



WG1 & WG2

SURVEYS METADATA TEMPLATE GUIDELINES DOCUMENT

Gathering information on migrant and ethnic minorities surveys

Laura Morales, Mónica Méndez, Anikó Bernat, and
Johannes Bergh (WG1 & WG2 co-leaders)

This guidelines document is based upon work from COST Action
ETHMIGSURVEYDATA, supported by COST (European Cooperation in Science and
Technology).

COST (European Cooperation in Science and Technology) is a funding agency for
research and innovation networks. Our Actions help connect research initiatives
across Europe and enable scientists to grow their ideas by sharing them with their
peers. This boosts their research, career, and innovation.

www.cost.eu

Introduction

This document outlines the data collection template and guidelines that will allow each national delegation of the COST Action to collect information on surveys specifically addressed to ethnic and/or migrant minority (EMM) populations and those in which these groups constitute a significant sub-sample.

This Word document aims at explaining important aspects relating to definitions, priorities, data collection processes, etc. However, please note that the actual template is the Excel file that we are circulating alongside this Word document (FINAL ETHMIGSURVEYDATA Metadata Template.xlsx). In this Excel file, each of the variables that are listed in section 2 below will be a row, whereas the information for each survey will be filled in by column (also called below as a “register”). Please make sure that you submit each survey as one column in the same sheet (either the sheet “National surveys” or the sheet “Subnational surveys”). Please do not send the file with a multiplication of the sheets/tabs, and please do not mix national surveys and subnational surveys in the same sheet. More detailed information about this can be found in section 4.2.

This document is structured into the following sections:

1. General guidelines on what survey data to search for and compile, and how to approach this task;
2. A table with the list of metadata variables to include in the Excel template;
3. More detailed annotations and clarifications on variables that require additional explanations;
4. Technical notes regarding the use of the Excel template; and
5. A checklist that needs to be completed and submitted when sending in the final Excel template (**deadline: at the latest 10 February 2019**)

1. General guidelines on what survey data to search for and compile, and how to approach this task

1.1. Time coverage

All national teams must systematically search and compile the information about surveys conducted since the year 2000. By 'conducted', we mean that the starting date of fieldwork was on or after 1 January 2000. This means that all countries will aim at uncovering existing surveys undertaken in the 2000s and 2010s.

However, please note that this is a 'minimal' requirement and national delegations (i.e., teams) can be more inclusive depending on: (i) how easy it is for them to include surveys prior to this date in their data compilation exercise and (ii) how relevant they think that surveys prior to that date are in the case of their own country.

When deciding what surveys to include and exclude from a timing point of view, please keep in mind the following considerations:

- In the case of longitudinal/panel (or multi-wave cross-sectional) surveys where the first waves were carried out before 2000: the waves/samples prior to 2000 should also be included in the list for completeness reasons. For longitudinal/panel/multi-wave cross-sectional surveys: please fill in one register per wave so that we have the full set of waves included in the Excel template.
- National teams should make it clear (in the “Annotations” sheet that is included in the Excel template) whether they have used 2000 as the starting point to do the systematic search/report of surveys, or an earlier one, and if so which date was used as a cut-off point.
- If an earlier date is used, the information provided in the “Annotations” sheet of the Excel template should clearly specify whether the inclusion of surveys prior to 2000 was comprehensive for the extended period covered or includes only a selection of the 'major' or 'noteworthy' surveys of interest.

1.2. Types of surveys to include and guidelines of minimal sample sizes

All throughout this document (and for the Action as a whole), we consider a **survey** to be **sample-based quantitative studies** that interview respondents based on a fully structured and harmonised questionnaire and for which data at the individual level would be available if the dataset were made available by the data producers. We are, thus, **not including** studies that are **exclusively based on qualitative interviews** using only a semi-structured interview outline/questionnaire. **Mixed-methods studies** that include a quantitative survey component can be added but the template should be filled **only for the quantitative survey** component of the study.

Moreover, we are only including surveys where respondents are ordinary individuals of ethnic or migrant origin. Hence, we are not including surveys to political, social, or economic elites, or surveys to the representatives of associations who are answering on behalf of an organisation. Equally, we are not including full population censuses—even for the countries where the census is undertaken with a sample of the population—or any aggregate-level statistical or administrative data.

Although surveys that have employed probability samples are preferred, we recognise that for this field many of the surveys depart from ideal probability samples. In countries that are data rich, teams should exercise discretion as to which surveys do not reach a minimal degree of quality in terms of sample design to be included in the compilation (e.g., online surveys with very strong self-selection biases). In countries that are not data rich, teams should try to be comprehensive in covering all surveys that exist, even if this will include non-probability samples.

All national teams must systematically search and compile the information about surveys with the following characteristics:

(i) Nation-wide surveys specifically addressed to ethnic and/or migrant minority (EMM) respondents

These are surveys that are designed to be representative of the target population: all or a selection of EMM groups in a given country. It is important that the sample design is such that the aim is to cover the target EMM group(s) that reside across the whole country. Certainly, many of these national surveys will include stratified and cluster-based sample designs, where the whole territory of the country is not included in the sample. However, **the important point here is that the sample was designed to be representative of the overall (or relevant) national territory where EMMs live.** In some cases, the eventually achieved sample might not represent equally well the whole country due to differential response rates or to fieldwork complications. In principle, we stick to the criterion that the initial design is what matters, unless the actual implementation departed so much from the initial plan that it can no longer be meaningfully considered as a nation-wide sample. For instance, if a sample is designed as a national one and ends up including only a very small subset of localities, then we would consider it as a sub-national survey if due to (e.g.) response rates or fieldwork complications it was unable to cover all of the target population it intended to cover. Instead, if a sample is designed as nationally representative but it just so happens that the minority population is only present in a few localities, then the sample will be nationally representative of the given minority population, and it is probably best treated as a national survey. When in doubt, the national team making decisions can get in touch with the WG leaders for a case-by-case discussion. In such cases, please also provide any details that might help users in the “comments” variable available for each survey coded.

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 larger (before weighting).

(ii) Sub-national surveys specifically addressed to ethnic and/or migrant minority (EMM) respondents

These are surveys that are designed to be representative of the population of all or a selection of EMM groups in a given city/metropolitan area/region/province/canton or in a number of cities/metropolitan areas/regions/provinces/cantons, but which are not designed to be representative of the overall population of the relevant EMM group(s) that reside in the country as a whole. For example, a survey that samples EMMs in

only two cities/municipalities of a given country will be classified as 'sub-national' unless close to 100% of the EMM population living in that country actually live in those two cities only. The distinction between national and sub-national surveys is still relevant in small (or micro-) countries (e.g., Luxembourg, Malta, or Cyprus), as even in these countries there is always more than one municipality. If in doubt as to whether to classify a survey as national or sub-national, please check with the WG leaders of WG1 and WG2.

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) Nation-wide or subnational general population surveys with substantial sub-samples of ethnic and/or migrant minority (EMM) respondents

These are surveys that are designed to be representative of the overall population of a given country (or a given subnational territory) and, as part of the normal sampling process or as a result of the design of 'boost' samples, a sizeable number of EMM respondents are included in the effective sample. The goal of ETHMIGSURVEYDATA is not (primarily) to collect such surveys, so national teams need to exert their discretion in this case and include these surveys if, at least, one of the following conditions 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)¹;
- 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 of EMM respondents (even if smaller than 400);
- There are no (or very few) alternative sources of survey data on EMM integration in that country and the survey includes a sizeable number of EMM respondents (even if smaller than 400).

In general terms, national teams should bear in mind that our **priority is to cover surveys described under (i) and (ii)**, so those described under (iii) should only be included after the former have been covered.

Additionally, please bear in mind that we are primarily interested in surveys that include EMMs residing in the countries we are studying. In other words, if you code for country A, we are interested in surveys that include respondents who reside in country A and are an ethnic or migrant minority in country A. Returning emigrants to country A are not our priority, although they can be included if very relevant in the given country, but only after a full coverage of the main focus is ensured.

It is also important to clarify that **we are not including** surveys that **cover only majority/autochthonous** population respondents. We cannot extend our exercise to

¹ It is important to clarify that countries with large minority populations can choose to include surveys where less than 800 but at least 400 EMM respondents are included but including the surveys with 800 or more is compulsory. We always refer to raw/absolute number of respondents (i.e., before any weighting). A threshold is always arbitrary to consider what is a relatively small or large minority population, but 15% of the overall population might be a useful rough guideline to be interpreted flexibly.

cover all surveys that ask majority populations about their attitudes/positions around immigration or immigrants. Other international databases do that already and we do not have the capacity to cover all potentially interesting surveys. Thus, please do not include such surveys as we will have to exclude them later on and it would be an unnecessary waste of time.

Finally, we should note that we are aiming to cover all of the surveys that meet any of the definitions/characteristics outlined above, regardless of whether micro-data is/will be available. Our first goal in the Action is to compile a comprehensive list of surveys covering EMM's integration, regardless of whether the actual data is made available to researchers other than those who produced the data. Hence, as long as minimal documentation about them exists to fill in most or part of the red variables in our list of metadata variables below, they should be included in the Excel template. A subsequent goal of the Action is to put in place the mechanisms that might persuade and support data producers to make their survey data available to a wider range of users; this will be addressed in future stages.

1.3. Which surveys to include in which files when multi-location surveys exist

In some cases, the survey was done in country A (or a subset of countries) but the survey institute or the data producers were based in a different country or only in one of the various countries where fieldwork was conducted. For example, a survey undertaken by a survey company in Brussels with fieldwork undertaken in Belgium, Spain, and Romania, could be coded by the Belgian, Spanish, or Romanian teams of COST.

When this happens, please proceed as follows:

- The team that has detected the survey first contacts the national teams of all of the countries affected and copies the WG leaders;
- All teams from the countries covered by this survey agree on who will compile the information for each or all countries, as to ensure there is no duplication of the information;
- At the end of the process, each country sample should only be included in its respective Excel template (in the example above, the Belgian sample needs to be coded in the Belgian template, the Spanish sample in the Spanish template, etc.)

The question of fieldwork location may also emerge in other cases. For example, a Polish-based research team might have undertaken a survey of Polish migrants in the UK, Germany, and Spain. In that case, if the Polish team becomes aware of such a survey, they need to contact the UK, German and Spanish teams to provide them with the information they need to input in their respective Excel templates.

1.4. How to search for surveys – data compilation strategy

All teams will need to ensure that **they cover at least the following steps in their data compilation strategy** so as to uncover the largest number of surveys that exist:

- 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 some countries, such a search strategy might produce a considerable number of surveys for which to code the metadata. Our advice, on how to prioritise which ones to start coding/inputting, is that you first focus on those for which you have readily available information. For those which you do not have readily-available information, you can include their names and minimal details in the Excel template, and then “fish” for the missing information.

Minimal dictionary of keywords for the searches (to be adapted to the search functions in each platform):

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

To gain in efficiency, the search can combine the terms above with “OR”, e.g.:
 (survey AND ethnic) OR (survey AND migrant) OR, etc.

1.5. How to deal with limited information on the surveys

Our aim is to identify and cover as many existing surveys on EMM integration. Thus, even if national teams cannot retrieve considerable portions of information about a given survey, as long as you can identify the essential pieces of information (name of the survey, data producers, approximate dates/year of fieldwork, sample size and groups covered), we would prefer you to include the survey in the Excel template.

1.6. Who should be filling in the Excel template

Each national team should decide among themselves who exactly will be doing the searches for the surveys and who (if different) will be filling in the Excel file template.

However, our (strong) advice is that the **person(s) filling in the Excel template should ideally be well trained in quantitative survey research** and be **very familiar with the technical language used throughout this document**. (S)he/they will need to properly understand the difference between the various sample designs and strategies, how response rates are calculated, what a DOI is, etc. The template also requires familiarity with the language used in ethnic and migrant studies; if the person(s) filling in the Excel template lacks familiarity with these subfields, (s)he/they might want to double-check the interpretation of certain concepts before starting.

This, in turn, means that even if you are asking data producers (e.g., think tanks, government departments) about surveys they have produced, if they do not have the technical skills, you should not be asking them to fill in the Excel template directly. You should schedule a phone/personal appointment with them to ask them about the necessary details and then you should fill in the Excel file yourselves. You can also find an alternative solution that works (like asking them to give you the contact of the survey institute that undertook the survey). Some of you might want to involve the data producers when compiling information on their surveys (e.g. by asking them to check whether you got the information right or to fill in any gaps that you were not able to find out). In such cases, please be mindful about the question on data quality (10.1) at the end of the template and sensitivities around it. The data producers may not always be comfortable with us evaluating, even if internally, the quality of their survey.

2. Table with the list of metadata variables to include in the Excel template

This table lists the variables we will include in the Excel template. Please note the following:

- We mark in red the variables we think are essential for all/most surveys.
- The column, “response categories”, has been finalised. In all open-ended questions when a response cannot be provided, please follow the specified coding instructions.

When we have the time and resources, we will: (i) check our list against the DDI standards of variables used by CESSDA and the key survey data archives, and (ii) we will re-format the table and the Excel template to conform to these standards.

1. General identification information about the survey		
Variables	Response categories	Comments
1.0. Country	(Open ended)	Select the appropriate ISO codes for the drop-down menu
1.1. Identification number for the survey in the CA16111 inventory		COST national teams to leave blank; will be assigned centrally by WG1/WG2 team
1.2. Survey acronym	(Open ended) -9 = Don't Know -99 = Information not available -999= Not applicable	If any used by data producers, only; do not make up one
1.3. Full name of the survey - in English	(Open ended)	
1.4. Full name of the survey - in national language(s)	(Open ended)	
1.5. Territorial scope of the survey	1. National 2. Sub-national -9. Don't Know	This variable refers to the territorial level at which the survey is designed to be representative
1.6. If subnational, which level?	(Open ended)	E.g., Local only; several municipalities and a metropolitan area; several provinces; etc.
1.7. If subnational, name of region(s) /city(s) covered by survey - in English	(Open ended)	Add all the cities/regions separated by a semi-colon (e.g., Barcelona province; Madrid province)
1.8. If subnational, name of region(s) /city(s) covered by survey - in national language(s)	(Open ended)	

1.8a. If subnational, add the NUTS/LAU code of the specific cities/regions	(Open ended)	Add all the NUTS (for regions/provinces) or LAU (for municipalities) codes for the cities/regions separated by a semi-colon (e.g., for Barcelona and Madrid in the example above this would be "ES511; ES300) Please note that this could be the LAU code if the territorial level is a local authority /municipality. The files with the LAU codes can be obtained from Eurostat here: https://ec.europa.eu/eurostat/web/nuts/local-administrative-units
1.9. Is the survey designed to be "representative" of the (sub)population in this territory?	1. Yes 0. No -9. Don't Know	The 'representativeness' of a survey is always tricky to ascertain, but here we are asking you to ascertain whether broadly-speaking the sample design is such that the survey will reasonably represent the target population, and/or it claims to be representative of such target population.
1.10. Is it a single cross-section survey or a repeated/longitudinal survey?	1. Single cross-section 2. Repeated cross-section (multiple waves with different samples) 3. Longitudinal/panel survey (multiple waves with the same or partially overlapping samples) 4. Other (mixed design, please specify) -9. Don't Know	Here we are only asking about this specific survey for the given location (i.e., city/region/country) where it is conducted; multi-site surveys can be identified in section 2 below
1.10a. If other in 1.10, specify	(Open ended)	
1.11. Starting date of the survey fieldwork	dd/mm/yyyy -99 = Information not available	For repeated or longitudinal surveys, here you should report only the fieldwork dates of the wave that this register corresponds to. If you do not know the day or the month add 00 where relevant.
1.12. End date of the survey fieldwork	dd/mm/yyyy -99 = Information not available	For repeated or longitudinal surveys, here you should report only the fieldwork dates of the wave that this register corresponds to. If you do not know the day or the month add 00 where relevant.
1.13. Main topic(s) covered by the survey.	<ul style="list-style-type: none"> - Asylum seekers and refugee issues - Citizenship and naturalization - Consumption and/or leisure - Demographic characteristics/behaviours - Discrimination, racism and/or xenophobia - Educational attainment/trajectory, human capital, skills - Family reunification, marriage, family relations - Gender relations, gender identity, sexuality - Housing/housing access - Health/health access - Human smuggling and trafficking - Identity (ethnic, national, racial, religious) and belonging - Income-related (and/or poverty) - Interethnic contact and conflict - Labour market integration - Language skills/training - Leisure, sports, arts - Legal status/administrative situation - Migration trajectory (past/future) 	<p>By 'main topics' we refer to topics that are given "considerable" space in the questionnaire. Of course, this will be relative to the length of the questionnaire, but typically we are referring to a set of at least 3-4 questions covering that topic. In other words, a topic is given some attention in the questionnaires and not just 1 or 2 short questions as generic 'control' variables.</p> <p>This variable is set up as a multi-response, such that several of these options can be marked. Please select the option 'Don't Know' if you are unsure if a main topic is covered by the survey.</p> <p>National teams can only ascertain the topic(s) if they get hold of the questionnaire used.</p>

	<ul style="list-style-type: none"> - Political inclusion and participation, and social/political attitudes - Public attitudes about migration and migrants - Reasons for migration/migration drivers - Return migration - Social cohesion and/or civic engagement and/or networks - Space use/spatial consequences - Time use - Transnational patterns (e.g., remittances, travel, engagement with 'home' country politics, etc.) and diasporas - Unaccompanied migrant minors - Religion - Other 	
1.13a. If other in 1.13, specify	(Open ended)	
1.14. Main purpose of the survey.	<ul style="list-style-type: none"> - Research/academic - Public policy - Government statistics - NGO/non-profit organisation - Commercial - Other 	This variable is set up as a multi-response one, such that several of these options can be marked. Please select the option 'Don't know' if you are unsure if a main topic is covered by the survey,
1.14a. If other in 1.14, specify	(Open ended)	
1.15. What is the coverage of the target population in terms of age?	<ul style="list-style-type: none"> 1. Children (up to 12 years-old) only 2. Youth (between 13 – 25 years-old) only 3. Children and youth only 4. Adult population (18+ or 15+) only 5. Elder population only (55+) 6. A combination of minors and adults -9. Don't Know -99.Information not available 	
1.16. What is the coverage of the target population in terms of the sex of the respondents?	<ul style="list-style-type: none"> 1. Men only 2. Women only 3. Both men and women -9. Don't Know -99.Information not available 	
1.17. Comments relevant to variables in section 1	(Open ended)	Use this section to add any relevant comments relating to variables in section 1. Such fields will be provided in every section.

2. Information about the inclusion of the survey in a larger study		
Variables	Response categories	Comments
2.1. Identification number for the larger Study in the CA16111 inventory		COST national teams to leave blank; will be assigned centrally by WG1/WG2 team
2.2. Study acronym	(Open ended) -9 = Don't Know -99 = Information not available -999= Not applicable	If any used by data producers, only; do not make up one
2.3. Name of the international/larger programme this survey is part of - in English	(Open ended) -999 = Not applicable	
2.4. Name of the international/larger programme this survey is part of - in national language(s)	(Open ended) -999 = Not applicable	
2.5. Name of other country(s)/region(s)/city(s) included in the international/larger programme - in English.	(Open ended)	
2.6. In case of repeated/longitudinal surveys: Date of first survey in the repeated/longitudinal series	dd/mm/yyyy -99 = Information is not available	
2.7. In case of repeated/longitudinal surveys: Frequency of waves/panels	(Open ended) -99 = Information not available	Here we want to know how often (approx.) each wave is undertaken (e.g. yearly, every two years, irregular but every 3-5 years, etc.)
2.8. In case of repeated/longitudinal surveys: Which wave number corresponds to this survey register?	(Number) -99 = Information is not available	Here you should specify if the survey for which you're coding this particular register is for wave 1, wave 2, etc.
2.9. In the case of surveys that are part of an international survey programme: Date from which the survey started to be part of an international survey programme for the first time	yyyy -99 = Information is not available	
2.10. In the case of surveys that are part of an international survey programme: Frequency of waves since then	(Open ended) -99 = Information is not available	Here we want to know how often (approx.) each wave is undertaken (e.g. yearly, every two years, irregular but every 3-5 years, etc.)
2.11. If sample is pooled: how many surveys pooled.	(Open ended)	Pooled surveys should be described separately in individual registers (see definition in guidelines above and please see Section 3.2 below on annotations relating to specific variables)
2.12. If sample is pooled: which other surveys pooled.		Pooled surveys should be described separately in individual registers (see definition in guidelines above and please see Section 3.2 below on annotations relating to specific variables)
2.13. Any qualitative studies linked to the survey?	1. Yes 0. No	Were there any qualitative studies done in connection with the survey?

	-9. Don't Know -99. Information not available	
2.14. If yes, please describe these qualitative studies	(Open ended)	Add here all relevant information about the linked qualitative studies
2.15. Comments relevant to variables in section 2	(Open ended)	

3. Ethnic and migrant minority target population		
Variables	Response categories	Comments
3.1. EMM Target population: which minority group(s)	(Open-ended)	<p>Please list here the best terms that describe the minority groups covered. This could be "immigrants" or "Roma minority" or "non-EU third country nationals", etc. Even if a 'majority' population sub-group is included, here we are only interested in the description of the EMM groups targeted.</p> <p>If you do not know how to complete this section or you do not have the information to fill in this section, please write this down.</p>
3.2. Was the EMM target population....	1. All foreign residents in the city/region/country 2. All residents of foreign origin in the city/region/country 3. All residents who are 1 st or 2 nd generation migrants in the city/region/country 4. All residents of ethnic minority identification in the city/region/country 5. A selection of residents of foreign/migrant origin or ancestry in the city/region/country 6. A selection of residents of ethnic minority identification in the city/region/country 7. Other (e.g., returning emigrants) -9. Don't Know -99. Information not available	Only one should apply
3.2a. If "Other" in 3.2, describe which	(Open ended)	
3.3. How is target population defined / operationalized / identified	- Country of birth of respondent - Country of birth of parents/grandparents - Citizenship/nationality of respondent (current) - Citizenship/nationality of respondent (at birth) - Citizenship/nationality of parents/grandparents (current) - Citizenship/nationality of parents/grandparents (at birth) - Ethnic self-identification of respondent (one	<p>Multiple responses possible. Mark the criterion-criteria used to define minority group(s). If you do not know if a criterion applies, select 'Don't Know'.</p> <p>Oftentimes, the definition of the target population is not offered in the questionnaire as such, but in additional technical documentation, or in core publications relating to the survey. If you have access to them, please check them to see if you find a definition / operationalization of the target population.</p> <p>By citizenship/nationality we mean the actual nationality rights (e.g., to a passport).</p>

	<ul style="list-style-type: none"> response allowed) - Ethnic self-identification of respondent (multiple responses allowed) - Ethnic self-identification of parents/grandparents - Mother tongue/language related question - Classification by interviewer (i.e., the interviewer assigns a subjective value about ethnicity/minority group adscription) - Classification by third agent/proxy (e.g., by a government authority, an NGO, a social/cultural mediator, etc.) - Classification by geographical location - Through other means/characteristics (including non-ethnic/migrant minority-related characteristics) - Information not available on definition of target population 	
3.3a. If "Other means" in 3.3., describe which	(Open ended)	
3.4. Migrant / minority related questions in the questionnaire	<ul style="list-style-type: none"> - Country of birth of respondent - Country of birth of parents - Country of birth of grandparents - Nationality of respondent (current) - Nationality of respondent (at birth) - Nationality of parents (current) - Nationality of grandparents (current) - Nationality of parents (at birth) - Nationality of grandparents (at birth) - Ethnic self-identification of respondent (one response allowed) - Ethnic self-identification of respondent (multiple responses allowed) - Ethnic self-identification of parents - Ethnic self-identification of grandparents - Mother tongue/language related question - Classification by interviewer (i.e., the interviewer assigns a subjective value about ethnicity/minority group adscription) - Information not available on migrant / minority related questions 	<p>Important: this question needs to be answered always, regardless of what you have answered in 3.3. above. Multiple responses possible. In cases when you do not know if a criterion applies, please select 'Don't Know'.</p> <p>When designing a survey, the researchers might have a core set of variables in mind to identify the target EMM population but can nevertheless add other relevant variables relating to migrant/minority status. Even if they are not used to define the target groups, please report whether the survey includes each question in the list.</p> <p>It may be sometimes that these questions are implicitly or explicitly included as part of the screening section of the questionnaire or as part of a general introduction to the survey where the respondent can say (s)he does not fit the profile searched. In those cases, mark as if these questions were included.</p>
3.5. Size of the EMM target population/s in the country/region/city	(Open ended -99 = Information not available)	Here we would like to get the information about the size of the population(s) on which the sample is likely to have been drawn. Typically, you would know this from either census information or local

as a whole (if available)		population registers, which are used as the sampling frame. If this is not available because the sample frame does not provide a count of the underlying population (e.g., the sample is addresses instead of individuals), write '-99' for information not available.
3.6. Does the survey include a sub-group of majority/autochthonous/general population?	1. Yes 0. No -9. Don't Know	
3.7. Comments relevant to variables in section 3:	(Open ended)	

4. Sampling method		
Variables	Response categories	Comments
4.1. Sampling strategy – closed	1. random sampling/selection (i.e., probability sampling, of some kind) 2. non-probability sampling (including snowball/network and purposive samplings) 3. mixed sampling procedures (in which there are elements of probability sampling, such as centre-location sampling) -9. Don't Know -99. Information not available	Please choose which category describes best the sampling strategy. Please see Section 3.2 below on annotations relating to specific variables.
4.2. Sampling strategy – open ended	(Open ended) -9 = Don't Know -99 = Information not available	Important: this question needs to be answered even if you have already answered 4.1. Add specifics that allow to understand better your answer in 4.1. If the answer is unknown, please enter in '-9'. If the information is not available, please enter in '-99'. Sample design strategy: for example, stratified random sample (and describe which strata and if oversample or proportional to population size), quota sampling, random route sampling, etc. Describe specific sample design technique: (open-ended field to describe with some detail). Please see Section 3.2 below on annotations relating to specific variables.
4.3. Sample design – full additional information	(Open ended)	Please add all the essential information for a reader to understand how the sample was designed without referring to the survey documentation.
4.4. Sampling frame(s)	(Open ended) -9 = Don't Know -99 = Information not available	This is the list, census of the population or addresses, or rolling population register/list of households from which sample extracted. Please also add the name of the list as such in the original country language for precise identification.

4.5. Sampling units	(Open ended) -9 = Don't Know -99 = Information not available	Individuals? Families? Households? Addresses? Census sections? etc. Please provide information on ALL sampling units used in the design (primary, secondary, etc.)
4.6. Comments and additional information on the sampling method	(Open ended)	Use this section to expand on any relevant details about sampling, especially if the documentation is partial and you can only provide information about effective population covered. Please see Section 3.2 below on annotations relating to specific variables.

5. Sample size for the overall survey		
Variables	Response categories	Comments
5.1. Total gross/issued sample	(Number) -9 = Don't Know -99 = information not available	Here we are interested in the full sample for this particular survey (wave). If this is a general population survey, you should include the overall sample size for all respondents. If this is a survey broken down by sub-samples to specific groups, we want here the overall number of respondents targeted across all groups.
5.2. Total net/achieved sample	(Number) -9 = Don't Know -99 = information not available	As above
5.3. Overall response rate	(Number) -9 = Don't Know -99 = information not available	Please provide the response rate reported by the data producers in the first instance (see Section 3.1 below on annotations relating to specific variables)
5.4. How was overall response rate calculated	1. By data producers with no mentioned formula; 2. AAPOR; 3. AAPOR RR1; 4. ESS; 5. Other -9. Don't Know -99. Information not available	
5.5. If response rate calculated as "5=Other", please describe	(Open ended)	Give as many details as possible about how response rates were calculated
5.6. Comments about any known issues with the sample and representativeness or biases	(Open ended)	Fill in only if available in documentation or existing research relating to this survey. We do not expect you to run your own analysis about the quality of the sample
5.7. Are any weights provided?	1. Yes 0. No -9. Don't Know -99. Information not available	
5.8. If there are any weights, please	(Open ended)	

describe		
5.9. Comments relevant to variables in section 5	(Open ended)	

6. Sample sizes for any subgroups in which the survey is partitioned		
Variables	Response categories	Comments
Subgroup #1 (= SG1)	<p>We will provide at least 5 repetitions of this battery of questions for up to 5 subgroups in which the survey is partitioned (6a to 6e).</p> <p>These variables only need to be filled in if the sample was partitioned by groups (e.g., 3-4 nationalities for migrant groups; 2-3 ethnic minorities; migrants vs majority population; etc.). Please see section 3.3 below in this Word document for clarifications.</p> <p>Important: if a survey is designed as a multi-city one but the samples are designed separately for each city (e.g., from different sample frames for each city), a separate record should be created for each city. Please see section 3.2 below on further annotations in this Word document for clarifications.</p> <p>Please sort by the achieved sample size.</p>	
6a1. SG1 name of sub-group	(Open ended)	<p>Write here the name of this subgroup as defined in the survey documentation.</p> <p>For example: "Turkish migrants", "Autochthonous population", "Refugees".</p>
6a2. SG1 gross/issued sample	(number) - 99 = Information not available	
6a3. SG1 net/achieved sample	(number) - 99 = Information not available	
6a4. SG1 response rate	(number) - 99 = Information not available	Please provide the response rate reported by the data producers in the first instance
6a5. SG1 How was overall response rate calculated	1. by data producers with no mentioned formula; 2. AAPOR; 3. AAPOR RR1; 4. ESS; 5. Other -9. Don't Know -99. Information not available	

6a6. SG1 If response rate calculated as "5=Other", please describe	(Open ended)	Give as many details as possible about how response rates were calculated
6a7. SG1 Comments about any known issues with the sample and representativeness or biases	(Open ended)	Fill in only if available in documentation or existing research relating to this survey. We do not expect you to run your own analysis about the quality of the sample
6b1-6b7 + 6c1-6c7 + 6d1-6d7 + 6e1-6e7 to be added in Excel file replicating above structure and replacing SG1 for SG2 - SG5		
6.8 Comments relevant to variables in section 6		

7. Data collection information		
Variables	Response categories	Comments
7.1. Name of person/institution/institute that undertook fieldwork	(Open ended) - 99 = Information not available	Please add here the full name of the institution(s) or survey institute(s) that carried out the fieldwork. This might be a university where the fieldwork was organised in-house, or a professional survey company, or a single individual (e.g., for surveys done in the context of PhD dissertations)
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 - Information not available on data collection mode	Multiple options possible. If the survey is mixed-mode, select 'Yes' for all that apply. If you are unsure if a collection mode was used or not, select 'Don't Know'. If you do not have any information on the data collection mode(s), select 'No' for all data collection modes and select 'Yes' for information not available.
7.3. For personal interviews: who interviewed?	1. Professional interviewers only 2. Cultural mediator only 3. Non-professional interviewers (e.g., students) 4. A mix -9. Don't Know -99. Information not available	
7.4. For personal interviews: interviewers spoke migrant or ethnic minority group languages?	1. Yes 2. No, but translator(s) present/available	

	3. No, nobody had targeted language skills -9. Don't Know -99. Information not available	
7.5. If yes, which?	(Open ended)	For later coding into international standard codes for languages (ISO)
7.6. Questionnaire available in languages of migrants or ethnic minorities other than official languages?	1. Yes 0. No -9. Don't Know -99. Information not available	
7.7. If yes, which?	(Open ended)	For later coding into international standard codes for languages (ISO)
7.8. Average duration/length of interview	(In minutes) -9 = Don't Know -99 = Information not available	
7.9. Number of questions in the questionnaire	(Number) -9 = Don't Know -99 = Information not available	By 'questions', we mean the number of 'question items' (i.e., every time the respondent needs to provide an answer to something). Thus, a question that asks for the opinion on (e.g.) 4 different institutions counts for 4. If you have access to the dataset, oftentimes the quickest way to establish the number of question items is to open the dataset with your statistical software of choice and query the software for the total number of variables in the dataset. If you are unable to tell what the number of questions is (e.g., for CAPI surveys, or dataset not available) and cannot even give a sensible approximation, please code as -9. If you do not have access to the information needed to respond to this question, please enter -99.
7.10. Comments relevant to variables in section	(Open ended)	Please add any further comment on data collection that could be relevant to users

8. Availability		
Variables	Response categories	Comments
8.1. Availability of the survey data matrix (the dataset with the individual responses) for individual research use	1. Yes, publicly available 2. Available through a COST Action member 3. Available by request 4. Unavailable 5. Unknown availability	
8.2. If publicly available or by request, where is the survey dataset stored	(Open ended)	If in a repository, please provide a URL for the respective data catalogue page, otherwise, describe as text who to contact

8.3. Identification number of the archive/repository where the dataset is stored	(Open ended)	If data are already in a repository, please provide the identification number or code with which the dataset can be found in the archive. If no identification number is available, and a name is provided instead, please add the name.
8.4. DOI for the dataset	(DOI format)	If available, write in DOI format
8.5. Access to complete dataset	1. Yes, micro-data available for download/direct access by researchers 2. Yes, but micro-data only available for online analyses 3. Yes, but micro-data only available to analyse on a secure office/premises of the data archive or data producer 4. Yes, but with other restrictions 5. No -9. Don't Know	Here we want to have the information about how the data can be accessed, if more information is relevant, please add in the Observations field at the end of this section.
8.6. Access to portions of the dataset	1. Yes 0. No -9. Don't Know -999. Not applicable (full dataset accessible)	Important: In the case of a survey which contains sensitive data, these data might be removed from the standard access and only available through a more restricted access, or removed completely; in other cases, researchers might choose to share only part of the dataset due to other confidentiality reasons.
8.7. Access to aggregate data results (e.g., via online databases)	1. Yes 0. No -9. Don't Know	
8.8. If there are any restrictions for data access in 8.5, 8.6 and 8.7, describe which	(Open ended)	
8.9. Language(s) in which dataset is available	(Open ended)	For later coding into international standard codes for languages (ISO)
8.10. Availability of the technical survey documentation for individual research use	1. Yes, publicly available 2. Available through a COST Action member 3. Available by request 4. Unavailable 5. Unknown availability	
8.11. If available, where is the technical survey documentation stored	(Open ended)	If in a repository, please provide a URL for the respective data catalogue page, otherwise, describe as text who to contact
8.12. Identification number of the archive/repository where technical documentation is stored		If data already in a repository
8.13. DOI for the documentation	(DOI format)	If available, write in DOI format
8.14. If the technical documentation is available, standard used for the documentation:	1. Data Documentation Initiative 2. Dublin Core 3. SDMX 4. No specific standard -9. Don't Know	
8.15. Language(s) in which technical survey documentation	(Open ended)	For later coding into international standard codes for languages (ISO)

is available		
8.16. Availability of the survey q'aire for individual research use	1. Publicly available 2. Available through a COST Action member 3. Available by request 4. Unavailable 5. Unknown availability	
8.17. If available, where is the survey q'aire stored	(Open ended)	If in a repository, please provide a URL for the respective data catalogue page, otherwise, describe as text who to contact
8.18. Identification number of the archive/repository where questionnaire is stored		If questionnaire already in a repository
8.19. DOI for the questionnaire	(DOI format)	If available
8.20. Language(s) in which survey questionnaire is available	(Open ended)	For later coding into international standard codes for languages (ISO)
8.21. Comments relevant to variables in section 8: on data availability, including any relevant legal/copyright requirements or restrictions for data access	(Open ended)	Please add any further comment on data availability that could be relevant to users. Important: If the data includes restrictions for further sharing of the dataset such that individual researchers can download/access the data but cannot share it further on, please mention it here.

9. Data producers, owners, distributors, and citations		
Variables	Response categories	Comments
9.1. Institution/team responsible for data production	(Open ended)	This is typically the team or institution/survey institute that undertook fieldwork
9.2. Institution/team that owns the data	(Open ended)	This is typically the team/institution that paid for the survey and holds copyrights/intellectual property rights. It can be more than one (e.g., if paid by a government unit but led by a research institution), and if so, add here both.
9.3. Institution/team distributing the dataset	(Open ended)	The institution/team in charge of sharing the data with future users
9.4. Contact details for queries/requests	(Open ended)	Who to contact to request access to the data, or for redistribution
9.5. Citation for dataset	(Open ended)	Add suggested dataset citation here, as a text field. Many repositories or data producers have such a suggestion
9.6. Citation for technical documentation	(Open ended)	Same as above
9.7. Citation(s) for any other publications	(Open ended)	Same as above
9.8. Comments	(Open ended)	

10. Additional Information		
Variables	Response categories	Comments
10.1. General subjective assessment of the data quality	From 0 "Extremely poor" to 10 "Extremely good" -9 = Don't Know	<p>This is a subjective assessment by the COST national teams and is intended as a guideline of whether future efforts of making the data available or pooling it with other datasets are worth the while. It will NOT be made public in this format with the metadata template. It will only be used for internal COST decision-making about prioritization of next steps with questionnaires and post-harmonization. Quality should be judged, primarily, on the basis of the availability of the relevant documentation, the transparency of the methods used, and the a priori appropriateness of the methods used. These are typically judgements about the 'input/process' of the survey.</p> <p>If the person coding the metadata template has used the dataset, then "output" quality (e.g., quality of the responses, response rates, etc.) can also be used to make this judgement.</p>
10.2. Additional Information about quality of the data	(Open ended)	Please add here any comment you think is relevant about the data quality
10.3. Sources of Information	(Open ended)	Add here the sources from which you obtained the information you have added in this template.
10.4. Any other comments about the survey	(Open ended)	Add any other comments you think are relevant and are not covered elsewhere

3. More detailed annotations and clarifications on some variables that require additional explanations

3.1. *Response rates*

How to report on response rates? Please fill in the response rate as it is documented in the survey technical report, specifying, where possible, the calculation formula used to report on response rates (i.e., AAPOR). You are not expected to calculate response rates yourselves unless you find that all necessary information for such calculations is provided in the documentation and that, for some reason, the data producers/distributors did not provide a response rate figure.

3.2. *Information on sampling*

Sample strategies and designs can differ considerably across cases, so we want each national team to provide as much information as they can about each survey so that any external reader can roughly make sense of how the sample was designed and actually implemented.

We consider one sample every territorial unit or (sub)target population group that was approached with a different sampling design or strategy, whether within a country or across countries.

For example, in cross-national surveys, each country typically constitutes as a separate sample because the sampling frames differ, and the sample design differs. But also, for example, within a single country, the minority population in city A might have been approached with a given sampling strategy/design that differs from that used in city B. In that case, the samples for city A and city B would be two different samples and need to be logged as two different registers in our database, even if they were conducted within the same national study. The data for city A and city B will normally be pooled within a single study or dataset and this is what we refer to as "pooled samples" in the sampling section.

3.3. *What counts as a sub-group*

We only consider as a sub-group those ethnic or migrant minority groups for which an explicitly different sample design/strategy, a well-defined quota or a target achieved sample was specified. In other words, we are not requesting that you break down the information for all the nationalities or ethnic groups that were actually covered but only for those where a separate/partitioned sample design or target was established a priori.

For example, if a survey to all ethnic minorities in country A has established a designed/issued sample size of 700 respondents, and no quota or specific sample design was established for the various ethnic minorities in the country, then you do not need to fill in any of the sub-group variables.

In another example, a survey was designed in country B to target: (i) an issued sample of 500 respondents who were born in a foreign country and (ii) an issued sample of 500 respondents who were born in country B. This survey has been designed by stratifying across these two groups (based on the country of birth) and

hence sub-group 1 consists of foreign-born respondents and sub-group 2 consists of country-born respondents.

In a final example, in city C a survey was designed to target 300 respondents born in Morocco, 300 respondents born in Ecuador, 300 respondents born in other Andean Latin American countries, and 300 respondents born in Spain. In this case, the survey has 4 sub-groups, and you should fill in the information relating to the variables for each of the 4 sub-groups.

Sub-groups can be defined in many ways: by nationality, national/country origin, ethnicity, generation, gender, country of residence, etc. For example, if a survey in country A includes a sub-group of EMM respondents who reside in country A and a comparison/control group of EMM respondents who never migrated to country A but are of the same country/ethnicity as those who reside in country A, these two are considered sub-groups for our purposes.

4. Technical notes

4.1. Saving format

Please make sure that the template is saved in the right Excel-format. The required suffix is “.xlsx”, which is the same format used in the blank template. Please also include the country name in the title of your Excel template.

4.2. A new column for each new survey

Each new survey should be entered into a new column on the appropriate sheet (i.e., “National surveys” or “Subnational surveys”). In the end, **every country will send one Excel template** with all of the identified surveys. See figure below for an example:

	A	B	C	D	E
1		SURVEY 1	SURVEY 2	SURVEY 3	...
2	1. General Identification Information about the Survey				
3	1.0. Country				
4					
5	1.1. ID number (leave blank)				
6					
7	1.2. Acronym				
8					
9	1.3. Survey Name Eng.				
10					
11	1.4. Survey Name Nat.	National Survey	Panel Survey	Health & Education Survey	
12					
13	1.5. Scope of survey	1.National	1.National	1.National	
14					
15	1.9. Representative of the population	1.Yes	0.No	1.Yes	
16					
17	1.10. Type of survey	1.Single cross-section	3.Longitudinal/panel survey	multiple waves	
18	1.10a. If "other" in 1.10. specify				
19					
20	1.11. Starting date of survey				
21					

CHECKLIST - ETHMIGSURVEYDATA METADATA COLLECTION ABOUT NATIONAL (WG1) and SUBNATIONAL SURVEYS (WG2)

Please fill this out BEFORE sending in the completed/final Excel template (deadline: at the latest 10 February 2019)

THINGS TO CHECK...

1. Do the national and subnational metafiles only contain data about surveys specifically addressed to ethnic minority/migrant populations or those in which these groups constitute a significant sub-sample?

☐

Tip: Only information about population surveys should be included, not about social statistics, censuses, etc (see sections 1.1-1.5 of this Word document)

2. Have you filled in the information required in the “Annotations” sheet?

☐

Tip: ALL QUESTIONS in the sheet need to be answered (see section 1.6 of this Word document)

3. Have you used one column to input the information for each survey?

☐

Tip: this applies both to the “National surveys” and “Sub-national surveys” sheets. DO NOT ADD EXTRA SHEETS TO THE EXCEL FILE.

4. Have you coded the required information in sections 1-10 of the Excel template (i.e., the questions marked in red) for each survey you have included?

☐

Tip: if the information required is not available, please follow the instructions provided (e.g., selecting the “Don’t Know” option). Use the “comments” section to give us more insights if necessary.