. It is also noteworthy that for surveys conducted with women only, the EMM target population was only classified by origin or legal / administrative status.

## C. MIGRANT/MINORITY RELATED QUESTIONS USED FOR SAMPLING AND/OR SCREENING OF TARGET POPULATION

Once we know the target population of a survey, we need to provide an operationalization of such definition in order to be able to select the sample. If a survey is addressed to all migrants residing in a particular region, it is likely that the variable used in order to know who can potentially be included in the sample will record whether respondents were born in a foreign country (in addition to other factors such as setting a lower age limit). Sometimes this information is available in population registers and these can be used as sampling frames, but in many countries this will not be possible and the question is included in a first part of a screening questionnaire used to "build" the sample to adjust to the definition of the target population. In other cases it is the "subjective" nature of the definition of the target population — e.g., "feelings of belonging to a particular minority" — that will require the use of a question in a screening questionnaire.

As was stated above, one of the variables in the metadata database reflects which characteristics of the respondents, their ancestors or the geographical area were taken into account in the sample design or questionnaire screening instruments to operationalize the target population as a specific EMM group or set of groups.

Table 10 shows the criteria used to operationalize the definition of the target group in the 161 surveys examined. The most frequently used criterion refer to factual issues such as the current citizenship/nationality of the respondent (used in 59% out of all the surveys analysed), the respondent's country of birth (32%) and/or the country of birth of the respondent's parents/grandparents (29%). Other factual criteria, such as nationality of the respondent at birth, or the mother language are less widely used (9% and 10% of the surveys examined, respectively). Subjective criteria such as self-identification as part of an ethnic/migrant group is used as a way to operationalize the selected target population in roughly a fifth of all the surveys examined in

the six counties (21% if we consider surveys in which only one self-identification criterion was used together with those in which several of them were applied).

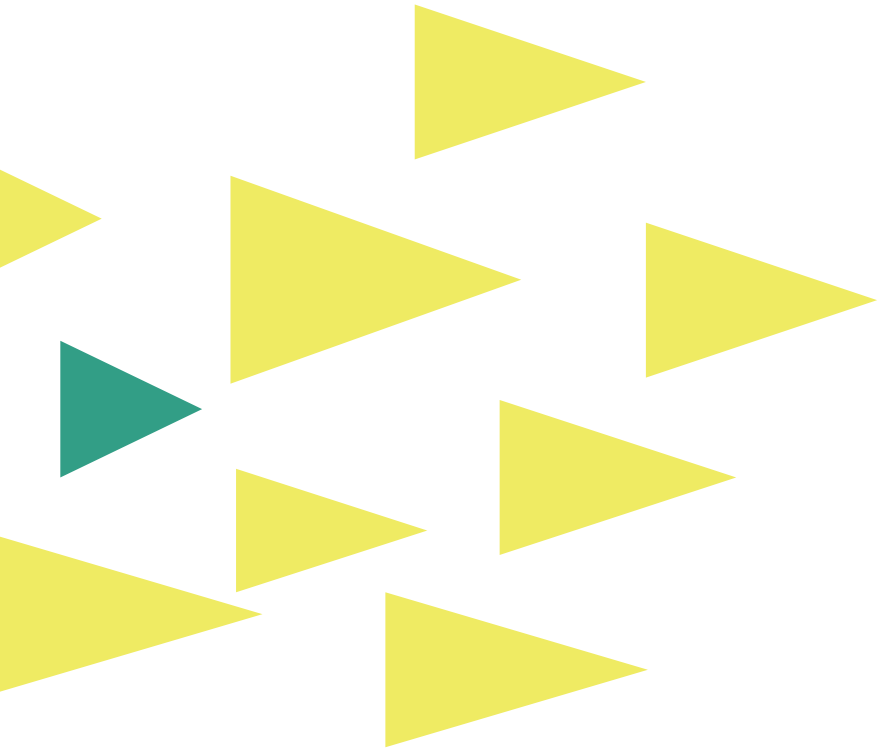
**Table 10. Migrant/minority related criteria used for sampling and/or screening the survey target population (variable 3.3), by country (%)**

| | CH | DE | HR | NO | RO | TR | Total |
|---|---|---|---|---|---|---|---|
| N (total) | 9 | 73 | 37 | 9 | 9 | 24 | 161 |
| Country of birth of respondent | 78.0 | 38.0 | 16.0 | 100.0 | 11.0 | 0.0 | 32.0 |
| Country of birth of parents/grandparents | 33.0 | 43.0 | 5.0 | 100.0 | 11.0 | 0.0 | 29.0 |
| Citizenship/nationality of respondent (current) | 67.0 | 56.0 | 76.0 | 100.0 | 11.0 | 42.0 | 59.0 |
| Citizenship/nationality of respondent (at birth) | 22.0 | 3.0 | 0.0 | 100.0 | 0.0 | 8.0 | 9.0 |
| Citizenship/nationality of parents/grandparents (current) | 0.0 | 8.0 | 3.0 | 100.0 | 0.0 | 0.0 | 10.0 |
| Citizenship/nationality of parents/grandparents (at birth) | 0.0 | | 0.0 | 89.0 | 0.0 | 0.0 | 6.0 |
| Ethnic self-identification of respondent (one response allowed) | 33.0 | 1.0 | 27.0 | 0.0 | 78.0 | 29.0 | 17.0 |
| Ethnic self-identification of respondent (multiple responses allowed) | 0.0 | 3.0 | 0.0 | 0.0 | 22.0 | 8.0 | 4.0 |
| Ethnic self-identification of parents/grandparents | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Mother tongue/language related question | 0.0 | 6.0 | 16.0 | 0.0 | 33.0 | 13.0 | 10.0 |
| Classification by interviewer | 0.0 | 0.0 | 24.0 | 0.0 | 11.0 | 0.0 | 6.0 |
| Classification by third agent/proxy (e.g. by a government authority) | 0.0 | 0.0 | 32.0 | 0.0 | 44.0 | 58.0 | 24.0 |
| Classification by geographical location | 78.0 | 0.0 | 62.0 | 44.0 | 22.0 | 4.0 | 23.0 |
| Through other means/characteristics (including non-ethnic/migrant minority) | 56.0 | 15.0 | 16.0 | 0.0 | 0.0 | 0.0 | 14.0 |
| Information not available on definition of target population | 0.0 | 1.0 | 5.0 | 0.0 | 11.0 | 4.0 | 3.0 |

Nationality and country of birth of the respondent are the most widely used criterion to operationalize the target population in all of the countries analysed, except for Romania. All of the surveys addressed to EMMs in Romania included in the dataset incorporated some sort of ethnic self-identification question, while respondents' nationality/country of birth was much less used. This is not surprising since most of the surveys from Romania included in this metadata file had the Roma minority as their target population, and it is a fairly common practice in those surveys to rely on ethnic self-identification.

Relying on the interviewers to select the respondents that comply with the definition of the target population is rare (only in 6% out of the 161 surveys). However, relying on external agents is more common, though this might have been interpreted as relying on sources external to the questionnaire that are sometimes also used as sampling frames, such as registers. Using other criteria for the operationalization of the target definition is not common, except in the case of Switzerland. When used, these criteria include reference to the migrant status (or residential status), year of arrival to the country, having experienced family reunification, religion or the type of school attended.

Geographical considerations are also used to operationalize the target population in a considerable number of surveys (around a fourth of the 161 analysed). The percentage that apply this criterion is very similar in subnational and national surveys (25% and 21%, respectively), so it is probably related to the sampling design.



## D. MIGRANT/MINORITY RELATED QUESTIONS INCLUDED IN THE QUESTIONNAIRE INSTRUMENTS

Migrant or minority related questions applied by the surveys in their data collection instruments were also explored in detail besides the operalization of target groups for the sample design and the pre-screening (Table 11). The most frequently used approach addressed factual questions to the respondent: half of the surveys identified in the six countries applied questions on the current nationality of the respondent (54%), on the mother tongue/language (50%) and/or on the country of birth of the respondent (47%), while nationality of the respondent at birth (17%) was asked less widely. Besides the factual questions, subjective questions such as the ethnic self-identification of the respondent were used less frequently (whether those that allow one response: 21% or multiple responses: 20%). Factual questions are often used in surveys to migrant minorities where the country of birth, nationality or mother tongue is much more relevant than in surveys targeting ethnic minorities, mainly the native Roma citizens of a country who were also born there, and thus the latter surveys are often relying on the subjective identification approach by asking self-identification.

The nationality of the respondents and the country of birth of the respondents and of the parents were included in all the Norwegien EMM surveys identified in the EMM Survey Registry, as well as in most Swiss surveys (100%, 89% and 67% respectively), and in more than half of the German surveys (63%, 56% and 55% respectively). The mother tongue as a migrant/minority indicator has been widely used in Switzerland, Germany and Turkey (in 89%, 68% and 63% of the surveys in the respective countries).

Questions on the parents' generation are mostly asked in relation to their country of birth (40%), while the current nationality of the parents (15%) or their nationality at birth (9%) are asked less commonly. Ethnic self-identification of the parents was asked only in 4% of the surveys (mainly in Romania: 11% and Germany: 6%), while no surveys included such questions with relation to the grandparents. Other questions targeting the grandparents generation were anyway rare.

External identification of migrant/minority background through classification by the interviewers is also rarely practiced, roughly 1 in 8 to 10 surveys applied this strategy in Croatia, Romania and Turkey. Information is not available on migrant/minority related questions in 12% of the surveys in the six countries, with a substantial variation across countries. There is no information on these questions in 1 in 3 surveys (32%) in Croatia, while it is available in all the Swiss and Norwegian surveys.

**Table 11. Migrant/minority related questions in the survey (variable 3.4.), by country (%)**

| | CH | DE | HR | NO | RO | TR | Total | N |
|---|---|---|---|---|---|---|---|---|
| N (Total) | 9 | 73 | 37 | 9 | 9 | 24 | | 161 |
| Country of birth of respondent | 89.0% | 56.0% | 30.0% | 100.0% | 44.0% | 13.0% | 47.0% | 76 |
| Country of birth of parents | 67.0% | 55.0% | 22.0% | 100.0% | 11.0% | 4.0% | 40.0% | 65 |
| Country of birth of grandparents | 0.0% | 10.0% | 0.0% | 0.0% | 0.0% | 4.0% | 5.0% | 8 |
| Nationality of respondent (current) | 100.0% | 63.0% | 43.0% | 100.0% | 33.0% | 17.0% | 54.0% | 87 |
| Nationality of respondent (at birth) | 56.0% | 12.0% | 0.0% | 100.0% | 22.0% | 8.0% | 17.0% | 27 |
| Nationality of parents (current) | 44.0% | 10.0% | 8.0% | 100.0% | 11.0% | 0.0% | 15.0% | 24 |
| Nationality of grandparents (current) | 0.0% | 1.0% | 0.0% | 0.0% | 0.0% | 0.0% | 1.0% | 1 |
| Nationality of parents (at birth) | 0.0% | 5.0% | 0.0% | 100.0% | 0.0% | 4.0% | 9.0% | 14 |
| Nationality of grandparents (at birth) | 22.0% | 1.0% | 0.0% | 33.0% | 0.0% | 4.0% | 4.0% | 7 |
| Ethnic self-identification of respondent (one re-sponse allowed) | 44.0% | 8.0% | 22.9% | 0.0% | 78.0% | 38.0% | 21.0% | 34 |
| Ethnic self-identification of respondent (multiple responses allowed) | 33.0% | 33.0% | 0.0% | 0.0% | 22.0% | 13.0% | 20.0% | 32 |
| Ethnic self-identification of parents | 0.0% | 6.0% | 3.0% | 0.0% | 11.0% | 0.0% | 4.0% | 6 |
| Ethnic self-identification of grandparents | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0.0% | 0 |
| Mother tongue/language related ques-tion | 89.0% | 68.0% | 14.0% | 0.0% | 22.0% | 63.0% | 50.0% | 80 |
| Classification by interviewer | 0.0% | 3.0% | 11.0% | 0.0% | 11.0% | 13.0% | 6.0% | 10 |
| Information not available on migrant/ minority re-lated questions | 0.0% | 8.0% | 32.0% | 0.0% | 11.0% | 4.0% | 12.0% | 20 |

## E. SURVEYS INCLUDING THE MAJORITY POPULATION AS A SUBGROUP

Of all the surveys studied, 43% (n=69) included a subgroup formed by the majority population (Table 12). Relative to the national-level surveys, a greater proportion of the subnational level surveys included a majority population subgroup, as this occured in 52% (n=41) of subnational surveys compared to 35% (n=28) of national surveys. Of the six countries, Switzerland (78%), Croatia (51%), and Germany (44%) had the highest proportion of surveys with a majority subgroup. Collectively, these three countries represented 84% of the surveys that had included a subgroup with the majority population.

**Table 12. National vs. subnational surveys with a majority population subgroup, by country (frequencies)**

| Country and territorial scope (national vs. subnational) | | Survey including a subgroup of the majority population (frequencies) | | | Total |
|---|---|---|---|---|---|
| | | Yes | No | Don't know | |
| **Croatia** | National | 6 | 12 | 0 | **18** |
| | Subnational | 13 | 6 | 0 | **19** |
| **Germany** | National | 11 | 24 | 0 | **35** |
| | Subnational | 21 | 17 | 0 | **38** |
| **Norway** | National | 1 | 7 | 0 | **8** |
| | Subnational | 1 | 0 | 0 | **1** |
| **Romania** | National | 3 | 2 | 4 | **9** |
| | Subnational | 0 | 0 | 0 | **0** |
| **Switzerland** | National | 3 | 2 | 0 | **5** |
| | Subnational | 4 | 0 | 0 | **4** |
| **Turkey** | National | 4 | 2 | 0 | **6** |
| | Subnational | 2 | 16 | 0 | **18** |
| **Total** | | **69** | **88** | **4** | **161** |

Interestingly, the inclusion of a majority population subgroup was, for the most part, not linked to the survey being designed as a general population survey. Of the 161 surveys, just 19 (12%) were classified as a general population survey (Table 13). What this, in turn, means is that over 70% of the surveys that had included a majority subgroup were surveys that had been designed purposefully to target an EMM population.

**Table 13. National vs. subnational surveys designed as a general population survey, by country (frequencies)**

| Country and territorial scope (national vs. subnational) | | Survey designed as a general population survey (frequencies) | | | |
|---|---|---|---|---|---|
| | | Yes | No | Don't know | Total |
| **Croatia** | National | 2 | 16 | 0 | **18** |
| | Subnational | 9 | 10 | 0 | **19** |
| **Germany** | National | 3 | 32 | 0 | **35** |
| | Subnational | 0 | 31 | 7 | **38** |
| **Norway** | National | 0 | 8 | 0 | **8** |
| | Subnational | 0 | 1 | 0 | **1** |
| **Romania** | National | 1 | 7 | 1 | **9** |
| | Subnational | 0 | 0 | 0 | **0** |
| **Switzerland** | National | 0 | 5 | 0 | **5** |
| | Subnational | 0 | 4 | 0 | **4** |
| **Turkey** | National | 4 | 2 | 0 | **6** |
| | Subnational | 0 | 18 | 0 | **18** |
| **Total** | | **19** | **134** | **8** | **161** |

# 6 THE TOPICS COVERED BY EMM SURVEYS IN EUROPE

When dealing with metadata on surveys it is not only interesting to focus on methodological aspects, the way surveys are done and how they define their target population. The metadata also provide the chance to analyse the type of information and the topics covered in these surveys.

To do so there is a section in the EMM Survey Registry metadata database that reflects the main topics examined by the surveys included, using a very detailed classification that covered issues specifically relating to migration such as refugee, asylum seeking, citizenship or naturalisation, and family reunification, but also many others such as education, health, social and political participation, discrimination, gender issues, and so on. This section will provide a first descriptive analysis of this information, offering an overview of the main topics that have been the subject of research through surveys specifically addressed to EMM groups.

It is, however, not straightforward to reach conclusions about the reasons behind the variations in coverage across topics. Which topics get to be the most researched can certainly reflect the issues relating to migrant and ethnic minorities that are considered a priority and have deserved the most attention in a given country, but they may also signal the topics for which information is not available on EMM populations through other sources unless a survey specifically addressed to these groups is carried out (e.g. about social and political attitudes). The prevalence of certain topics can also reflect the specific situation of a particular country regarding other sources to collect information on EMM populations, such as population registers, official statistics or various types of administrative registers regarding labour status, health, education and other social and economic characteristics. It may also reflect the trajectory of a country regarding migration, with countries of recent immigration showing more interest in dealing with general issues, while countries with a long immigration history having a more varied array of topics of interest.

When it comes to the actual topics covered with the current metadata scheme, 29 different topics were proposed and an additional residual "other" category was also provided (Q13)1. This provided a wide range of possible answers and made it possible to identify most of the common trends in EMM studies. Indeed, despite the low number of cases in most of the six included countries some patterns emerge and some conclusions can be drawn.

The most common survey topics were demographic characteristics and behaviors, which were part of more than 80% of all registered studies. More than 60% of the studies tracked educational attainment and trajectories, human capital and skills. Furthermore, more than half of the studies included items on labour market integration, identity related issues, political inclusion and participation, social as well as political attitudes, income and poverty related questions as well as interethnic contact and conflict (see Table 14)

However when we look at the development of topic prevalence over time additional information comes to the foreground (Figure 3). The strongest observation is tied to the topic relating to political inclusion, participation, and social/political attitudes. Items that track this kind of information have dropped in relevance in surveys on EMM populations since the beginning of the century. Once part of more than 70% of the studies, in the most recent period analyzed this topic has only appeared in approximately 40% of the surveys covered.

Furthermore, between 2005 and 2009 issues of identity and interethnic communication as well as conflict gained prominence, highlighting the rising interest in questions of belonging and multiculturalism. This comes as no surprise, as this period is marked by the "war on terror" and the rising tensions around religion and other identity markers.

Both drop in the next period studied – 2010 to 2014 – as interest in economic issues like labour market integration starts to slowly rise. In this period more than 60% of the identified surveys cover the latter topic and after that it continues to grow slightly. However, it seems that the interest in the topics that were most popular in the 2005-2009 period seems wane. Understandably so, as the 2010-2014 period is marked by the aftermath of the financial crisis that hit most of Europe.

By the start of 2015 we see another shift in the research topics that receive most attention in surveys to EMM populations. Not only are economy related topics like labor market integration and poverty-related issues covered more and more, the interest in human capital, skills as well as the educational attainment of ethnic and migrant minorities rises as well, and such topics are included in three quarters of the documented surveys. We also observe a rapid rise in the percentage of surveys that deal with forced migration and asylum seekers. Understandably so, as 2015 saw the beginning of the so-called "refugee crisis" that went on to become a political and social challenge across all of Europe2. While it was a fringe topic before 2015 that was covered in approximately 12% of the surveys, it suddenly became part of around 40% of all the identified surveys.

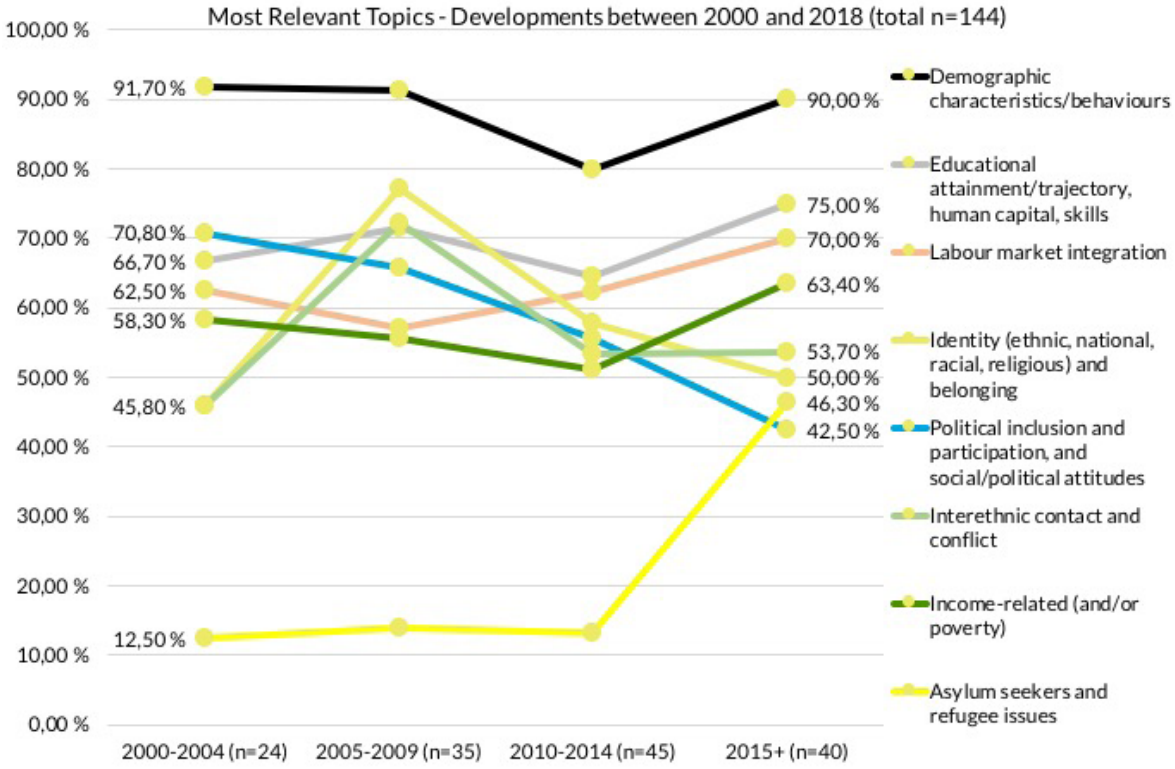
1 The "other" category was eventually used to capture eight different topics: life satisfaction, child care, military service, household assets, freedom of expression, wellbeing, physical violence and victimisation.

2 Hess, S., & Kasparek, B. (2017). "De-and restabilising Schengen. The European border regime after the summer of migration". *Cuadernos Europeos de Deusto*, 56(2017), 47-77.

## Table 14. The topics covered by the surveys by country, rank-ordered

| Topic | Total (n=161) | | Croatia (n=37) | Norway (n=9) | Switzerland (n=9) | Romania (n=9) | Germany (n=73) | Turkey (n=24) |
|---|---|---|---|---|---|---|---|---|
| | Rank | Percentage | | | | | | |
| Demographic characteristics/behaviors | 1 | 81.0% | 84.0% | 100.0% | 100.0% | 22.0% | 80.0% | 92.0% |
| Educational attainment/trajectory, human capital, skills | 2 | 63.0% | 32.0% | 78.0% | 100.0% | 67.0% | 73.0% | 63.0% |
| Labor market integration | 3 | 58.0% | 41.0% | 44.0% | 100.0% | 78.0% | 62.0% | 54.0% |
| Identity (ethnic, national, racial, religious) and belonging | 4 | 53.0% | 57.0% | 89.0% | 89.0% | 33.0% | 47.0% | 50.0% |
| Political inclusion and participation, and social/political attitudes | 5 | 53.0% | 38.0% | 100.0% | 89.0% | 56.0% | 58.0% | 33.0% |
| Income-related (and/or poverty) | 6 | 53.0% | 41.0% | 100.0% | 89.0% | 89.0% | 38.0% | 75.0% |
| Interethnic contact and conflict | 7 | 53.0% | 51.0% | 33.0% | 78.0% | 44.0% | 53.0% | 54.0% |
| Social cohesion and/or civic engagement and/or networks | 8 | 46.0% | 24.0% | 89.0% | 100.0% | 56.0% | 45.0% | 42.0% |
| Family reunification, marriage, family relations | 9 | 45.0% | 11.0% | 44.0% | 100.0% | 22.0% | 59.0% | 46.0% |
| Language skills/training | 10 | 44.0% | 3.0% | 44.0% | 89.0% | 22.0% | 62.0% | 42.0% |
| Discrimination, racism and/or xenophobia | 11 | 42.0% | 38.0% | 44.0% | 89.0% | 67.0% | 37.0% | 38.0% |
| Housing / Access to housing | 12 | 42.2% | 32.0% | 56.0% | 44.0% | 78.0% | 41.0% | 42.0% |
| Religion | 13 | 35.0% | 50.0% | 89.0% | 67.0% | 44.0% | 38.0% | 33.0% |
| Migration trajectory | 14 | 33.0% | 24.0% | 100.0% | 56.0% | 44.0% | 15.0% | 58.0% |
| Health/ Access to health services | 15 | 30.0% | 24.0% | 44.0% | 44.0% | 78.0% | 21.0% | 42.0% |
| Citizenship and naturalization | 16 | 29.0% | 2.70% | 44.0% | 100.0% | 11.0% | 34.0% | 25.0% |
| Consumption and/or leisure | 17 | 29.0% | 11.0% | 33.0% | 22.0% | 11.0% | 47.0% | 8.0% |
| Leisure, sports, arts | 18 | 26.0% | 2.7% | 33.0% | 56.0% | 11.0% | 40.0% | 13.0% |
| Public attitudes about migration and migrant | 19 | 24.0% | 8.0% | 44.0% | 67.0% | 0.0% | 19.0% | 46.0% |
| Reasons for migration/migration drivers | 20 | 22.0% | 14.0% | 33.0% | 56.0% | 22.0% | 15.0% | 42.0% |
| Legal status/administrative situation | 21 | 22.0% | 32.0% | 44.0% | 89.0% | 11.0% | 7.0% | 21.0% |
| Asylum seekers and refugee issues | 22 | 21.0% | 11.0% | 44.0% | 22.0% | 0.0% | 8.0% | 71.0% |
| Transnational patterns (e.g. remittances, travel, engagement with 'home') | 23 | 17.0% | 5.0% | 44.0% | 78.0% | 0.0% | 18.0% | 8.0% |
| Return migration | 24 | 16.0% | 14.0% | 22.0% | 44.0% | 0.0% | 4.0% | 50.0% |
| Space use/spatial consequences | 25 | 8.0% | 5.0% | 44.0% | 44.0% | 0.0% | 3.0% | 0.0% |
| Time use | 26 | 6.0% | 0.0% | 44.0% | 0.0% | 0.0% | 8.0% | 0.0% |
| Gender relations, gender identity, sexuality | 27 | 3.0% | 0.0% | 0.0% | 0.0% | 11.0% | 6.0% | 0.0% |
| Unaccompanied migrant minors | 28 | 1.0% | 0.0% | 0.0% | 11.0% | 0.0% | 0.0% | 4.0% |
| Human smuggling and trafficking | 29 | 1.0% | 0.0% | 0.0% | 0.0% | 0.0% | 1.0% | 0.0% |
| Others | | 22.0% | 3.0% | 0.0% | 0.0% | 11.0% | 43.0% | 8.0% |

**Figure 3. Topics of Interest and Trends**




While Table 14 provides a more detailed overview concerning the relevant topics it has to be noted that a country-by-country and overtime analysis is not possible due to the small number of surveys identified in most of the countries currently covered in the dataset. However, even when comparing the data available between the three western countries (Germany, Switzerland and Norway) to the three eastern ones (Romania, Turkey and Croatia) only some minimal ranking differences emerge, which can be attributed to the reduced number of cases. Nevertheless, it has to be noted that Romania continues to be an outlier, as only two of its surveys cover socio-demographic characteristics and behaviours at length.

# 7 TECHNICAL CHARACTERISTICS OF THE EMM SURVEYS IN EUROPE

A survey is a data collection tool that can be used to systematically gather information and, in turn, glean insights about a population of interest. As surveys are set up to collect information directly from an entity (e.g. an individual) that is part of the "targeted" population, they can often provide a (more) complete and nuanced picture about a population in relation to a specific topic. Moreover, they are flexible in design where they can include a longitudinal element — where information can be collected over time — and/or a comparative component — where information is also gathered from other population groups to allow for meaningful comparisons.

For the study of EMM integration, survey data can therefore serve as an invaluable resource. Not only can surveys be leveraged to generate a wealth of information about EMM populations, they can also offer insights about the multi-faceted, complex, intertwined integration process experienced by EMMs (that is often missed or oversimplified when looking at just static statistics). Thus, as part of our metadata schema, we wanted to adeptly summarize the choices made by the survey producers to make their respective survey(s) an instrument for studying EMM integration. As such, a sizeable number of the metadata variables have been dedicated to documenting the technical characteristics of a survey: how the survey was designed and subsequently implemented by the survey producers. The following are the specific sections and/or variables of our metadata schema that aim to capture the technical characteristics:

- 1.9 Representativeness of the survey (in terms of the target (sub)population)
- 1.10 Type of survey (single cross-section / repeated cross-section / longitudinal or panel / other)
- Section 4 - Sampling method: 4.1 sampling strategy (closed response), 4.2 sampling strategy (open response), 4.3 sample design (full information), 4.4 sampling frame(s), 4.5 sampling unit(s)

- Section 5 - Sample size of the overall survey: 5.1 total gross/issued sample, 5.2 total net/achieved sample, 5.3 overall response rate, 5.4-5.6 response rate calculation used, 5.7-4.8 sampling weights
- Section 6 - Sample sizes for any subgroups in which the survey is partitioned (up to five subgroups): 6a-e1 total gross/issued sample, 6a-e2 total net/achieved sample, 6a-e3 overall response rate, 6a-e4 and 6a-e5 response rate calculation used
- 7.2 Data collection mode (face to face (PAPI), face to face (CAPI), telephone, web / email survey, paper self-administered (collected), paper self administered (postal), other)
- 7.3 Person conducting the personal interview (professional interviewers, cultural mediators, non-professional interviewers, a mix)
- 7.4 Interviewers spoke an EMM language
- 7.5-7.5a EMM languages spoken by the interviewer
- 7.6 Questionnaire in EMM language(s)
- 7.7-7.7a EMM languages in which the questionnaire was made available

In the sections below, we provide a detailed analysis of the technical characteristics of the identified EMM surveys using the sections and variables of the metadata schema listed above.

## A. TYPES OF SURVEYS

The 161 surveys from the six countries were either: single cross-section (n=80), repeated cross-section (n=40), or longitudinal (n=41). In Germany, longitudinal surveys were most common (n=34), whereas single cross-section was most common in Switzerland (n=7), Croatia (n=24), and Turkey (n=20), and repeated cross-section was most common in Norway (n=7) and Romania (n=7).

When examining the surveys by territorial scope, single cross-section was the most prevalent type, representing 38 surveys at the national level and 42 surveys at the subnational level. However, national and subnational surveys differed in terms of their second and third rankings of survey type. National level surveys (after single cross-section) had more repeated cross-section (n=31) than longitudinal (n=12), whereas for subnational level surveys longitudinal (n=29) was more common than repeated cross-section (n=9).

## B. REPRESENTATIVENESS OF THE SURVEYS IN TERMS OF THE TARGET POPULATION

In examining the 161 surveys from the six countries, 45% (n=73) were identified as being representative of the target population of the survey. A detailed overview of the representativeness of the surveys by country can be found in Table 15 below. In Germany and Romania, a similar trend was observed as almost half of their respective surveys (n=33 in Germany and n=4 in

Romania) were deemed as representative. By contrast, all surveys in Switzerland (n=9) were classified as representative, as well as almost all the surveys in Norway (n=8). And in Turkey, 75% of their surveys (n=18) were indicated as not being representative.

**Table 15. Representativeness of the survey, by country (frequencies)**

| Country | Representativeness of the survey (frequencies) | | | |
|---|---|---|---|---|
| | Yes | No | Don't know | Total |
| Croatia | 35.1 | 43.2 | 21.6 | **100%** |
| | 13 | 16 | 8 | **37** |
| Germany | 45.2 | 50.7 | 4.1 | **100%** |
| | 33 | 37 | 3 | **73** |
| Noway | 88.9 | 0.0 | 11.1 | **100%** |
| | 8 | 0 | 1 | **9** |
| Romania | 44.4 | 22.2 | 33.3 | **100%** |
| | 4 | 2 | 3 | **9** |
| Switzerland | 100% | 0.0 | 0.0 | **100%** |
| | 9 | 0 | 0 | **9** |
| Turkey | 25.0 | 75.0 | 0.0 | **100%** |
| | 6 | 18 | 0 | **24** |
| **Total** | **45.4** | **45.4** | **9.3** | **100%** |
| | **73** | **73** | **15** | **161** |

The representativeness of a survey can also be analysed by territorial scope. A greater number of national level surveys were identified as being representative than subnational level surveys. Specifically, 59% of the national level surveys (n=48) and 31% of the subnational level surveys were considered to be representative of the target population.

## C. SAMPLE DESIGN

Of the 161 surveys, 45% (n=72) used random sampling/selection (i.e. probability sampling of some kind). The sample designs for the remaining 89 surveys were identified as follows: 22% as mixed sampling procedures, in which there are elements of probability sampling (n=36); 21% as non-probability sampling (n=34); and 12% as having no information available on the sample design (n=19).

At the country level, random sampling/selection was most common in Norway (100%), Romania (89%), and Germany (52%). Croatia was the only country for which non-probability sampling was the most prevalent (n=14), representing 38% of all their identified surveys. In Switzerland, mixed sampling procedures was the top sample design type, as it was utilised by six surveys (67%). Surveys conducted in Turkey largely used mixed sampling procedures (38%) or non-probability sampling (n=8, 33%). While sample design information was largely available and documented across the surveys identified in the six countries, 30% of Croatian surveys did not have this information reported.

As part of the sample design metadata, information as to whether or not a survey had partitioned subgroups focusing on specific target sub-populations (up to five different subgroups could be identified) was also captured. These are those situations where a survey sample is designed to be partitioned by origin or any other significant categorization of the EMM populations. For example, a sample may be formed by 1,200 respondents equally distributed across four subgroups of 300 respondents each as to allow comparing (e.g.) migrants from Turkey, Morocco, Portugal and Romania.

Of the 161 surveys, 54% (n=86) had at least one partitioned subgroup as part of its survey (see Table 16). The presence of a partitioned subgroup appeared to be equally prevalent at the national and subnational level, as it represented 56% (n=45) of national level surveys and 51% (n=41) of subnational level surveys.

While almost all of the surveys in Switzerland (89%) included a partitioned subgroup, the opposite trend was found in surveys identified in Norway and Romania; specifically, surveys with a partitioned subgroup only constituted 22% of the Norwegian surveys and 33% of the Romanian surveys. As for Germany, Croatia, and Turkey, there seemed to be a balance between surveys with and without a partitioned subgroup.

In looking strictly at surveys with partitioned subgroups (n=86), 14% (n=12) had one subgroup identified, 30% (n=25) had two subgroups identified, 10% (n=9) had three subgroups identified, 15% (n=13) had four subgroups identified, and 31% (n=27) had five subgroups identified.

**Table 16. National vs. subnational surveys with at least one partitioned subgroup (frequencies and percentages)**

| Territorial scope of the survey | Survey had at least one partitioned subgroup | | |
|---|---|---|---|
| | Yes | No | Total |
| National | 55.6 | 44.4 | 100% |
| | 45 | 36 | 81 |
| Subnational | 51.2 | 48.8 | 100% |
| | 41 | 39 | 80 |
| Total | 53.4 | 46.6 | 100% |
| | 86 | 75 | 161 |

Note 1: Surveys could have had more than one sampling unit identified (e.g. PSU and SSU identified).
Note 2: Not all the identified sampling units are listed. This table only captures the sampling units that were more commonly identified across the 161 surveys.

## D. SAMPLING FRAME AND SAMPLING UNITS

Different types of sampling frames were used by the 161 surveys. For the purposes of analysis, the following broad categories (of sampling frames) were identified: census, government database, electoral roll, list of foreigners, telephone list/directory/records, population register/registry, other, and no information provided. Across the 161 surveys, population registers/registries were the most common (19%), followed by the census (11%) and then the telephone list/directory/records (10%). It is also notable that the sampling frame information was not available for 32% (n=52) of the surveys, with this being most likely in Turkey (63%) and in Romania (44%).

At the country level (see table 17), population registers/registries were the most prevalent in Germany and Switzerland, representing 29% (n=21) and 56% (n=5) of their surveys respectively. The electoral roll was the most common sampling frame in Norway only (56%) and the census in Croatia only (35%).

When examining the surveys by territorial scope, the top three sampling frames for national level surveys were the telephone list/directory/records (15%), the census (14%), and population registers/registries (10%). To compare, population registers/registries (28%), schools (11%) and the census (9%) were the top three for subnational level surveys. For both national and subnational surveys, a sizeable number of them had no sampling frame information available, with 36% (n=29) for the former and 29% (n=23) for the latter.

**Table 17. Categories of sampling frames, by country and territorial scope (frequencies)**

| Sampling frame, by category | Country and territorial scope (frequencies) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Croatia | | Germany | | Norway | | Romania | | Switzerland | | Turkey | | Total |
| | Nat'l | Subnat'l | Nat'l | Subnat'l | Nat'l | Subnat'l | Nat'l | Subnat'l | Nat'l | Subnat'l | Nat'l | Subnat'l | |
| Census | 6 | 7 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 18 |
| Government database | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| Electoral roll | 0 | 1 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 8 |
| List of foreigners | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 |
| Phone | 0 | 0 | 10 | 4 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 16 |
| Population register / registry | 2 | 0 | 2 | 19 | 1 | 1 | 0 | 0 | 3 | 2 | 0 | 0 | 30 |
| School | 0 | 4 | 6 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 15 |
| Other | 1 | 3 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 16 |
| None identified | 7 | 4 | 13 | 7 | 2 | 0 | 4 | 0 | 0 | 0 | 3 | 12 | 52 |
| **Total** | **18** | **19** | **35** | **38** | **8** | **1** | **9** | **0** | **5** | **4** | **6** | **18** | **161** |

As for sampling units (primary sampling units (PSUs) and/or secondary sampling units (SSUs) could be documented), a variety was used across the 161 surveys, with the most common ones being: households, families, individuals, schools, students, and locations (e.g. neighborhoods, municipalities). Out of these six categories, the most prevalent was individuals, as it was used as a sampling unit in 74% of the surveys (n=119); when dissecting by territorial scope, this in turn translated into 77% (n=62) at the national level and 71% (n=57) at the subnational level. After individuals, households was the next most commonly used sampling unit, with 25% of surveys having identified this type. At the national level, households were used as a sampling unit in 26% of surveys (n=21) and at the subnational level in 24% of surveys (n=19).

At the country level (see Table 18), the majority of the identified surveys within a country used individuals as the sampling unit. The only exception was Romania, where households (55%) was the most frequently identified sampling unit. Broadly speaking, Germany seemed to have the most diversity when it came to the sampling units used; for example, it was the only country for which households, families, individuals, schools, students, and locations had been identified as a sampling unit (either PSU or SSU) in at least one of the identified surveys.

## Table 18. Commonly used sampling units, by country (frequencies and percentages)

| Sampling frame, by category | Country | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Croatia | | Germany | | Norway | | Romania | | Switzerland | | Turkey | | Total | |
| | No. | % | No. | % | No. | % | No. | % | No. | % | No. | % | No. | % |
| **Households** | | | | | | | | | | | | | | |
| Yes | 6 | 16.2 | 9 | 12.3 | 0 | 0.0 | 5 | 55.6 | 8 | 88.9 | 12 | 50.0 | **40** | **24.8** |
| No | 31 | 83.8 | 64 | 87.7 | 9 | 100.0 | 4 | 44.4 | 1 | 11.1 | 12 | 50.0 | **121** | **75.2** |
| **Total** | **37** | **100.0** | **73** | **100.0** | **9** | **100.0** | **9** | **100.0** | **9** | **100.0** | **24** | **100.0** | **161** | **100.0** |
| **Families** | | | | | | | | | | | | | | |
| Yes | 0 | 0.0 | 6 | 8.2 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 1 | 4.2 | **7** | **4.3** |
| No | 37 | 100.0 | 67 | 91.8 | 9 | 100.0 | 9 | 100.0 | 9 | 100.0 | 23 | 95.8 | **154** | **95.7** |
| **Total** | **37** | **100.0** | **73** | **100.0** | **9** | **100.0** | **9** | **100.0** | **9** | **100.0** | **24** | **100.0** | **161** | **100.0** |
| **Individuals** | | | | | | | | | | | | | | |
| Yes | 36 | 97.3 | 44 | 60.3 | 9 | 100.0 | 3 | 33.3 | 8 | 88.9 | 19 | 79.2 | **119** | **73.9** |
| No | 1 | 2.7 | 29 | 39.7 | 0 | 0.0 | 6 | 66.7 | 1 | 11.1 | 5 | 20.8 | **42** | **26.1** |
| **Total** | **37** | **100.0** | **73** | **100.0** | **9** | **100.0** | **9** | **100.0** | **9** | **100.0** | **24** | **100.0** | **161** | **100.0** |
| **Schools** | | | | | | | | | | | | | | |
| Yes | 1 | 2.7 | 12 | 16.4 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | **13** | **8.1** |
| No | 36 | 97.3 | 61 | 83.6 | 9 | 100.0 | 9 | 100.0 | 9 | 100.0 | 24 | 100.0 | **148** | **91.9** |
| **Total** | **37** | **100.0** | **73** | **100.0** | **9** | **100.0** | **9** | **100.0** | **9** | **100.0** | **24** | **100.0** | **161** | **100** |
| **Students** | | | | | | | | | | | | | | |
| Yes | 0 | 0.0 | 2 | 2.7 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | **2** | **1.2** |
| No | 37 | 100.0 | 71 | 97.3 | 9 | 100.0 | 9 | 100.0 | 9 | 100.0 | 24 | 100.0 | **159** | **98.8** |
| **Total** | **37** | **100.0** | **73** | **100.0** | **9** | **100.0** | **9** | **100.0** | **9** | **100.0** | **24** | **100.0** | **161** | **100.0** |
| **Locations** | | | | | | | | | | | | | | |
| Yes | 0 | 0.0 | 15 | 20.5 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | **15** | **9.3** |
| No | 37 | 100.0 | 58 | 79.5 | 9 | 100.0 | 9 | 100.0 | 9 | 100.0 | 24 | 100.0 | **146** | **90.7** |
| **Total** | **37** | **100.0** | **73** | **100.0** | **9** | **100.0** | **9** | **100.0** | **9** | **100.0** | **24** | **100.0** | **161** | **100.0** |

Note 1: Surveys could have had more than one sampling unit identified (e.g. PSU and SSU identified).
Note 2: Not all the identified sampling units are listed. This table only captures the sampling units that were more commonly identified across the 161 surveys.

## E. KEY DESCRIPTIVES OF SAMPLE SIZE

We principally targeted national surveys with more than 300 respondents and subnational level surveys with 150 respondents or more. Most are above these thresholds, though the Croatian team have collected surveys with lower numbers of respondents (Table 19). The units for which information is available in some of those Croatian surveys are schools rather than individuals even if the latter were the final sampling units, which explains the unusually low N. The mean Croatian ethnic minority survey has a good sample size of over 1400. The average local level survey has around 600 respondents. Medians are a bit lower, suggesting that there are a number of smaller surveys in the Croatian metadata, which includes 37 surveys in all.

Norway, Switzerland and Romania have eight or nine surveys each; all with a sizeable national level sample size. The median survey in Norway and Switzerland includes more than 2000 Ethnic Minority respondents, about 1400 in Romania. Sample sizes in Germany are similar to those in Norway and Switzerland, but there are a lot more surveys done in Germany. The six national level surveys in Turkey have very large sample sizes.

There is substantial variation in the size of the samples in these surveys, within each country and overall. Local level surveys are generally smaller, which is not surprising. The largest sample is a Turkish survey with close to 50 000 respondents.

**Table 19. Sample size**

| Country | Level | Min | Max | Mean | Median | Std | N |
|---|---|---|---|---|---|---|---|
| Croatia | Subnational | 15 | 1671 | 620 | 421 | 541 | 19 |
| | National | 31 | 4975 | 1461 | 1028 | 1623 | 18 |
| Norway | Subnational | 1200 | 1200 | 1200 | 1200 | 0 | 1 |
| | National | 437 | 6350 | 2467 | 1980 | 2121 | 8 |
| Switzerland | Subnational | 643 | 1203 | 889 | 855 | 262 | 4 |
| | National | 1184 | 5973 | 2426 | 1764 | 2014 | 5 |
| Romania | Subnational | | | | | | 0 |
| | National | 500 | 1857 | 1236 | 1398 | 487 | 8 |
| Germany | Subnational | 378 | 4888 | 1701 | 1247 | 1352 | 28 |
| | National | 473 | 15801 | 3101 | 2038 | 2865 | 35 |
| Turkey | Subnational | 153 | 5790 | 1264 | 818.5 | 1404 | 18 |
| | National | 1230 | 47958 | 12610 | 6810 | 17632 | 6 |

## F. DESCRIPTIVES OF RESPONSE RATES AND METHOD USED

Calculating response rates in surveys can be challenging. Sometimes one does not have a clearly defined gross sample (the denominator when calculating response rates), and this depends on how sampling is done. There are different formulas at use in the academic literature for calculating response rates, and these sometimes differ from those used by commercial survey companies. We therefore have very limited data on response rates in the metadata. We do not have sufficient information on the methods used to calculate response rates to present it here – that information was provided with respect to just a handful of surveys.

The Norwegian metadata includes response rates for all surveys, whereas the Swiss and German data includes this information with respect to a majority of EMM surveys. Response rates vary widely, as one would expect (Table 20). The German data, especially, includes some surveys with very low response rates, and some with unusually high rates of response, probably due to unique features in the design of those surveys. Though response rates are not an ideal measure of the quality of a survey, they could be used as one of several indicators of data quality. On that score, the average response rate is quite high, suggesting that there are a number of high quality surveys in the EMM Survey Registry.

**Table 20. Response rates**

| Country | Level | Min | Max | Mean | Median | Std | N |
|---|---|---|---|---|---|---|---|
| Croatia | Subnational | | | | | | 0 |
| | National | 60.0 | 60.0 | 60.0 | 60 | | 1 |
| Norway | Subnational | 52.6 | 52.6 | 52.6 | 52.6 | | 1 |
| | National | 35.0 | 64.2 | 48.0 | 47.35 | 11.0 | 8 |
| Switzerland | Subnational | 36.0 | 47.5 | 41.8 | 41.75 | 8.1 | 2 |
| | National | 35.5 | 42.0 | 38.2 | 37.2 | 3.4 | 3 |
| Romania | Subnational | | | | | | 0 |
| | National | 46.0 | 81.4 | 63.8 | 64 | 17.7 | 3 |
| Germany | Subnational | 8.1 | 97.9 | 51.9 | 43 | 31.3 | 17 |
| | National | 6.4 | 85.8 | 42.3 | 41 | 26.7 | 19 |
| Turkey | Subnational | 40.0 | 93.0 | 66.5 | 66.5 | 7.5 | 2 |
| | National | | | | | | 0 |

## G. DATA COLLECTION MODE

All types of data collection modes are used for these surveys (Table 21). We know that collecting high quality survey data on immigrant and ethnic minority populations can be challenging. One way of ensuring high quality data is to use expensive but reliable data collection modes such as face-to-face interviews. A substantial share of the surveys are in fact done face-to-face. Telephone, web- and postal surveys are also frequently used. Except for the relatively high frequency of face-to-face interviews, there is no particular pattern. Researchers and others that collect survey data on ethnic and immigrant minorities use all the tools in the data collection toolkit.

## Table 21. Data collection modes

| Country and territorial scope (national vs. subnational) | | Survey designed as a general population survey (frequencies) | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Face to face (PAPI) | | Face to face (CAPI) | | Telephone | | Web/e-mail survey | | Paper self--administered (collected) | | Paper self--administered (postal) | | Other | | Total | No info |
| | | % | N | % | N | % | N | % | N | % | N | % | N | % | N | N | N |
| **Croatia** | Subnational | 5.6 | 10 | 5.3 | 1 | 0.0 | 0 | 0.0 | 0 | 31.6 | 6 | 0.0 | 0 | 10.5 | 2 | **19** | 1 |
| | National | 55.0 | 11 | 0.0 | 0 | 0.0 | 0 | 20.0 | 4 | 25.0 | 5 | 0.0 | 0 | 0.0 | 0 | **20** | 0 |
| **Norway** | Subnational | 0.0 | 0 | 0.0 | 0 | 100.0 | 1 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | **1** | 0 |
| | National | 23.1 | 3 | 0.0 | 0 | 38.5 | 5 | 15.4 | 2 | 0.0 | 0 | 23.1 | 3 | 0.0 | 0 | **13** | 0 |
| **Switzerland** | Subnational | 0.0 | 0 | 33.3 | 2 | 33.3 | 2 | 0.0 | 0 | 33.3 | 2 | 0.0 | 0 | 0.0 | 0 | **6** | 0 |
| | National | 0.0 | 0 | 0.0 | 0 | 50 | 4 | 37.5 | 3 | 0.0 | 0 | 12.5 | 1 | 0.0 | 0 | **8** | 0 |
| **Romania** | Subnational | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | **0** | 0 |
| | National | 87.5 | 7 | 12.5 | 1 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | **8** | 1 |
| **Germany** | Subnational | 7.7 | 3 | 25.6 | 10 | 23.1 | 9 | 5.1 | 2 | 10.3 | 4 | 25.6 | 10 | 2.6 | 1 | **39** | 4 |
| | National | 25.6 | 11 | 16.3 | 7 | 30.2 | 13 | 9.3 | 4 | 7.0 | 3 | 9.3 | 4 | 2.3 | 1 | **43** | 0 |
| **Turkey** | Subnational | 75,0 | 15 | 15 | 3 | 0.0 | 0 | 10 | 2 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | **20** | 2 |
| | National | 83.3 | 5 | 16.7 | 1 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | 0.0 | 0 | **6** | 0 |

## H. SURVEYS INCLUDING PERSONAL INTERVIEWS

We have also collected some information about how personal, face-to-face interviews were conducted. Given the relatively small number of surveys in this category, we have aggregated this information across countries (Table 22). Professional interviewers are more common than non-professional ones, and a majority of interviewers know migrant languages. About half of the surveys had questionnaires in migrant languages, and the most common ones were Arabic and English.

**Table 22. Interviewing protocols**

| | N | % |
|---|---|---|
| **Who interviewed?** | | |
| Professional interviewers only | 36 | 63.16 |
| Non-professional interviewers | 5 | 8.77 |
| A mix | 16 | 28.07 |
| Total | 57 | 100.0 |
| Don't know | 1 | |
| Information not available | 11 | |
| **Interviewers spoke migrant languages** | | |
| Yes | 35 | 70.0 |
| No, but translator(s) present/available | 7 | 14.0 |
| No, nobody had targeted language skills | 8 | 16.0 |
| Total | 50 | 100.0 |
| Don't know | 5 | |
| Information not available | 16 | |

| | N | % |
|---|---|---|
| **Questionnaire in migrant language** | | |
| Yes | 30 | 50.0 |
| No | 30 | 50.0 |
| Total | 60 | 100.0 |
| Don't know | 7 | |
| Information not available | 21 | |
| **If yes, which? Top 5 languages** | | |
| 1. Arabic | 10 | |
| 2. English | 9 | |
| 3. Turkish | 5 | |
| 4. Albanian, Italian, Bosnian, Norwegian, Urdu | 4 | |
| 5. Serbian, Croatian | 3 | |

# 8 ACCESSIBILITY AND REUSABILITY OF THE EMM SURVEYS IN EUROPE

Ethnic and Migrant Minorities (EMMs) are typically classified as hard to reach as well as hard to research populations, when it comes to survey research. Not only is it often hard to define a fitting sampling frame, cultural and language issues are additional known factors that complicate research in this field. Thus, thorough documentation of collected data and rigorous data sharing practices are essential to help gain a deeper and more encompassing understanding of the field. In short: To gain a deeper understanding of the underlying issues it is necessary for survey data producers focusing on EMM populations to share their insights not only in the format of scientific publications, presentations or policy papers, but also in the form of research data. Especially in a period of time that is defined by the so-called "datafication of society", which produces a seemingly endless stream of quantitative data that needs to be judged and discussed in regards to its validity and accuracy.

This section will address how the current surveys that are referenced by the EMM Survey Registry are documented, archived and made accessible, if persistent identifiers are available and which kind of material is available. While the first part of this section will simply address on a country basis how many studies are available, the later parts will give a more detailed insight into the documentation practices found in this field of survey research.

When it comes to simple availability of the referenced data, approximately 40% of the data is available either publicly via data archives or upon request from a researcher. However, a third is unavailable, and the status of a quarter of the studies is unknown (Figure 4).

Yet a more detailed look at the country distribution shows that there is considerable cross-national variation. The status of surveys identified in Croatia (89%) and Romania (78%) is mostly unknown. In Turkey more than 90% of the studies are unavailable. In comparison to this in Germany two thirds of the data sets are available, either publicly or by request. In Norway and Switzerland all the studies are available (see Table 23).

This data is nearly a direct mirror in regards to data archiving practices. In Norway all of the nine identified studies are archived. In Switzerland seven of the nine studies are archived

and only two sub-national ones are not found in an archive. In Germany two thirds of the studies are available via GESIS – DAS. This archive hosts 60% of the national studies and more than 70% of the subnational studies. No study from the other three countries was found in a data archive.

**Figure 4. Data Access**


As in the case of data sharing practices, the documentation of the data collection tools and of data management helps other researchers to understand the underlying survey data, hence fostering data re-use. Among the six countries studies, three count with consolidated social data archives: FORS (Switzerland), GESIS (Germany) and NSD (Norway; where also the CESSDA HQ is located) are not only embedded in large social science infrastructures, but are also long-established players in the field of data management and archiving. They are seen as reference models for other European countries trying to establish their own social science research infrastructure. In two of the countries studied, despite the existence (e.g. RoDA in Romania) or the ongoing efforts (The Croatian Data Archive Services for the Social Sciences) to establish a national data service provider, their reach and resources are limited.

This situation obviously has an impact on the chance to identify and find data as well. While in Norway and in Switzerland all the surveys have either a link or a DOI to identify the studies, the other countries do not fare so well. In Germany more than 60% of the studies (44 out of 73) have a link and 80% (46) have a DOI. Interestingly enough nearly all but one of the subnational studies in Germany have a DOI, while only two thirds of the national ones have such a persistent identifier. This hints at the fact that most subnational level surveys identified are actually mapped and identified consistently, while a group of national ones lack this kind of information. On the flipside only two studies in Romania — national studies — and three in Croatia — one national and two subnational — have a link. Not a single study in Turkey did so.

The same pattern is true for the chance to find citable documentation and data. Again, the best practice examples are the datasets from Norway and Switzerland. Only one identified subnational survey in Norway lacks citable documentation and the dataset itself is also uncitable. In Switzerland all the survey documentation is available, however only four of the nine datasets themselves can be cited. In Germany 40% of the national and 50% percent of the subnational surveys lack citable documentation. However, two thirds of the national datasets can be cited and 55% of the subnational ones. In Turkey only one national survey has citable documentation. For Romania and Croatia all surveys lack this type of citing information. When it comes to the datasets themselves, Romania, Croatia and Turkey all lack citable datasets.

Finally, when it comes to the availability of documentation in English — a very important factor to translate national insights into globally usable knowledge — a more diverse picture emerges. While all nine projects identified in Switzerland have English language documentation not a single one of the Turkish studies does so. The other countries offer a mixed picture. In Croatia only one national survey is available in English, while in Norway five out of eight national surveys have English documentation. The single subnational survey is only documented in Norwegian. In Germany 49% of the national and 46% percent of the subnational surveys have English documentation, and in Romania four of the nine national surveys have English language materials available.

**Table 23. Availability by country and geographical level (percentages)**

| Status of Data | Croatia | | | Romania | | | Turkey | | |
|---|---|---|---|---|---|---|---|---|---|
| | National | Subna-tional | Total | National | Subna-tional | Total | National | Subna-tional | Total |
| Available | 0% | 0% | 0% | 22% | 0% | 22% | 0% | 0% | 0% |
| Available upon request | 6% | 11% | 8% | 0% | 0% | 0% | 0% | 0% | 0% |
| Unavailable | 6% | 0% | 3% | 0% | 0% | 0% | 100% | 89% | 92% |
| Unknown | 89% | 90% | 89% | 78% | 0% | 78% | 0% | 11% | 8% |
| N | 18 | 19 | 37 | 9 | 0 | 9 | 6 | 18 | 24 |

| Status of Data | Norway | | | Switzerland | | | Germany | | |
|---|---|---|---|---|---|---|---|---|---|
| | National | Subna-tional | Total | National | Subna-tional | Total | National | Subna-tional | Total |
| Available | 0% | 100% | 11% | 60% | 50% | 56% | 40% | 21% | 30% |
| Available upon request | 100% | 0% | 89% | 40% | 50% | 44% | 31% | 40% | 36% |
| Unavailable | 0% | 0% | 0% | 0% | 0% | 0% | 29% | 37% | 33% |
| Unknown | 0% | 0% | 0% | 0% | 0% | 0% | 0% | 3% | 1% |
| N | 8 | 1 | 9 | 5 | 4 | 9 | 35 | 38 | 73 |

# 9 CONCLUSIONS AND RECOMMENDATIONS

The EMM Survey Registry produced by COST Action 16111 Ethmigsurveydata, with the support of the H2020 project SSHOC is a new and rich tool to allow the discovery, facilitate the access and foster the reuse of survey data focusing on EMMs across Europe and neighbouring countries. The EMM Survey Registry aims at making survey data on EMMs as FAIR as possible: Findable, Accessible, Interoperable and Reusable. To this end, it collects and displays the metadata of all surveys that have been detected in 35 European and neighbouring countries since 2000, and provides detailed information about the characteristics of these surveys — target population, sample design, data collection modes, availability, citations, etc. At the same time, the rich metadata compiled through it allows us to gain a detailed and comprehensive picture of the strengths and weaknesses in our knowledge of the complex processes and multidimensional nature of the integration of EMMs.

This first version of the report on the EMM Survey Registry metadata has focused on the six countries for which the metadata were both complete and controlled for quality: Croatia, Germany, Norway, Romania, Switzerland and Turkey. Future versions of the report will include metadata from more countries up to a total of at least 30 European and neighbouring countries. Whereas the current version of the report analyses the metadata of 161 surveys, we anticipate that the final version of the report will include metadata for more than 800 surveys. As such, the current version of the report portrays a snapshot that may vary somewhat with the inclusion of metadata from more countries. Nevertheless, some meaningful patterns have already emerged from the analysis of six countries that are sufficiently heterogeneous in terms of the size of the EMM population stocks and flows, and in terms of the survey practices.

**First**, we observe that there is no single pattern of dominance of a preferred geographical level of focus. Around half of all the surveys conducted in the six countries studied are undertaken through nationwide samples, and the other half through subnational samples that only cover one or a few localities or regions. This balance is also prevalent in Croatia, Germany and Switzerland, whereas in Norway and Romania the vast majority are nationwide and in Turkey the vast majority are subnational. Where there are subnational surveys, they tend to be in large cities or middle-sized towns. Surveys in rural areas exclusively are very uncommon and limited to Croatia, but they are more common in situations where geographical sampling includes a mix of various geographical settings (around a third).

**Second**, surveys that are part of a larger cross-national study whereby the same questionnaire and instruments are applied to samples of EMM populations in several countries in a concerted fashion are a minority (20% of all studied), and they are unevenly covering the whole set of countries considered here.

**Third**, the overwhelming majority of the surveys targeting EMM populations focus on a subset of either migrant minorities or ethnic minorities in any given geographical setting (90%). Hence, they rarely aim at covering the whole migrant or ethnic minority population in the selected locations. This means that the surveys undertaken until now provide us only with very partial pictures of the integration of EMM populations, as they refer to only specific subsets of these. In particular, much of the knowledge base in some countries is focusing on very specific groups; for example, surveys in Germany focusing on the Turkish-origin migrants are very abundant but those focusing on the whole migrant-origin population are rarer. Equally, in some countries only either the ethnic or the migrant minorities are well studied through surveys and the knowledge base for the other set of minorities is missing.

**Fourth,** the knowledge base generated by surveys to EMM populations is predominantly dominated by academic research needs, though a large share also aims at informing public policy. However, only a minority serve to feed official government statistics or the information needs of NGOs and non-profit organisations.

**Fifth**, EMM populations are categorised in a vast and heterogeneous fashion, but generic categories such as 'migrants', 'ethnic minorities', 'refugees' or 'foreigners' are more common. Among all the forms of categorisation used in the surveys studied, those referring to origin or to legal / administrative status were the most frequently used. Moreover, the categorisation used in the surveys tends to coincide with the origin or ethnicity of the largest minorities in each country, though we find some exceptions in countries were the categories used most often are more generic and inclusive ones, such as "second generation migrants" (e.g. in Norway) and "refugees" (e.g. in Turkey). Categorisations around religion-related labels or socio-economic status are much less common but still present across the surveys studied. Additionally, the way the various groups might be identified or self-identified is currently quite restrictive, and in many countries it relies on only a few variables. For example, to identify migrant-origin minorities typically the country of birth or the (current) nationality of the respondent is used; rarely is the country of birth of the parents or grandparents considered or the citizenship at birth of the respondent and/or parents/ grandparents. Similarly, to identify ethnic minorities it is more common to provide questions that allow only to provide one self-identification than to allow for multiple identities.

**Sixth**, surveys to EMM populations overwhelmingly focus on the adult population (60%) or a combination of adult and minor respondents (23%). From this point of view, the number of surveys that focus particularly on the issues that are specific to the youth and children of migrant or ethnic minority backgrounds is very small. As a consequence, our ability to understand the processes of integration and the challenges to equality throughout the various stages of the life cycle is hampered. By contrast, the coverage of the EMM population by sex of the respondents

is very balanced, with the vast majority of surveys including both female and male respondents. Additionally, around half of the surveys include a subgroup formed by the majority population, and this is more common for surveys conducted at the subnational level than for nationwide surveys.

**Seventh**, the topics covered by the surveys have evolved considerably over time and they vary somewhat across countries but some topics — such as demographic characteristics, educational attainment and human capital, labour market integration, identity and belonging, as well as political inclusion, participation and attitudes — have dominated the scene of surveys to EMM populations consistently. Since 2015 surveys focusing on issues relating to asylum seekers and refugees have considerably increased in number.

**Eighth**, in terms of the technical characteristics of the surveys, a mixed picture emerges in terms of the extent to which best practices are followed. Overall, there is a good balance between the proportion of surveys that are single-cross section studies and those that include an over-time component — repeated cross sections or longitudinal studies, but this distribution varies considerably across countries. A certain weakness can be pointed out in relation to the ability to design samples that can be representative of the underlying target population, as only 45% were identified as such and these are highly concentrated in some countries. This is mostly due to the fact that a considerable proportion of surveys (33%) employ non-probability sampling forms or provide no information on sample design. Although face-to-face interviewing is the norm in the field, only around half of the surveys conducted to EMM populations included a questionnaire translated into the mother tongue of the minorities studied, which is very likely to reduce the quality of the data collected as well as response rates in general. Additionally, the vast majority of the surveys on EMM populations do not report on response rates.

**Nineth**, the surveys focusing on EMM populations are — for the most part — not easily accessible and are very often not well documented and citable. Only 40% of the surveys identified are publicly available through data archives or by request to a researcher, a majority are not findable through a persistent identifier (e.g. DOI), for many the survey documentation is not publicly available (if it exists), and there are no data citations available to ensure that the data producers and distributors are appropriately credited and referenced. Although we find considerable variation across countries in these respects, overall EMM survey data faces serious challenges in terms of future preservation and accessibility.

## Given these findings, we can offer a number of tentative recommendations subject to the completion of the analysis of the remaining countries:

1. Future surveys should improve the knowledge base on the integration of EMMs in towns and rural areas, as these are considerably less covered by the existing sets of surveys.

2. Cross-national programmes of surveys targeting EMM populations should be fostered to allow for a better comparative knowledge base.

3. Surveys that aim at covering the entire migrant or ethnic minority population are required for a more comprehensive understanding of the integration of EMM populations.

4. Countries with intense survey research exclusively on migrant minorities should consider expanding the focus to ethnic minorities as well, and vice versa.

5. Funders may want to promote multi-purpose surveys to EMM populations that contribute to the knowledge needs of academics, policy makers and civil society organisations through more integrated approaches to survey design that do not create knowledge silos.

6. Governments — at all levels, EU, national, regional, local — may want to consider including at least one periodical survey (repeated cross-section or longitudinal) specifically targeting the EMM population within their jurisdiction as part of their official statistics portfolio.

7. Surveys that focus on EMMs that are also religious minorities are not very common and, hence, future funding efforts could be directed at studying whether these groups face specific challenges in terms of equal opportunities for integration and inclusion.

8. A future balanced focus on the EMM youth and children would be particularly useful, as the current knowledge base on these age groups is very limited across most countries.

9. When designing the pre-screening and questionnaire instruments, survey data producers should consider a wider range of options to define ethnic and migrant minorities. Survey instruments that allow for a richer set of variables gauging migrant background and ethnic identities will result in more nuanced analyses about orientations, experiences and outcomes.

10. Surveys targeting specifically the EMM population should always consider the opportunity and benefits of including a subgroup of the majority population as a point of reference. Typically, surveys to EMMs are designed to inform about experiences and challenges of integration but it is very difficult to gauge these without a point of reference to the non-minority population.

11. Surveys to EMM populations do not seem to be sufficiently capturing certain topics that would seem a priori critical for the proper understanding of the integration of these populations. In particular, surveys directed at migrant populations should always include sufficient information about the reasons and drivers for migration, as well as about the legal status and administrative situation.

12. The study of important topics — such as the situation of unaccompanied migrant minors and human smuggling and trafficking — should be promoted in future survey research as the knowledge base is very limited.

13. A number of countries should consider investing in surveys that can introduce a temporal element with repeated cross-sectional and longitudinal designs, as this adds value to survey data collection and enhances the knowledge base.

14. Survey research on EMM populations relies too much on samples that are not statistically representative of the underlying population and survey data producers should, in the future, learn from the best practices in the field to adapt to local circumstances while retaining the beneficial properties of probability sampling.

15. In the future, survey data producers of studies on EMM populations need to make a greater effort at providing detailed documentation and transparency in relation to the technical characteristics of the surveys, including (but not restricted to) sample design, questionnaire administration, fieldwork operations and response rates.

16. Survey data producers should bear in mind that the translation of the questionnaires into the mother tongue of the minorities studied is not only good practice but also highly advisable to ensure acceptable response rates that will reduce the bias in the outcomes obtained from the survey.

17. International and national efforts are required to ensure that a larger number of surveys can be fully documented, archived and made available and citable for future re-use. National data archives as well as CESSDA should play proactive roles in fostering the future preservation of the existing surveys focusing on EMM populations.

18. Research funders, data producers and data users should become more aware about the collective benefits of making EMM survey data FAIR and embed the FAIR principles into their ordinary research practices.

ethmig survey data

ethmigsurveydata@sciencespo.fr

www.ethmigsurveydatahub.eu

@EthMigSurveyDat

https://www.facebook.com/ethmigsurveydata/