



HAL
open science

Guidelines Document: Metadata Template Version Control

Laura Morales, Ami Katherine Saji

► **To cite this version:**

Laura Morales, Ami Katherine Saji. Guidelines Document: Metadata Template Version Control. 2021. hal-03845803

HAL Id: hal-03845803

<https://sciencespo.hal.science/hal-03845803>

Preprint submitted on 22 Feb 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Guidelines Document: Metadata Template Version Control

****INTERNAL USE ONLY****

Written by: Ami Saji and Laura Morales

Last updated: 01 September 2021

Version: 0.1

Acknowledgments

This guidelines document, which details how metadata compiled using an Excel-based metadata template should be version controlled, has been jointly produced by the H2020 project, [SSHOC \(Social Sciences and Humanities Open Cloud\)](#), and [COST Action 16111 – ETHMIGSURVEYDATA](#).

SSHOC



SSHOC, “Social Sciences and Humanities Open Cloud”, has received funding from the EU Horizon 2020 Research and Innovation Programme (2014-2020); H2020-INFRAEOSC-04-2018, under the agreement No. 823782.

The content of this guidelines document does not represent the opinion of the European Commission and the European Commission is not responsible for any use that might be made of such content.

COST Action 16111 - ETHMIGSURVEYDATA

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



Table of Contents

Introduction	4
1. Overview of the Quality Control Process	5
2. First Stage Quality Control	9
3. Second Stage Quality Control	9
4. Final Quality Control	9
5. Data Cleaning	18
Conclusion	27
Appendix A: Structure and Use of the Team Google Drive folder, COST WG1_WG2 docs	28
Appendix B: Sample Word Document (Template Feedback for First and Second Stage Quality Control)	31
Appendix C: Sample Emails (Template Feedback for First and Second Stage Quality Control)	34
Appendix D: Sample Word Document (Template Feedback for Final Quality Control)	37
Appendix E: Sample Emails (Template Feedback for Final Quality Control)	40
Appendix F: Protocol for Coding Variables with “Missing” Responses	43
Appendix G: Standardization Rules for Proper Nouns	55

Introduction

A core responsibility for Working Group (WG) 1 and 2 of COST Action 16111 - ETHMIGSURVEDYATA is to compile metadata on existing surveys to ethnic and migrant minorities (EMMs) in each of the [35 COST Action participating countries](#). As such, national representatives for each of these 35 countries have searched for, collected, and documented surveys to EMMs by: (i) using an Excel metadata template developed by the WG1 and 2 leaders ([FINAL ETHMIGSURVEYDATA Metadata Template.xlsx](#)) and (ii) following the procedures and rules outlined in the *Surveys Metadata Template Guidelines Document* ([FINAL CA16111 ETHMIGSURVEYDATA WG1 WG2 Metadata Template.pdf](#)).

To ensure that the national representatives have correctly coded metadata (for their respective identified surveys) on to their template, each template needs to be regularly reviewed by the WG1 and WG2 leaders and the Sciences Po team (i.e. Laura Morales— Action Chair of ETHMIGSURVEYDATA and Task Lead of Task 9.2 of SSHOC,¹ Ami Saji— a junior researcher funded by SSHOC and supporting Task 9.2, Network Coordinator of the COST Action, Short-term Scientific Mission (STSM) participants with ETHMIGSURVEYDATA,² and research assistants (RAs) hired to work at Sciences Po and funded via SSHOC). **This guidelines document, therefore, outlines how each template should be reviewed and handled to ensure proper version control and is structured into the following sections:**

1. Overview of the Quality Control Process: A detailed explanation of how the overall quality control process is structured
2. First Stage Quality Control: A step-by-step guide on when and how to conduct the first stage quality control of a template
3. Second Stage Quality Control: A step-by-step guide on when and how to conduct the second stage quality control of a template
4. Final Quality Control: A step-by-step guide on how to conduct the final quality control of a template
5. Data Cleaning: A step-by-step guide on how to clean the template, so it can be deposited to the EMM Survey Registry and be used for analysis

For anyone contributing to and assisting with the version control of the templates, they must first read this guidelines document to ensure that they are executing and completing the steps correctly. Any questions about the version control of the templates can be directed to Laura Morales (laura.morales@sciencespo.fr) and Ami Saji (amikatherine.saji@sciencespo.fr).

¹ The [SSHOC \(Social Sciences and Humanities Open Cloud\) project](#) is a large-scale and collaborative initiative to develop and launch a new cloud-based infrastructure, where data, tools, and training will be made available to researchers and scholars in the social sciences and humanities communities. Within SSHOC, Task 9.2 of Work Package 9 is dedicated to the Ethnic and Migration Studies Data Community.

² The STSM participants of ETHMIGSURVEYDATA, who are hosted at Sciences Po (via Laura Morales), often develop a work plan based on supporting the survey metadata compilation work.

1. Overview of the Quality Control Process

Each of the 35 COST Action participating countries are required to submit their template for review to ensure quality control. The quality control process for the templates consists of the following steps (see [Figure 1](#) for a visual representation):

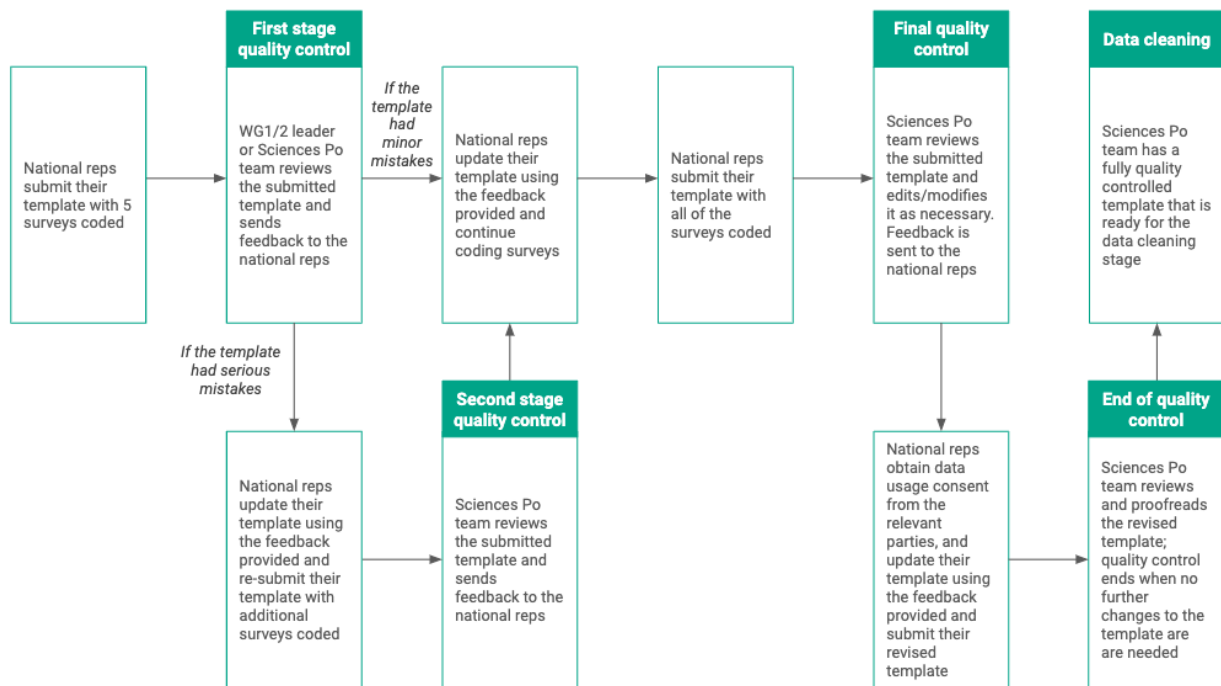
1. The national representatives of the respective COST Action participating country codes 5 surveys onto the Excel metadata template and submits it for an initial quality control (i.e. first stage quality control). The target completion date for this step was set for 10 January 2019.
2. A WG1 or 2 leader or a Sciences Po team member completes the first stage quality control by reviewing the submitted template and providing (via email) detailed feedback to the national representatives. The target turnaround time is within 2-3 weeks of the template being submitted for review.
3. The national representatives update their template using the feedback provided. If their template had serious mistakes (which is communicated to them by the reviewer), the national representatives code additional surveys onto the template and submit their template for a second round of quality control (i.e. second stage quality control).
4. The national representatives finish coding all of the surveys found and submit the final version of their template for review (i.e. final quality control). The target completion date for this step was set for 30 April 2019.
5. A Sciences Po team member conducts the final quality control of the submitted final template by reviewing and making any straightforward edits/modifications to it; for any edits/modifications that require the national representatives' input and clarification, the Sciences Po team member contacts them directly to request their assistance. The Sciences Po team member also liaises with the national representatives to obtain consent from the relevant survey producers/distributors/owners³ to store and reuse the survey metadata. Once the template has completed the final quality control process, including the steps where the national representatives are involved, the template can progress to the data cleaning stage (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 ready to be uploaded onto the EMM Survey Registry and be used for analysis).

As every template needs to complete the full quality control process, the Sciences Po team is responsible for guiding each of the 35 COST Action participating countries through each of the above steps. If any country is delayed or not progressing with the process, the Sciences Po team will be responsible for following up with the national representatives of that country.

³ Explicit consent to store and use the survey metadata is needed if the survey has been deposited to a data archive/repository; in such a case, the data archive/repository will be asked to provide this consent. For surveys that are not publicly available through a non-data archive/repository platform, implicit consent to store and use the survey metadata is needed from the survey owner (e.g. independent researcher/research team).

The Sciences Po team and the WG1 and WG2 leaders are keeping copies of the received templates (i.e. templates submitted by the national representatives for the first, second, and/or final quality control) in the following shared Sciences Po-based Team Google Drive folder: *COST WG1_WG2 docs* ([Appendix A](#)).⁴ Templates that have been quality controlled (all stages) are also being housed in the above-mentioned folder, with those that have completed the final quality control process being backed up in a Google Drive folder, *A2. Final Metadata Quality Control (COPY)*, which is only accessible to authorized users. For all these various aforementioned folders, Ami Saji is responsible for managing them and keeping them up to date.

Figure 1: Quality Control Workflow



⁴ As this folder is set up as a Team Google Drive, a shareable link cannot be provided. Instead, users (i.e. team members) are added to the folder, so they can edit/upload/access files (including the received templates). Currently, only the WG1 and WG2 leaders and the Sciences Po team members have access to this folder.

2. First Stage Quality Control

The first stage quality control, which is when a template is reviewed for the first time, begins with the national representatives submitting their template with at least 5 surveys to the WG1 and WG2 leaders and the Sciences Po team. A specific individual (i.e. a WG1 or WG2 leader, or a Sciences Po team member) is then tasked with reviewing the submitted template and providing detailed feedback to the national representatives.

To provide correct and appropriate feedback, **the assigned reviewer should first (re)read the *Surveys Metadata Template Guidelines Document* ([FINAL CA16111 ETHMIGSURVEYDATA WG1 WG2 Metadata Template.pdf](#))** to ensure they are fully familiar with how national representatives have been asked to code surveys onto the template. **After reading this document, the assigned reviewer should complete all of the following steps:**

1. Open the Word template for the template feedback ([Template Feedback](#)) and create a new Word document entitled *Template Feedback Country Name*. This newly created Word document is where all the reviewer's feedback should be documented (see [Appendix B](#) for an example of how this document should be filled out by the reviewer).
2. Go to the sub-folder, *2. 5 Surveys Metadata*, of *COST WG1_WG2 docs*. Download and open the template submitted by the national representatives (an Excel file) and rename it as, *Country Name (Submission Date)⁵_5 Surveys_FINAL ETHMIGSURVEYDATA Metadata Template_Reviewed (Reviewer's Initials).xlsx*, with the "Submission date" formatted as *dd.mm.yy*.
3. Check that the national representatives used the correct version of the template (i.e. [FINAL ETHMIGSURVEYDATA Metadata Template.xlsx](#)). [Note: Prior to the release of the final version of the template, a pilot version was circulated to the national representatives in spring-summer 2018. Moreover, as a companion to the final version of the template, another Excel file ([FINAL ETHMIGSURVEYDATA Metadata Template\(Changes Marked\).xlsx](#)) was created to show/mark all of the changes made from the pilot to final version of the template.⁶]
 - a. If the correct template was used, proceed to step 4.
 - b. If the pilot version of the template was used, notify Ami Saji so that the national representatives can re-code all the surveys onto the correct version. The template submitted by the national representatives cannot be reviewed for quality control at this point.
 - c. If the template called, [FINAL ETHMIGSURVEYDATA Metadata Template\(Changes Marked\).xlsx](#), was used, copy all of the coded surveys onto the correct template ([FINAL ETHMIGSURVEYDATA Metadata Template.xlsx](#)) and provide a comment

⁵ This is the date that is already listed on the template submitted by the national representatives.

⁶ While the national representatives received several communications about using the final version of the template (i.e. [FINAL ETHMIGSURVEYDATA Metadata Template.xlsx](#)) for compiling metadata on existing surveys to EMMs, it is possible that the pilot or changes-marked version of the template was accidentally used.

about this in the *General Comments* section of the above mentioned Word document. The reviewer can then proceed to step 4.

4. Review the *ANNOTATIONS* sheet of the template and check that the national representatives correctly coded responses for all the variables following the instructions provided in the *Surveys Metadata Template Guidelines Document*. For any variable with a missing or unclear response, provide a comment (i.e. why the variable needs to be filled out and/or how the national representatives can correct their mistake) in the *Annotations sheet* section of the Word document and highlight the corresponding cell (of the template) in yellow (see [Appendix B - Montenegro Template \(5 Surveys\).xlsx](#)⁷ for an example).
5. Review the *NATIONAL SURVEY* and *SUBNATIONAL SURVEY* sheets of the template. For each survey, check that the national representatives: (i) provided responses for the compulsory variables (i.e. those in red text), (ii) correctly coded the responses following the instructions provided in the *Surveys Metadata Template Guidelines Document*, and (iii) provided responses that are generally logical and consistent (e.g. if they provide a response rate calculation, they should have also provided a response rate). For any variable with a missing or unclear response, provide a comment (i.e. why the variable needs to be filled out and/or how the national representatives can correct their mistake) in the *National Survey Sheet/Subnational Survey Sheet* section of the Word document and highlight the corresponding cell in yellow. [NOTE: Do not add any new variables during this stage of quality control. Specifically, this means that no extra variables can be added in section 6 to accommodate for more minority groups as sub-samples, as this has technical implications that we need to avoid. If the national team has mentioned that more sub-groups than the 5 for which there is space were sub-sampled, then please ask the national team to add as much information as possible on the corresponding cell for variable 6.8. Hence, there should be no more variables between 6e7 and 6.8 and if either the national team or anyone else has added more, they must be deleted, and the information contained in them should be added to the corresponding cell of 6.8.]
6. Once the *ANNOTATIONS*, *NATIONAL SURVEY*, and *SUBNATIONAL SURVEY* sheets of the template have been fully reviewed, provide, in the *General Comments* section of the Word document, a 1-2 sentence description of the overall quality of the template, including an explanation of what the general mistakes/errors were. Also provide a brief description of how the feedback (Word document plus Excel file with the highlighted cells) has been structured.
7. Proofread the Word document and Excel file with the highlighted cells. Double check that the comments provided in the Word document are clear and that they correspond to a highlighted cell in the Excel file.
8. Email the finalized feedback (Word document and Excel file with the highlighted cells) to the national representatives and CC Ami Saji (amikatherine.saji@sciencespo.fr) and the Sciences Po team email (sshoc.project@sciencespo.fr). In this email: (i) explain what the Word document and Excel file are; (ii) remind the national representatives to use the feedback provided to update/edit the already coded surveys and to inform how they code

⁷ For security purposes, the link to this file has been removed in this published version of this guidelines document.

new surveys; (iii) advise the national representatives that the reviewer, the WG1 and WG2 leaders, and the Sciences Po team are available to answer any questions about the feedback; and (iv) ask the national representatives to provide an estimated submission date for the final template. For any template with serious mistakes (i.e. many of the compulsory variables were missing responses without any explanation and/or the variables were regularly coded incorrectly), also advise the national representatives that they will need to submit another template for quality control, before they submit the final version of their template. A sample email on how to send the finalized feedback can be found in [Appendix C](#). Once this email has been sent by the reviewer, Ami Saji or the reviewer will upload the finalized feedback onto the Team Google Drive folder, *COST WG1_WG2 docs*, under the sub-folder, *3. 5 Surveys Metadata Quality Control*.

3. Second Stage Quality Control

A template will only undergo a second stage of quality control if it was deemed as having serious mistakes during the first stage of quality control. As indicated in the section above, the national representatives will be informed of this when receiving feedback for the first stage of quality control.

Only a Sciences Po team member will conduct a second stage quality control and they will follow the steps outlined in the section above (i.e. [First Stage Quality Control](#)).

For this quality control stage, the template submitted by the national representatives can be found on the Team Google Drive folder, *COST WG1_WG2 docs*, under the sub-folder, *4. 2nd Submission Metadata*. The Word feedback document should be labelled as, *2 QC Template Feedback Country Name*, and the Excel feedback file should be labelled as, *Country Name (Submission Date)_2 QC_FINAL ETHMIGSURVEYDATA Metadata Template_Reviewed (Reviewer's Initials).xlsx*, with the "Submission date" formatted as *dd.mm.yy*. Ami Saji will upload both feedback items onto the Team Google Drive folder, *COST WG1_WG2 docs*, under the sub-folder, *5. 2nd Submission Metadata Quality Control*.

4. Final Quality Control

The objective of the final quality control is to prepare the template for the data cleaning stage (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 ready to be uploaded onto the EMM Survey Registry and to be used for analysis). The final quality control is consistent with the first and second stage quality control processes, but with the added requirements of the reviewer having to rigorously check that the metadata compiled for each survey is logical/coherent AND having to make direct edits/modifications to the template whenever feasible. Below are all the steps the reviewer will need to complete:

1. Open the Word template for the template feedback ([Template Feedback](#)) and create a new Word document entitled *Final Template Feedback* **Country Name**. This newly created Word document is where the reviewer will document any responses that require follow up and/or clarification by the national representatives, as well as any general questions and/or concerns the reviewer may have about the surveys coded onto the template (see [Appendix D](#) for an example of how this document should be filled out by the reviewer).
2. Go to the sub-folder, 7. *Final Metadata*, of *COST WG1_WG2 docs*. Download and open the template submitted by the national representatives (an Excel file) and rename it as, **Country Name (Submission Date⁸)_FINAL ETHMIGSURVEYDATA Metadata Template_Reviewed (Reviewer's Initials).xlsx**, with the "Submission date" formatted as *dd.mm.yy*.
3. Check that the national representatives used the correct version of the template (i.e. [FINAL ETHMIGSURVEYDATA Metadata Template.xlsx](#)). [Note: Prior to the release of the final version of the template, a pilot version was circulated to the national representatives in spring-summer 2018. Moreover, as a companion to the final version of the template, another Excel file ([FINAL ETHMIGSURVEYDATA Metadata Template\(Changes Marked\).xlsx](#)) was created to show/mark all of the changes made from the pilot to final version of the template.⁹]
 - a. If the correct template was used, proceed to step 4.
 - b. If the pilot version of the template was used, notify Ami Saji so that the national representatives can re-code all the surveys onto the correct version. The template submitted by the national representatives cannot be reviewed for quality control at this point.
 - c. If the template called, [FINAL ETHMIGSURVEYDATA Metadata Template\(Changes Marked\).xlsx](#), was used, copy all of the coded surveys onto the correct template ([FINAL ETHMIGSURVEYDATA Metadata Template.xlsx](#)) and provide a comment about this in the *General Comments* section of the above mentioned Word document. The reviewer can then proceed to step 4.
4. Review the *ANNOTATIONS* sheet of the template and check that the national representatives correctly coded responses for all the variables following the instructions provided in the *Surveys Metadata Template Guidelines Document*.
 - a. For any variable that has a response with a typing/data entry,¹⁰ or a small/obvious spelling, grammatical, and/or punctuation error,¹¹ correct the mistake directly onto the template and highlight it in pink. If the response provided is too difficult to correct (i.e. requires extensive corrections or editing), please do not touch this cell and highlight it instead in red; such cells will be handled by Ami Saji or another individual who is able to perform a linguistic check.

⁸ This is the date when the national representatives submitted the final version of the template to the Sciences Po team.

⁹ While the national representatives received several communications about using the final version of the template (i.e. [FINAL ETHMIGSURVEYDATA Metadata Template.xlsx](#)) for compiling metadata on existing surveys to EMMs, it is possible that the pilot or changes-marked version of the template was accidentally used.

¹⁰ This refers to any typing or data entry errors that can be easily corrected by the reviewer. Examples include changing a N/A response to -999. Not applicable (e.g. variable 1.2) or removing the % symbol from a response rate (e.g. variable 5.3).

¹¹ For spelling, grammar and punctuation, UK or US writing conventions can be used.

- b. For any variable with a missing or unclear response, provide a comment (i.e. why the variable needs to be filled out and/or how the national representatives can correct their mistake) in the *Annotations sheet* section of the Word document and highlight the corresponding cell (of the template) in yellow.

An example of how the template should be highlighted in yellow, pink, and red can be found here: [Appendix D - Norway Template \(Final\).xlsx](#).¹²

- 5. To the template, add the new metadata variables listed below to the *CODES- DO NOT CHANGE OR DELETE* sheet first (see [Codes for New Metadata Variables.xlsx](#)¹³ for the codes to be used) and then update the *NATIONAL SURVEYS* and *SUBNATIONAL SURVEYS* sheets accordingly. For every survey, try to code a response for each of these new variables. For any variable where a response cannot be provided by you (except for variable 1.8b., as explained in point a) and/or input from the national representatives is needed, highlight the respective cell in yellow and add a comment to the Word document so the national representatives can provide assistance.
 - a. 1.8b. If subnational, type of subnational area [Note: This variable only applies to the subnational sheet. Conversion from a LAU/NUTS code to a subnational area for most of the 35 COST Action countries can be found here, [EU 28 - Subnational Area Types \(LAU and NUTS\).xlsx](#), OR here, [Methodological manual on territorial typologies](#). For the former document, please locate the sheet for the country you are reviewing (the sheets are labelled by country name in ISO code) and consult the DEGRUBA column (1 = cities, 2 = suburbs/towns, 3 = rural); in the case this column is blank or has been filled out as “n.a”, notify Ami Saji via email, mark the affected cells in red, and on the feedback document explain that the Sciences Po team is looking into the correct conversion as this information has not been made readily available by EUROSTAT. For the latter document, please consult the table, *NUTS: classification of territorial units for statistics* (pages 116-126); if the conversion is not provided in this table, notify Ami Saji via email, mark the affected cells in red, and on the feedback document explain that the Sciences Po team is looking into the correct conversion as this information has not been made readily available by EUROSTAT.]
 - b. 1.12a. Survey in development/ not yet completed [Note: Select the appropriate response from the list options.]
 - c. 2.12a. Pools emigrants from more than 1 country [Note: Select the appropriate response from the list options.]
 - d. 3.1a. EMM target population with terms standardized [Note: This variable is being post-coded by breaking down the EMM target population description into larger/general key terms (Example: If the EMM target population was defined as “First generation immigrants of Pakistani or Moroccan origin”, it would be post-coded as: Pakistani; Moroccan; First generation migrants). Use the following

¹² For security purposes, the link to this file has been removed in this published version of this guidelines document.

¹³ For security purposes, the link to this file has been removed in this published version of this guidelines document.

Google Drive sheet, [EMM Target population standardized](#),¹⁴ to ensure that this variable is post-coded correctly; more detailed instructions have also been provided on the “Instructions” tab and it is imperative that you read the contents on this tab BEFORE post coding this variable.]

- e. 3.6a. Survey designed as a general population survey [Note: Select the appropriate response from the list options.]
 - f. 7.5a. Migrant languages in ISO code [Note: ISO 639-3 should be used, available [here](#). ISO codes should be listed in all caps and separated with a semicolon]
 - g. 7.7a. Questionnaire in migrant language in ISO code [Note: ISO 639-3 should be used, available [here](#). ISO codes should be listed in all caps and separated with a semicolon]
 - h. 8.9a. Dataset languages in ISO code [Note: ISO 639-3 should be used, available [here](#). ISO codes should be listed in all caps and separated with a semicolon]
 - i. 8.15a. Language(s) of technical survey documentation in ISO code [Note: ISO 639-3 should be used, available [here](#). ISO codes should be listed in all caps and separated with a semicolon]
 - j. 8.20a. Language(s) of survey questionnaire in ISO code [Note: ISO 639-3 should be used, available [here](#). ISO codes should be listed in all caps and separated with a semicolon]
6. Review the *NATIONAL SURVEY* and *SUBNATIONAL SURVEY* sheets of the template:
- a. For each survey (i.e. you need to **review by column by column**), check that the national representatives: (i) provided, at a minimum, responses for the compulsory variables (i.e. those in red text), (ii) correctly coded responses following the instructions provided in the *Surveys Metadata Template Guidelines Document*, and (iii) provided responses that are logical and consistent with one another (e.g. if they provide a response rate calculation, they should have also provided a response rate; if they indicated the survey was stratified by EMM categories, then they should have provided the sub-sample information). For any variable with a missing (again, only those that are in red text) or unclear response, provide a comment (i.e. why the variable needs to be filled out and/or how the national representatives can correct their mistake) in the *National Survey Sheet/Subnational Survey Sheet* section of the Word document and highlight the corresponding cell in yellow. For any variable that has a response with a typing/data entry,¹⁵ or a small/obvious spelling, grammatical, and/or punctuation error,¹⁶ correct the mistake directly onto the template and highlight it in pink. If the response provided is too difficult to correct (i.e. requires extensive corrections or editing) and you are not a native speaker of English, please do not touch this cell and highlight it instead in red; such cells will be handled by Ami Saji or another individual who is able to perform a linguistic check.

¹⁴ For security purposes, the link to this file has been removed in this published version of this guidelines document.

¹⁵ This refers to any typing or data entry error that can be easily corrected by the reviewer. Examples include changing a N/A response to -999 (e.g. variable 1.2) or removing the % symbol from a response rate (e.g. variable 5.3).

¹⁶ For spelling, grammar and punctuation, UK or US writing conventions can be used.

- b. For each survey, ensure that every single variable has a response coded, EXCEPT the variables for assigning a unique identifier to the survey (1.1) and larger study (2.1),¹⁷ which you need to assign using the coding instructions provided in the points below; this means that variables that were non-compulsory and/or were not required to be filled out by the national representatives per the *Surveys Metadata Template Guidelines Document* need to have a response provided, too. Use the table found in [Appendix F](#) (view the Excel version here: [Appendix F - Final Metadata Template: Coding Missing Responses](#)¹⁸) to code variables with any empty responses and mark the affected cells in green (if they do not require input from the national representatives) or yellow (if they require input from the national representatives, in which case a comment should also be added to the Word document).
- c. For each survey, ensure that 1.1 is assigned a unique identifier by using the [2 letter ISO code for the country](#) followed by N (for national surveys) or S (for subnational surveys) and then by consecutive numbers (001, 002, 003...) starting with column B and then continuing until the last survey in this sheet.
- d. For the sheet, check all the open response variables to ensure that the national representatives have used proper nouns that are consistent with the standardization rules outlined in [Appendix G](#). When the template has a mistake of this nature, correct the cell-in-question, and then highlight the cell in blue.
- e. For the sheet as a whole, make sure that all the variables with a -9 / -99 / -999 response have been coded as -9.Don't know / -99.Information not available / -999.Not applicable. When the template has a mistake of this nature, correct the cell-in-question, and then highlight the cell in blue.
- f. For the sheet as a whole, please pay special attention to the format of dates (they should be dd/mm/yyyy) and that all figures relating to sample sizes do not include any comma or dot to separate the thousands (e.g. change 1,235 to 1235, and 1.235 to 1235) and that they do not provide any unnecessary text. Please correct all cells that do not respect the aforementioned date and number formats; any cell that has required corrections should also be highlighted in pink.
- g. For the sheet as a whole, double check the open response variables (particularly the comments section for each section), to ensure that the information provided can be displayed on our publicly available EMM Survey Registry (e.g. if in 10.2 the data quality is described as being “very bad because of the poor survey design”, the comment should be adapted to be more diplomatic) . Please edit and modify the language whenever necessary and mark affected cells in yellow; please also explain to the national representatives why the contents of the cell were modified and to have them review and validate the edits made. If you are unsure or unable

¹⁷ This file explains how to assign and subsequently keep track of the unique identifiers for surveys belonging to a larger study. Detailed information about this file and how it should be used is also discussed in section 5, as the process of checking that we have assigned unique identifiers to such surveys is part of the data cleaning stage.

¹⁸ For security purposes, the link to this file has been removed in this published version of this guidelines document.

to edit and modify the language of a cell and such changes are needed, please highlight the affected cell in red so Ami Saji can make the needed changes.

- h.** For the sheet as a whole, please make sure that variables 8.4, 8.13 ,and 8.19 have coded the DOI in http format (e.g. <https://doi.org/10.1109/5.771073>). Please correct all DOIs that are NOT written in http format and highlight such cells in pink.
- i.** For the sheet as a whole, double check the responses in section 2 to make sure they have been correctly filled out. For example, if a survey is identified as being a single cross-section, then the sub-section in section 2 for longitudinal/repeated surveys (i.e. section 2b) should not be filled out. Another example is that if a survey is part of an international survey program, then not only should section 2c be filled out, but 2.5 should be filled out, too. Please remember that if any cell requires input/action from the national representatives, it should be highlighted in yellow, and a corresponding comment should be added to the Word-based feedback document.
- j.** For the sheet as a whole, ensure that surveys belonging to an international survey program have coded 2.6 and 2.9 to reflect the starting date of the country. In other words, 2.6 and 2.9 should not list the starting date for the international survey program as a whole UNLESS the country had the same starting date. If any cell requires input/action from the national representatives (i.e. it's not clear if the date listed is for the country or not), it should be highlighted in yellow, and a corresponding comment should be added to the Word-based feedback document.
- k.** For the sheet as a whole, double check if there are any longitudinal surveys with multiple cohorts. If yes, please make sure that section 2 is coded so that all waves, regardless of the cohort being surveyed, is treated as "belonging" to the same "larger study". For example, if survey X had 2 cohorts, one that started in 1998 (respondents surveyed in 2000 and 2002) and another that started in 2004 (respondents surveyed in 2006 and 2007), then: (i) 2.6 would be coded so that all waves of the 1998 and 2004 cohorts have the same start date (i.e. the start date of the 1998 cohort, 00/00/2000); (ii) 2.7 would be coded to reflect how frequently the survey has occurred taking into account both the 1998 and 2004 cohorts (i.e. irregular, every 1-2 years); (iii) 2.8 would be coded so that the first iteration of the 1998 cohort is wave 1, the second iteration of the 1998 cohort is wave 2, the first iteration of the 2004 cohort is wave 3, etc.; and (iv) 2.15 would be coded to include a comment about the wave order for each cohort (e.g. for wave 3, you would code, "wave 3 of survey X is the first time the 2004 cohort was surveyed), given that we're counting the waves collectively for the 2 cohorts in 2.8. Please keep in mind that if any cell requires input/action from the national representatives, it should be highlighted in yellow, and a corresponding comment should be added to the Word-based feedback document.
- l.** For all general population surveys (though ideally for all surveys), please double check that variable 3.4 has at least 1 "Yes" response; this is because at least 1 EMM-related question is needed in order for a survey to count as an EMM survey.

If 3.4 has not been coded with at least 1 “Yes” response, please make sure you highlight this section in yellow and then provide a comment in the Word-based feedback document.

- m.** For the sheet as a whole, double check that the general population surveys are coded as follows. If the survey **has been partitioned** to have an EMM subgroup (and/or any other subgroups), make sure that: (i) section 5 captures the sample size information for ALL the survey respondents and not just the EMM respondents; (ii) variable 5.9 includes a comment explaining that section 5 provides the sample size information for the survey as a whole and that section 6 is where the EMM-specific sample size information (as well as all other subgroups) can be found; and (iii) section 6 captures the sample size information for the EMM respondents (as well as any other subgroups that have resulted from partitioning). If, however, the survey **has NOT been partitioned** to have any subgroups, including EMM respondents, make sure that: (i) section 5 captures the sample size information for ALL the survey respondents; (ii) variable 5.9 includes a comment explaining that section 5 provides the sample size information for the survey as a whole; and (iii) variable 5.9, instead of section 6, captures the (estimated) EMM sample size information. Please remember that if any cell requires input/action from the national representatives, it should be highlighted in yellow, and a corresponding comment should be added to the Word-based feedback document. Also, if you have any questions about how to handle general population surveys or need further assistance, please contact Ami Saji.
- n.** For the sheet as a whole, double check surveys that have included a partitioned majority subgroup (i.e. a “yes” response to variable 3.6). For all surveys with a partitioned majority subgroup, make sure that: (i) section 5 captures the sample size information for ALL the survey respondents, (ii) variable 5.9 includes a comment explaining that section 5 provides the sample size information for the survey as a whole and that section 6 is where the partitioned subgroup sample size information can be found; and (iii) section 6 captures the sample size information for the majority subgroup (as well as any other subgroups that have resulted from partitioning). Please remember that if any cell requires input/action from the national representatives, it should be highlighted in yellow, and a corresponding comment should be added to the Word-based feedback document.
- o.** For the sheet as a whole, double check all surveys that DO NOT fall under points m and n above, AND have had section 6 filled out by the national team (because the survey partitioned the EMM respondents). Make sure that for all such surveys section 5 captures the sample size information for ALL the survey respondents and section 6 captures the sample size information for ALL of the partitioned subgroups (if there are more than 5 subgroups, then all remaining subgroups should be documented in the comments section of section 6). Please remember that if any cell requires input/action from the national representatives, it should be

highlighted in yellow, and a corresponding comment should be added to the Word-based feedback document.

- p. For the sheet as a whole, check to see if it includes any EU-LFS waves or Roma-targeted surveys by UNDP and/or FRA. If yes, highlight in red the survey/wave name in English (row 8) for all applicable surveys/waves on the sheet and also highlight in red all the comments you have made on the Word-based feedback document about such surveys/waves. This will allow Ami Saji to do an extra careful check of these surveys/waves and update the Word-based feedback document if needed. [NOTE: This step was added in because there was repeated confusion about how to properly document these surveys/waves across the different countries.]
- q. For the sheet as a whole, re-check variable 4.4 to ensure that a proper sampling frame has been listed. During this re-check, you should also re-review information provided in variables 4.2-4.3 to see if the sampling frame coded in variable 4.4 seems appropriate. For example, if in variables 4.2-4.3 the sampling method is described as using random route, then a census/population register type sampling frame would not be suitable; instead, the sampling frame listed should be a list of addresses or location-based starting points. Please remember that if any cell requires input/action from the national representatives, it should be highlighted in yellow, and a corresponding comment should be added to the Word-based feedback document.

An example of how the template should be highlighted in yellow, pink, red, green, and blue can be found here: [Appendix D - Norway Template \(Final\).xlsx](#).¹⁹

- 7. After reviewing the *NATIONAL SURVEY* and *SUBNATIONAL SURVEY* sheets, consult the Excel file, [List of International Surveys](#).²⁰ For any survey that appears on this list, please check if it has: (i) been included on the template (i.e. the survey is or is not “missing” or “omitted”) and (ii) has been correctly coded in section 2 (Information about the inclusion of the survey in the larger study).
 - a. For point (i), list all the “missing” or “omitted” surveys on the template by coding the survey name (variable 1.3) onto the first “free” column (i.e. the first column after the column containing the last survey coded) and highlighting the corresponding cell in yellow. Then add a comment to the Word document asking that the national representatives check if the survey meets our inclusion criteria (make sure to provide a link to the survey) and, in the case it does, to code the survey information. In the case the survey does not meet our inclusion criteria, explain that the national representatives should delete the column containing the survey name you have coded (e.g. if you added in the EU-SILC survey to column AA and the national representatives determine that the EMM sample for this survey does not meet our sample size threshold, the national representatives

¹⁹ For security purposes, the link to this file has been removed in this published version of this guidelines document.

²⁰ For security purposes, the link to this file has been removed in this published version of this guidelines document.

should delete column AA) and provide an explanation as to why the survey did not meet the inclusion criteria.

- b. For point (ii), check the Excel file, [Unique ID Tracker: Larger Studies](#),²¹ to see if any other country has already coded this survey onto the template. If yes, use the information coded onto the other country's template (only those uploaded to the EMM Survey Registry and/or at the data cleaning stage) to check and correct (if necessary) the entries for section 2.

Also, if you identify any international surveys on the template that have not been included on the [List of International Surveys](#),²² please add it in. Please then send an email to Ami Saji (amikatherine.saji@sciencespo.fr) and the Sciences Po team email (sshoc.project@sciencespo.fr) explaining which survey has been added to the list and which countries are affected.

8. Once the *ANNOTATIONS*, *NATIONAL SURVEY*, and *SUBNATIONAL SURVEY* sheets of the template have been fully reviewed and edited/modified, provide in the *General Comments* section of the Word document, a 1-2 sentence description of the overall quality of the template, including an explanation of what the general mistakes/errors were and why additional follow up and/or clarification is needed from the national representatives. Also provide a brief description of how the feedback (Word document plus Excel file with the highlighted cells) has been structured.
9. Proofread the Word document and Excel file with the highlighted cells. Double check that the comments provided in the Word document are clear and that they correspond to a cell (highlighted in yellow only) in the Excel file.
10. Email the finalized feedback (Word document and the Excel file with the highlighted cells) to Ami Saji (amikatherine.saji@sciencespo.fr) and the Sciences Po team email (sshoc.project@sciencespo.fr). Ami Saji will then send an email that: (i) explains what the Word document and Excel file are; (ii) asks the national representatives to review each comment in the Word document, make the necessary edits/modifications on the template, and remove the yellow highlighting from each cell where the necessary edits/modifications have been made; (iii) requests that the national representatives provide an estimated turnaround time for responding to the reviewer's comments and making the necessary edits/modifications; (iv) advises the national representatives that the reviewer and the Sciences Po team are available to answer any questions about the comments provided; (v) asks the national representatives to send the revised final template to the reviewer, Ami Saji, and the Sciences Po team email; (vi) asks the national representatives and anyone else who helped filled in the metadata template (first variable of the Annotations sheet) if they provide us with consent to list their names on the registry; and (vii) asks the national representatives to obtain metadata usage consent from data archives/repositories that are storing/managing surveys coded onto the template (see [Consent Email Template - Use of Metadata for EMM Survey Registry](#)²³ for an example of

²¹ For security purposes, the link to this file has been removed in this published version of this guidelines document.

²² For security purposes, the link to this file has been removed in this published version of this guidelines document.

²³ For security purposes, the link to this file has been removed in this published version of this guidelines document.

how the national representatives can request the needed consent).²⁴ Once this email (see Appendix E for a sample) has been sent, Ami Saji will upload the finalized feedback onto the Team Google Drive folder, *COST WG1_WG2 docs*, under the sub-folder, *8. Final Metadata Quality Control*.

11. Once the national representatives have submitted the fully revised template, **review the edits/modifications made** by the national representatives and **double-check that there are no additional typing/data entry, spelling, grammatical, and/or punctuation errors** (see [example](#) of a fully quality controlled template). Please also make sure that the national representatives have consulted and added in ALL the “missing” or “omitted” surveys (i.e. those that you had identified via the [List of International Surveys](#)²⁵) that meet our inclusion criteria.²⁶
 - a. If further edits/modifications are still needed by the national representatives, produce new feedback documents that use the same format/structure as the original feedback documents; Ami Saji will then send these new feedback documents to the national representatives. This process of producing new feedback documents will continue until no further edits/modifications are needed by the national representatives/and or the reviewer. Once this is achieved, proceed to point b below.
 - b. When no further edits/modifications are needed by the national representatives and/or the reviewer: (i) remove any remaining color highlights from the template, (ii) rename the template as, **Country Name (Re-submission Date)_FINAL ETHMIGSURVEYDATA Metadata Template_Corrected (Reviewer's Initials).xlsx**, with the “Re-submission date” formatted as *dd.mm.yy*, and (iii) email the re-named template to Ami Saji and the Sciences Po team email. Ami Saji or the reviewer will then upload the fully quality controlled template onto: (i) the Team Google Drive folder, *COST WG1_WG2 docs*, under the sub-folder, *8. Final Metadata Quality Control* and (ii) Google Drive folder *SSHOC (WP9 - Task 9.2)*, sub-folder *3. SSHOC (WP9 - Task 9.2) Metadata Compilation*, and sub-sub-folder *B. Final Metadata Quality Control (COPY)* as a security backup (NOTE: If you have not been given access to the Team Google Drive folder or the SSHOC Google Drive folder, warn Ami Saji or Laura Morales that you have now completed stage 4 so that they can upload the finalized file).

5. Data Cleaning

After a template has completed all of the final quality control steps outlined and described in the section above, it will undergo a data cleaning process, so it is fully readable and ready to be

²⁴ All metadata consent emails from the data archive/repository need to be stored by country: i.e. the country-specific folder found in *8. Final Metadata Quality Control*, of the Team Google Drive folder, *COST WG1_WG2 docs*.

²⁵ For security purposes, the link to this file has been removed in this published version of this guidelines document.

²⁶ As a general rule of thumb, the general population surveys that are listed on the [List of International Surveys](#) may not meet our sample size threshold for most countries. However, the EMM specific surveys, such as EU-MIDIS, should, for the most part, not be omitted from the template. If you have any doubts, please email Ami Saji at amikatherine.saji@sciencespo.fr.

uploaded onto the EMM Survey Registry and be used for analysis. The files generated at this stage of the process need to be saved in the Team Google Drive folder, *COST WG1_WG2 docs*, under the sub-folder, *9. Final Metadata Data Cleaning*. Once completing all of the data cleaning steps for a template (i.e. the steps specified in the table below), you will produce 1 Excel file that will be used for data analysis, as well as 3 Excel files and 2 CSV files for the EMM Survey Registry.

As referenced above, the table below provides step-by-step instructions on how to perform the data cleaning process for a template. Please make sure that each step is implemented fully and correctly.

Successful execution of the data cleaning steps requires an intermediate proficiency of Microsoft Excel. If you have had limited experience with Microsoft Excel or if you have not used Microsoft Excel in quite some time, it may be helpful to consult the training videos produced by Microsoft ([click here](#)) or to search for video tutorials on Youtube.

Specific step		Comments
ADDING NEW VARIABLES		
1	In the Subnational surveys sheet: Create a new variable 1.7a right after 1.7 (leaving one row above/below empty as for the rest of the file) and code in that row a single region name in English that will be used for sorting on the online database. Call it "1.7a. Region (for sorting)" (See Norway cleaned file if needed)	Which region to choose? Choose in the first instance the region that seems more important in the study in terms of sample size (in variable 5.2 or 5.9). If all regions had the same sample size or no information about sub-sample sizes is available, then choose the one with a larger target population (in variable 3.5), and if this information is not available the first one in alphabetical order in English (variable 1.7).
2	For some rows with sub-headings that are currently not numbered, add the numbering system indicated in the Comments cell to the right, for both national and sub-national sheets.	<ul style="list-style-type: none"> • For row before variable 2.6 (In case of repeated /longitudinal surveys), we will rename them in all Excel files as 2b.; • For row before variable 2.9 (In the case of surveys that are a part of an international survey program), we will rename them as 2c.; • For row before variable 2.11 (If sample is pooled) we will rename them as 2d.; • For row before variable 7.3 (For personal interviews:), we will rename them as 7.2b. <p>Please make sure that you use the numbering 2b./2c./2d. and NOT any other variation like 2.b/2.c/2.d. If you still have doubts, please consult column D of the sample Norway file (Youngminds version - step 27) : Norway(15.11.19)Youngminds.xlsx.²⁷</p>
3	For the Annotations sheet, only keep rows that will be shown on the front-end	<p>The following rows need to be deleted for the Youngminds version:</p> <ul style="list-style-type: none"> • Email(s) of the person(s) filling in this file • Contact person(s) for this country for WG1 (national surveys) • Email(s) of contact person(s) for this country for WG1

²⁷ For security purposes, the link to this file has been removed in this published version of this guidelines document.

		<ul style="list-style-type: none"> • Contact person(s) for this country for WG2 (subnational surveys) • Email(s) of contact person(s) for this country for WG2 • Prior experience of the person(s) filling in this file with designing, conducting, or analyzing surveys. Please provide a short description • Challenges or difficulties faced in completing the template, including specific reasons. Please comment here
4	Place the Annotations information in each of the columns of the National and Subnational Surveys sheets	After the last variable in section 10 (i.e. 10.4) in both the national and subnational sheets, you will need to add in information about who compiled the metadata on the Excel file by creating a new section: 11. Information on this compilation of metadata. Please see how this is done for the Norwegian final metadata file (dated 15.11.19²⁸) in folder "9. Final Metadata Data Cleaning - subfolder 9.1 Common Steps" to replicate exactly the same patterns of variables. To fill out this new section, you will need to extract information from the Annotations sheet; please keep in mind that this means that not ALL the information displayed on the Annotations sheet is covered in this new section. Please also pay special attention to the fact that in the National sheet, we only include the information relevant for the National surveys and in the Subnational sheet only the info relevant for the Subnational surveys. This includes deleting 11.6 for the National sheet and copying the value -999.Not applicable for 11.5 in the Subnational sheet. Finally, please make sure to code any response of Not applicable in text as -999.Not applicable.
ADAPTING EXISTING VARIABLES		
5	Re-code dates using the established protocol	<p>Please make sure that all of the variables for dates are shown with the full day, month, and year format (DD/MM/YYYY). If needed change the cell format to date = DD/MM/YYYY if needed for variables:</p> <ul style="list-style-type: none"> • 1.11 • 1.12 • 2.6 • 2.9 • 11.2 • 11.3 <p>If a day and/or month is not known code as 00.</p> <p>All dates in the annotations sheet should also respect the aforementioned format.</p>
6	Check the unique identifiers for each survey (1.1) and larger study (2.1)	<p>Unique identifier format for surveys (1.1): 2 letter ISO country code, followed by N for national level or S for subnational level, and then a 3-digit number (start with column B, continuing until the last survey; numbers should be consecutive starting with 001). <i>Example for a national level survey conducted in Croatia and coded in column B: HRN001</i> <i>Example for a national level survey conducted in Croatia and</i></p>

²⁸ For security purposes, the link to this file has been removed in this published version of this guidelines document.

		<p>coded in column C: HRN002</p> <p>Unique identifier format for larger study (2.1): 3-letter abbreviation based on the English name of the larger study, followed by the start year of the larger study <i>Example for the EU-MIDIS survey: MID2008</i></p> <p>Surveys with multiple/recurring waves should have the same unique identifier. <i>Example for ESS surveys conducted in 2000 (assuming this is the first wave), 2001 and 2003: All ESS surveys should be labelled as ESS2000</i></p> <p>Before adding in a unique identifier for the larger study, the file, Unique ID Tracker: Larger Studies,²⁹ will need to be consulted and used, as it details how to assign and keep track of the unique identifiers.</p>
7	Check that the EMM target population description variable (3.1) has been properly post-coded into variable 3.1a with terms separated by semicolon (;)	<p>This variable is currently being post-coded by breaking down the EMM target population description into larger/general key terms. [Use the following Google Drive sheet to check that the same wording has been applied to the same groups and that all the groups mentioned are listed: EMM Target population standardized³⁰]. <i>Example: If the EMM target population was defined as “First generation immigrants of Pakistani or Moroccan origin”, it would be post-coded as: Pakistani; Moroccan; First generation migrants</i></p> <p>If you have any questions about how the EMM target population should be broken down, please contact Ami Saji.</p>
8	Check and correct sample sizes and other numerical variables	<p>3.5 Size of the EMM target pop: Remove any commas for thousands (e.g. write 1100 and NOT 1,100) and remove any text (except if it is -9.Don't know or -99.Information not available) that has been provided (copy any additional text that was included in 3.5 and add it to 3.7). If more than one figure is listed here, please select the one that is most relevant or better captures the identified EMM target population.</p> <p>5.1 Total gross/issued sample and 5.2. Total net/achieved sample: We only want numbers, no text (except for -9.Don't know and -99.Information not available); please do not use commas for thousands (e.g. write 1100 and NOT 1,100). Ensure that only one set of sample size numbers is included here, the lowest level one (typically individuals). If you get both households and individuals, keep only individuals, and copy the rest of the information into variable 5.9 where you can just paste any additional text that was included in 5.1/ 5.2.</p> <p>Section 6: Follow the instructions provided for 5.1/5.2 for all the numerical variables.</p>
9	Standardize the variables with a series of values by listing and separating the values via semicolons (;)	<p>The variables that require this standardization are the following:</p> <ul style="list-style-type: none"> • 1.6. Which subnational level

²⁹ For security purposes, the link to this file has been removed in this published version of this guidelines document.

³⁰ For security purposes, the link to this file has been removed in this published version of this guidelines document.

		<ul style="list-style-type: none"> • 1.7. Name of region(s) Eng., 1.8 Name of region(s) Nat.: [Also sort all names in alphabetical order] Make sure that any given region is always spelled in the same way and with the same contents. For example, "region X (cities A, B, C)" and "region X (cities B, C, D)" is not acceptable because "region X" appears twice with different contents. In such a case it is better to only list the region as a whole and provide (copy/paste) the detailed information about sampled cities in variable 1.17. You will need, however, to pay attention to what the level of analysis was in variable 1.6 (whether regional or local), if regional prioritize the names of the regions, if local prioritize the names of the localities. See an example of how this is done in the cleaned file for Croatia in folder 9. • 1.13.30a. If "other" in 1.13, specify • 1.14a. If "other" in 1.14, specify • 2.5. Name of other countries/regions/cities Eng. • 3.2a. If "other" in 3.2., describe which • 3.3a. If "other means" in 3.3., describe which • 4.5. Sampling units • 7.5a. Migrant languages in ISO code • 7.7a. Questionnaire in migrant language in ISO code • 8.9a. Dataset language(s) in ISO code • 8.15a. Language(s) of technical survey documentation in ISO code • 8.20a. Language(s) of survey questionnaire in ISO code • 11.1. Person(s) filling in this file <p>Whenever possible and whenever it makes sense, please try to semicolon separate entries for 4.4. Sampling frame(s). This use of the semicolon is particularly important for data analysis purposes</p>
10	Double check that all variables have a response provided (i.e. there are no missing values). Correct as needed	The table showing/explaining how to handle missing values can be found in Appendix F and here: Appendix F - Final Metadata Template: Coding Missing Responses . ³¹
11	Double check that variables where the requested information is unknown, not available, or not applicable are coded using the predetermined and specified format. Correct as needed	The response options are: -9.Don't know, -99.Information not available, or -999.Not applicable.
12	Double check that proper nouns are being consistently used within a template and across templates. Correct as needed	For the different types of proper nouns, specific data entry formats have been specified (see Appendix F).
13	Post-code 1.6 and the "other" variables (1.10a, 1.13.30a, 1.14a, 3.2a, 3.3a, 5.5)	1.10a, 1.14a, 3.2a, 5.5 The response provided should be checked to see if it can be grouped into one of the "non-other" options (i.e. the set of variables above the "other" variable). If yes, delete the response provided, change the response to "No", and mark the relevant "non-other" option as "Yes". Only if the "other" variable needs to be used, should the response be post-coded by using 1-3 keywords; if you have any questions about how to identify the appropriate keywords to use, please contact Ami Saji. [NOTE: These variables have not been identified as sorting and/or filtering variables for the EMM Survey Registry. However, they are being post-coded in this way in the case this

³¹ For security purposes, the link to this file has been removed in this published version of this guidelines document.

		<p>changes.]</p> <p>1.6 This a filtering variable for the EMM Survey Registry and needs to be post-coded as follows: 1. Consult the Google Drive sheet, Post-coding: 1.6. Which subnational level³² and check if each of the subnational levels identified in the cell are already listed. If yes, please make sure that all of the subnational levels identified in the cell match exactly the listed options. If any of the subnational levels identified in the cell DO NOT appear as one of the listed options (i.e. it is a “missing” subnational level), proceed to step 2-3 below. 2. Ensure that the “missing” subnational level is written in the plural form (e.g. Cities; Municipalities) whenever possible. 3. Add the “missing” subnational level to the Google Drive sheet by respecting the alphabetical order.</p> <p>1.13.30a, 3.3a These variables are sorting/filtering variables for the EMM Survey Registry. As such, these variables need to be post-coded as follows: 1. Check if the response provided can be grouped into one of the “non-other” options (i.e. the set of variables above the “other” variable). If yes, move the response provided to the comments section for reference, code the now empty cell as “-999.Not applicable”, change the response for 1.13.30 as “No”, and mark the relevant “non-other” option(s) as “Yes”. Only if the “other” variable needs to be used, should you proceed to step 2 below. 2. Open the Google Drive sheet, Post-coding: “Other” variables (1.13.30a, 3.3a)³³ and consult the tab, 1.13.30a, for variable 1.13.30a and the tab, 3.3a, for variable 3.3a. -Check if the response provided can be grouped into one of the response options listed on the tab. If yes, edit the response provided to exactly match the wording for the relevant option found on the tab. If no, postcode the response using 1-3 keywords; if you have any questions about how to identify the appropriate keywords to use, please contact Ami Saji.</p>
14	Double check that surveys belonging to the same larger survey are coded consistently (section 2)	Please only reference waves that have already been uploaded to the EMM Survey Registry and/or have completed the full final quality control process.
15	Correct typos for any of the original variable names (1.13.29, 1.5, 3.3, 7.8) and check that all variable names, including the headings/sub-headings, have not been altered by the national teams	<p>(in the Subnational Surveys sheet) change:</p> <ul style="list-style-type: none"> ● 1.13.39 for 1.13.29 ● 1.5: The word territorial is misspelled ● 3.3: The word operationalization is misspelled ● 7.8: The unit (minutes) is missing; re-label as “Average duration/length of interview (minutes)” <p>To check the variable names and headings/subheadings, please consult this Google Drive sheet:</p>
16	Correct the “-9.Don’t Know” responses to be “-9.Don’t know”	

³² For security purposes, the link to this file has been removed in this published version of this guidelines document.

³³ For security purposes, the link to this file has been removed in this published version of this guidelines document.

17	Remove tabs/hard enters inside of cells	Use the CLEAN function.
18	Remove extra spaces before and after text (i.e. leading and trailing spaces)	You can use the TRIM function.
19	Revise all cells with URLs and make sure there is one space before and one after each URL link: Revise all the "Additional comments rows (1.17, 2.15, 3.7, 4.6, 5.9, 6.8, 7.10, 8.21, 9.8,) as well as all the relevant availability cells (8.2, 8.11, 8.17, 9.4, 10.3)	For example (but without the quotes!): " https://www.google.com " NOTE: The URLs need to be formatted in this way, so they appear as "Go to link" on the EMM Survey Registry.
20	Remove the empty rows	For the national sheet this should result in 241 rows. For the subnational sheet this should result in 248 rows.
21	Remove the window panes/sections	Check the (un)freeze panes and "split" icons of the "View" tab of the Excel ribbon.
22	Complete one final spell check of the file **DO NOT SKIP THIS STEP**	While grammar and linguistic errors should be checked as part of the final quality control process, this final spell check of the file will help in catching any errors that may have been accidentally overlooked. Please make sure that you use the spell check feature of Excel when doing this final grammatical and linguistic check. Also, please be careful when using spell check that you correct/react to all the errors that spell check finds.
23	**ATTENTION: STEP REQUIRES NEW FILE VERSION TO BE SAVED AND STORED** At this stage, save the file as CountryName(dd.mm.yy).xlsx [for example Norway(15.11.19).xlsx]	Save and keep the file on your computer for now. This file will be sent to Ami Saji (see step 24 below) who will add it to "folder 9. Final Metadata Data Cleaning - subfolder 9.1 Common Steps" once the file has completed step 24. To view an example of a file produced at this step, see Norway(15.11.19).xlsx . ³⁴ NOTE: For the purposes of data analysis, the files saved in this folder will be used.
24	Conduct a final quality check of the metadata using a STATA do file	Send the file produced in step 23 above to Ami Saji who will run the STATA do file as a final quality check of the metadata. She will make ALL necessary corrections to the file, add the checked and subsequently validated file to "folder 9. Final Metadata Data Cleaning - subfolder 9.1 Common Steps," AND complete ALL the remaining steps.
25	Create and use a copy of the file produced in step 23 and checked/validated in step 24	Using the version uploaded to "folder 9. Final Metadata Data Cleaning - subfolder 9.1 Common Steps," create a copy and use it to execute steps 26-29.
26	Remove all quotes (") and all apostrophes (') in the National Surveys and Subnational Surveys sheets	To remove all the quotes and apostrophes, use the "find all" and "replace all" commands in Excel.
27	Insert back in the apostrophes (') for the -9 responses	After completing the step above, the response option, "-9.Don't know" would have been changed to "-9.Dont know". Please revert the "-9.Dont know" responses back to "-9.Don't know" by using the "find all" and "replace all" commands in Excel.
28	For variable 5.2, make sure only numbers are used (NO text)	If a cell has been coded as -9.Don't know or -99.Information not available, please change to -9 and -99 respectively.

³⁴ For security purposes, the link to this file has been removed in this published version of this guidelines document.

29	<p>**ATTENTION: STEP REQUIRES NEW FILE VERSION TO BE SAVED AND STORED** Add the Survey Fields from Youngminds into the main "CountryName(dd.mm.yy)Youngminds.xlsx" file</p>	<ol style="list-style-type: none"> 1. Download and open the file SurveyFields(15.11.19).³⁵ 2. In the National Surveys sheet of the country template you are working with: Select all the columns that contain the variable names and the survey metadata and cut them and paste them into column D (i.e. All of the variables and survey metadata should be moved so that the variables are listed starting in column D and the metadata for the first survey is listed in column E. Columns A-C will therefore be made blank/empty). 3. Copy columns A to C in the National Survey Fields sheet of the SurveyFields file into columns A to C of the National Surveys sheet. 4. Check row by row that the numbering of the variables in columns C and D coincides (disregard differences in punctuation, etc.). Check that it coincides with the layout of the example in the file: Switzerland(28.1.20)Youngminds.xlsx.³⁶ 5. Repeat steps 2 to 4 for the Subnational Surveys sheet in your main file but using the Subnational Survey Fields sheet of the SurveyFields file. Check that it coincides with the layout of the example in the file: Switzerland(28.1.20)Youngminds.xlsx.³⁷ 6. Save the file and upload it to its own country folder in "9. Final Metadata Data Cleaning - subfolder 9.2 EMM Survey Registry (For Youngminds)". Please make sure that you save and store this file in the aforementioned folder BEFORE proceeding to the remaining steps below.
30	<p>**ATTENTION: STEP REQUIRES NEW FILE VERSIONS TO BE SAVED AND STORED** Generate the .csv versions for uploading into the online database</p>	<ol style="list-style-type: none"> 1. Select the National Surveys sheet of your "CountryName(dd.mm.yy)Youngminds.xlsx" file and right-hand click on "Move or Copy". In the emerging window, tick on "Create a copy" and in the drop-down menu of where to save it select "New File". You should get a new emerging file just with the National Surveys sheet. Save it as CountryName(dd.mm.yy)NatCSVYoungminds.xlsx [e.g. Norway(15.11.19)NatCSVYoungminds.xlsx] 2. Select column D and remove it. Save the file (still in excel format) and upload it to its own country folder in "9. Final Metadata Data Cleaning - subfolder 9.2 EMM Survey Registry (For Youngminds)" 3. Now save the file as .csv (attention: using comma as the separator NOT semicolon! You need to make sure that this is properly set up in your computer, as if you use commas as the decimal separator, it may not work out, for example if you are using Excel in a French language computer!) with the same name. Please make sure that you are only saving the file with the columns that contain information, as sometimes all the empty columns may be saved, so delete any empty columns that show up (check with TextEdit or any similar text file opener). Upload it to its own country folder in "9. Final Metadata Data Cleaning - subfolder 9.2 EMM Survey Registry

³⁵ For security purposes, the link to this file has been removed in this published version of this guidelines document.

³⁶ For security purposes, the link to this file has been removed in this published version of this guidelines document.

³⁷ For security purposes, the link to this file has been removed in this published version of this guidelines document.

		<p>(For Youngminds)"</p> <p>4. Repeat steps 1 to 3 but for the Subnational Surveys sheet and calling the new files "CountryName(dd.mm.yy)SubnatCSVYoungminds.xlsx" or .csv depending on the step.</p> <p>5. Email Ami to let her know you've completed steps 1-28 of the data cleaning process, so that she can review all the files that have been produced. Once she validates all the files, they will be added to the EMM Survey Registry and/or used for data analysis.</p>
31	Fill out the tracker to indicate the unique ID number (variable 1.1) assigned to the last survey in the National or the Subnational sheet.	<p>The tracker can be accessed here.³⁸ Please read the instructions tab first before filling out the tracker.</p> <p>The purpose of the tracker is to facilitate the process of assigning new IDs to survey records added directly to the back-end.</p>
END OF DATA CLEANING PROCESS		

³⁸ For security purposes, the link to this file has been removed in this published version of this guidelines document.

Conclusion

This guidelines document outlines how each template should be reviewed and handled to ensure version control. As indicated in this document, the quality control process consists of a first stage quality control (compulsory for all templates), a second stage quality control (only for templates rated as having “serious mistakes” during the first stage of quality control), and a final quality control (compulsory for all templates). Only after a template has undergone the full final quality control process will it proceed to and complete the data cleaning process.

For anyone contributing to and assisting with the version control of the templates, they must read this guidelines document first. Any questions about the version control of the templates can be directed to Laura Morales (laura.morales@sciencespo.fr) and Ami Saji (amikatherine.saji@sciencespo.fr).

Appendix A: Structure and Use of the Team Google Drive folder, *COST WG1_WG2 docs*

The Team Google Drive folder, *COST WG1_WG2 docs*, was created to house all files related to the metadata compilation task undertaken by WG1 and WG2. Below is a table that shows and describes all of the folders in *COST WG1_WG2 docs* that are relevant to the version control process.

Folder Name	Use
0. Pilot	This folder houses the pilot version of all of the metadata compilation materials and any templates that were completed during the pilot stage. The files in this folder should serve only as a reference for the version control process.
1. Metadata Compilation Materials (FINAL)	This folder houses the final version of all of the metadata compilation materials. The files in this folder are what the national representatives have been asked to use when compiling the survey metadata for their respective country.
2. 5 Surveys Metadata	<p>This folder houses all of the metadata templates with 5 survey coded that have been submitted by the national representatives.</p> <p>All the templates should be labelled using the following format: <i>Country Name (Submission Date)_5 Surveys_FINAL ETHMIGSURVEYDATA Metadata Template.xlsx.</i></p>
3. 5 Surveys Metadata Quality Control	<p>This folder houses all of the feedback (Word document and Excel file with the mistakes/errors highlighted) provided to the national representatives for their respective template with 5 surveys.</p> <p>Each country has its own subfolder (labelled as, <i>Country Name</i>) with their respective Word document and Excel file. The Word document should be labelled as, <i>Template Feedback Country Name</i>, and the Excel file as, <i>Country Name (Submission Date)_5 Surveys_FINAL ETHMIGSURVEYDATA Metadata Template_Reviewed (Reviewer's Initials).xlsx.</i></p>
4. 2nd Submission Metadata	<p>This folder houses all of the metadata templates that have been submitted by the national representatives for the second stage quality control.</p> <p>All the templates should be labelled using the following format: <i>Country Name (Submission Date)_2 QC_FINAL ETHMIGSURVEYDATA Metadata Template.xlsx.</i></p>
5. 2nd Submission Metadata Quality Control	<p>This folder houses all of the feedback (Word document and Excel file with the mistakes/errors highlighted) provided to the national representatives, who submitted their template for the second stage quality control.</p> <p>Each country has its own subfolder (labelled as, <i>Country Name</i>) with their respective Word document and Excel file. The Word</p>

	<p>document should be labelled as, <i>2 QC Template Feedback Country Name</i>, and the Excel file as, <i>Country Name (Submission Date)_2 QC_FINAL ETHMIGSURVEYDATA Metadata Template_Reviewed (Reviewer's Initials).xlsx</i>.</p>
<p>6. Midpoint Metadata</p>	<p>This folder houses all of the midpoint metadata templates that have been submitted by the national representatives. As preparation for the COST Action meetings in Palermo (6-8 March 2019) and Rome (8-9 July 2019) respectively, a midpoint version was requested from the national representatives if they were unable to submit the final version of their template by the meeting start date.</p> <p>All the templates should be labelled using the following format: <i>Country Name (Submission Date)_Midpoint_FINAL ETHMIGSURVEYDATA Metadata Template.xlsx</i>.</p>
<p>7. Final Metadata</p>	<p>This folder houses all of the final metadata templates that have been submitted by the national representatives.</p> <p>All the templates should be labelled using the following format: <i>Country Name (Submission Date)_FINAL ETHMIGSURVEYDATA Metadata Template.xlsx</i>.</p>
<p>8. Final Metadata Quality Control</p>	<p>This folder houses all of the feedback (Word document and Excel file with the mistakes/errors highlighted) provided to the national representatives, who submitted the final version of their template. It also houses the final template that has been revised by the national representatives (using the feedback provided) and reviewed/proofread by a Sciences Po team member, as well as the metadata consent.</p> <p>Each country has its own subfolder (labelled as, <i>Country Name</i>) and contains the following files:</p> <ul style="list-style-type: none"> - The Word feedback document (labelled as, <i>Final Template Feedback Country Name</i>) - The Excel feedback file (labelled as, <i>Country Name (Submission Date)_FINAL ETHMIGSURVEYDATA Metadata Template_Reviewed (Reviewer's Initials).xlsx</i>) - The final template revised by the national representatives and reviewed/proofread by a Sciences Po team member (labelled as, <i>Country Name (Re-submission Date)_FINAL ETHMIGSURVEYDATA Metadata Template_Corrected (Reviewer's Initials).xlsx</i>) - The metadata consent emails (labelled to include the country name and type of consent)
<p>9. Final Metadata Data Cleaning</p>	<p>This folder houses all of the templates that have completed all of the data cleaning steps and is further organized into 3 sub-folders: 9.1 Common Steps for Data Cleaning and 9.2 EMM Survey Registry (For Youngminds).</p> <p>9.1 Common Steps for Data Cleaning</p>

	<p>This sub-folder houses all of the templates that will be produced after completing step 22 of the data cleaning process. The templates should be labelled using the following format: Country Name(Actual Date).xlsx</p> <p>9.2 EMM Survey Registry (For Youngminds) This sub-folder houses all the templates that are related to the EMM Survey Registry.</p> <ul style="list-style-type: none"> - For the template that is produced after completing step 27 of the data cleaning process, label it as: Country Name(Actual Date)Youngminds.xlsx - For the template that is produced after completing sub-step 1 of step 28, label it as: Country Name(Actual Date)NatCSVYoungminds.xlsx - For the template that is produced after completing sub-step 3 of step 28, label it as: Country Name(Date)NatCSVYoungminds.csv - For the templates that is produced after completing sub-step 4 of step 28, label it as, Country Name(Actual Date)SubnatCSVYoungminds.xlsx, for the Excel version, and as, Country Name(Actual Date)SubnatCSVYoungminds.csv, for the CSV version.
<p>10. References and Resources</p>	<p>This folder contains miscellaneous files that have been developed or consulted by WG1 and 2 for the purposes of the metadata compilation work.</p>
<p>NOTE: The date format for all the templates should be dd.mm.yy.</p>	

Appendix B: Sample Word Document (Template Feedback for First and Second Stage Quality Control)

Below is an example of how the template feedback document should be filled out by the reviewer when completing the first or second stage quality control. This specific example is based off of the template with 5 surveys submitted by the national representatives for Montenegro; their template, with the cells highlighted can be found here: [Appendix B - Montenegro Template \(5 Surveys\).xlsx](#).³⁹

TEMPLATE FEEDBACK (MONTENEGRO)

General Comments

- You have generally done a good job, but there are some issues with completeness or misunderstandings that need to be corrected for the 5 surveys you have already located and described, as well as for any additional surveys you will be adding to this template.
- The feedback is structured by sheet: Annotations, National Surveys, and Subnational Surveys. For each sheet, each of the variables and the responses provided were reviewed. Detailed comments have been provided below for any cell (i.e. variable or response) that requires your attention. These cells have also been highlighted in yellow in the attached Excel file.

Annotations Sheet

- We appreciate that this is not the final version of the excel file for Montenegro yet, but for the final version it would be useful to provide some comments on cell B9. This could be limited access to relevant surveys in Montenegro, no feedback received from the data producers, etc. If you faced no challenge, then write “No major challenge faced”.
- Cell B10 will need to contain a date in the next version.
- Cell B11 should have contained a date already, so that we understand from which date of data production you have started searching for surveys. As indicated in the guidelines, you should be searching for a survey that (at least) dates back since 1st January 2000, so at a minimum the date here should be 01/01/2000.
- Cells B13 to B15 also must be filled in already.

National Surveys Sheet

- Cells D6 & E6 need to contain some value. If the survey used no acronyms, then used the relevant missing code indicated in the codebook and on the excel file. Please note that the missing values are with a negative sign in front (e.g. -999 not 999)
- Row 69: If you have no comments to add here, it is best if you write “No comment”
- E74: as for the previous section, add -999 if no acronym was used
- E88-89: you indicate that this survey is part of an international survey, however you do not list the other countries that participate in this larger study on cell E80, and that’s contradictory. Maybe you added the information in the wrong cells and meant that the survey is longitudinal

³⁹ For security purposes, the link to this file has been removed in this published version of this guidelines document.

(multiple waves) only for Montenegro? If so, that would mean that the info should be filled in in cells E83-E85

- Row 98: ideally, if you have filled this section at all, you should use this cell to describe briefly what the larger study consisted of, just one sentence is enough.
- E119, E120, E121, C141, E141: I think you just missed these ones!
- In row 143 you use acronyms (e.g. RAE) that are not defined previously, it would be useful if you defined them somewhere, perhaps in the title of the survey (where you also use these acronyms), or in 3.7.
- C152, D152, E152: these cells need to be provided as well for these surveys.
- C154: should be filled in.
- B156, D156: should be filled in. If no sampling frame say “no sampling frame used”
- C160, D160, E160: please fill something here, even if it is “no additional comments on sampling”
- Row 163: the correct code for not applicable would be -999, consistent with other variables.
- Row 165: should only be the numbers, if you need to specify where these were individuals or households, please do so either on variable 4.6 or on variable 5.9
- Row 167: the absence of a response rate should be code as -99 and not 99 or it will seem that these surveys obtained a 99% response rate!
- E169: there is an inconsistency between this response (AAPOR RR1) and the fact that E167 says that you don't know what the response rate is. Can you double-check, please?
- Section 6: Here I think there might be some misunderstanding. It's been left empty for column B. However, I understand that this survey targeted three different groups (Roma, Ashkali and Egyptians), so these should be treated as 3 separate subgroups (unless in Montenegro are considered a single group? Which would be odd, I guess, but in which case you should at least fill in for sub-group 1). In columns C and D you have left empty most of the cells after just naming the relevant subgroups, but here you should add either some information about the samples for these subgroups or at least fill in with the relevant missing values for each cell.
- D234, D239: It seems you've missed this one.
- Row 244: here you should only add numbers and not the word “minutes” and the 99 should be -99 if you want to say not known, as otherwise it will be interpreted as 99 minutes.
- Row 246: please check that the 9 and 99 should not be -9 and -99. Are these 9 items or “don't know”?
- D248: please check the word marked in red as it seems a mistake?
- E253: info missing here
- Row 255: please try to add here some info on who should be contacted to gain access to the data when it is available by request
- Row 257: if not know, leave empty
- Row 267: I would imagine that the datasets will be at the very least available in the native language of Montenegro (which I think is Serbo-Croat), right? Please add the ISO code for the national language at least.
- D271: As by request, please add here who should be contacted for requests.
- Row 279 and 289: as above, at least in Serbo-Croat? If not sure, use variable 8.21 to explain why not

- Row 283: here info should be provided for those you've said are available somehow
- D295-D296: wouldn't this be UNICEF as well?

Subnational Surveys Sheet

- B19: you said n.a. (not available) for the NUTS/LAU code for Konik. You need, however, to put a code here. At the very least this would be the LAU1 code for Podgorica (as I understand that Konik is within the municipality of Podgorica), but it might be that MOSTAT has given Konik a LAU2 code as they have created codes also for "settlements" (see here: <https://www.monstat.org/userfiles/file/klasifikacije/statistical%20regions%20.pdf>). So I would suggest that you contact MOSTAT and request the list of codes for LAU1 and LAU2.
 - Section 2: please add the not applicable codes for the variables highlighted in yellow
 - B159: try to describe the sampling strategy used. From what you say in B161 it seems as if this was an attempt at undertaking a full census? If so, describe as such.
 - Section 4: please fill in the missing cells
 - B170: please provide the number for the attempted sample. If you think this was the original estimation of 4000 inhabitants of the camp, give that number.
 - B174: check if this should be -99. (All missing values start with a - sign).
 - B251: check if this should be -9
 - D276: if it is unavailable, how were you able to calculate the number of questions in the qu'aire?
 - D270: given what you say in D298, shouldn't this be yes?
-

Appendix C: Sample Emails (Template Feedback for First and Second Stage Quality Control)

First Stage Quality Control (No Serious Mistakes)

The email template below can be used to send feedback for a template that underwent a **first stage of quality control** and was rated as having **no serious mistakes** (i.e. all or almost all of the compulsory variables had responses and the variables were regularly coded correctly).

Subject: CA16111 Metadata Template Feedback - 5 Surveys (*Country Name*)

Dear (*Names of national representatives*),

Thank you for submitting your template for review. Overall, your template was (*quality of the template*) and I appreciate all the work and effort you have put in.

To this email, I have attached a Word document that provides detailed feedback about your template. I have also attached a copy of your template (see Excel file) with the entries/cells (where each entry/cell corresponds to a specific comment on the Word document) requiring your attention highlighted in yellow.

Thank you again for your contributions and please do not hesitate to reach out to me (*Reviewer's email address*) and the Sciences Po team (sshoc.project@sciencespo.fr) if you have any questions and/or concerns.

Warm regards,
(*Reviewer's name*)

First Stage Quality Control (Serious Mistakes)

The email template below can be used to send feedback for a template that underwent a **first stage of quality control** and was rated as having **serious mistakes** (i.e. many of the compulsory variables were missing responses without any explanation and/or the variables were regularly coded incorrectly).

Subject: CA16111 Metadata Template Feedback - 5 Surveys (*Country Name*)

Dear (*Names of national representatives*),

Thank you for submitting your template for review. Your template is a good starting point, though there were some issues with completeness or misunderstandings that will need to be corrected for the already coded surveys, as well as any new ones you add to your template. Once you have corrected your template using the feedback provided and have added a few more surveys to the template, please submit your template for a second round of review so we can provide you with some additional support and feedback as you progress with this metadata compilation task.

To this email, I have attached a Word document that provides detailed feedback about your template. I have also attached a copy of your template (see Excel file) with the entries/cells (where each entry/cell corresponds to a specific comment on the Word document) requiring your attention highlighted in yellow.

Thank you again for your contributions and please do not hesitate to reach out to me (*Reviewer's email address*) and the Sciences Po team (sshoc.project@sciencespo.fr) if you have any questions and/or concerns.

Warm regards,
(*Reviewer's name*)

Second Stage Quality Control

The email template below can be used to send feedback for a template that underwent a **second stage of quality control**.

Subject: CA16111 Metadata Template Feedback - 2nd Quality Control (*Country Name*)

Dear (*Names of national representatives*),

Thank you for submitting your template for a second review. Overall, your template was (*quality of the template and the overall progress made*). I sincerely appreciate all the work and effort you have put in.

To this email, I have attached a Word document that provides detailed feedback about your template. I have also attached a copy of your template (see Excel file) with the entries/cells (where each entry/cell corresponds to a specific comment on the Word document) requiring your attention highlighted in yellow.

Thank you again for your contributions and please do not hesitate to reach out to me (*Reviewer's email address*) and the Sciences Po team (sshoc.project@sciencespo.fr) if you have any questions and/or concerns.

Warm regards,

(*Reviewer's name*)

Appendix D: Sample Word Document (Template Feedback for Final Quality Control)

Below is an example of how the template feedback document should be filled out by the reviewer when completing the final quality control. This specific example is based off of the final template submitted by the national representatives for Norway; their template, with the cells highlighted, can be found here: [Appendix D - Norway Template \(Final\).xlsx](#).⁴⁰

FINAL TEMPLATE FEEDBACK (NORWAY)

General Comments

- The final template is in excellent shape. You have captured a decent amount of detail for each of the surveys. The comments below are more points of clarification that require your input/verification to finalize this final template.
- The feedback is structured by sheet: Annotations, National Surveys, and Subnational Surveys. For each sheet, each of the variables and the responses provided were reviewed. For any cell (i.e. variable or response) that requires your attention, detailed comments have been provided in the sections below. These cells have also been highlighted in yellow in the attached Excel file. Once a cell highlighted in yellow has been validated/corrected by you, please remove the highlight.
- The aforementioned Excel file also includes cells that are highlighted in a color other than yellow. These cells do not require any action by you, but a key/explanation is provided below for your reference:
 - For any cell that had a typing/data entry, spelling, grammatical, and/or punctuation error, it has been corrected by the reviewer and highlighted in pink. No comments have been provided in the sections below.
 - For any cell (open-ended responses only) that had terminology inconsistencies (within the cell itself and/or for the template as a whole), it has been corrected and highlighted in blue. No comments have been provided in the sections below.
 - For any cell that had no response, because the variable was non-compulsory and/or was not required to be filled out per the instructions provided by the Surveys Metadata Template Guidelines Document, it has been filled out by the reviewer and highlighted in green. No comments have been provided in the sections below.
 - For any cell that requires future data cleaning, but no action has been taken during this stage (e.g. the specific data cleaning step has not yet been defined), it has been highlighted in red. No comments have been provided in the sections below.

Annotations Sheet

- No comments.

⁴⁰ For security purposes, the link to this file has been removed in this published version of this guidelines document.

National Surveys Sheet

- B16-I16: All of the surveys are classified as a repeated cross-section. Is this correct? If this is the case, all of these surveys should have section 2 completed as a repeated cross-section survey is considered to be a part of a larger study (since it had multiple waves). Moreover, the other waves for each of the surveys are not currently coded on this template; if any of the other waves for each of the surveys meet the inclusion criteria set forth in the guidelines document, please add them to the template.
- Row 22: This is a new variable that was added based on feedback from the Palermo meeting and other COST Action meetings. Please review the responses to ensure they have been coded correctly.
- Row 149: This is a new variable that was added based on feedback from the Palermo meeting and other COST Action meetings. Please review the responses to ensure they have been coded correctly.
- G151: You indicated here that the survey coded in column G is like a sub-sample of a larger study. What do you mean by this? Also, which of the columns is the larger study? Please correct this cell accordingly.
- H151: You indicated here that there was a smaller sub-sample of Norwegians born to immigrant parents. Is this correct? If so, section 6 should be filled out for this survey.
- F-G158, I158: These cells were left blank. Please double check that they have been coded correctly.
- G-I243, G-I247: What seems to be listed (for the most part) are countries as opposed to languages. Do you know the actual languages, as opposed to countries? Please correct these cells to list languages only.
- G286, I286: What seems to be listed (for the most part) are countries as opposed to languages. Do you know the actual languages, as opposed to countries? Please correct these cells to list languages only.
- G-I291: You indicated that the survey questionnaire is available for individual research by request. Do you know who should be contacted to make such a request? Please code this cell accordingly.
- B-I293, C-I295: These variables were not required to be filled in per the guidelines document. However, for the metadata database, the programmers require these variables to have a response. Please double check that these cells have been correctly coded.
- G-I297: What seems to be listed (for the most part) are countries as opposed to languages. Do you know the actual languages, as opposed to countries? Please correct these cells to list languages only.
- B-I317: This variable was not required to be filled in per the guidelines document. However, for the metadata database, the programmers require these variables to have a response. Will you kindly provide a response for each survey?

Subnational Surveys Sheet

- Row 20: This is a new variable that was added based on feedback from the Palermo meeting and other COST Action meetings. Please double check that the response was correctly coded.

- Row 30: This is a new variable that was added based on feedback from the Palermo meeting and other COST Action meetings. Please double check that the response was correctly coded.
- B74, B76: This variable was not required to be filled in per the guidelines document. However, for the metadata database, the programmers require these variables to have a response. Will you kindly provide a response for both these cells?
- B78: To double check, since you were unable to locate and access the technical documentation for Norway, how were you able to obtain the information for this survey? Could you add a note about this in this cell?
- Section 2: Since LOCALMULTIDEM is a cross-national survey, information has been coded in this section. [NOTE: This section is highlighted in pink to denote that no action is needed by you.]
- B111: What is meant by immigrants from Norway? Does it refer to individuals who are born in Norway but have immigrant parents? Does it refer to a control group of the majority population (i.e. Norwegians who do not have a migrant origin)? Please provide a corrected/revised response in this cell.
- B154: This variable was not required to be filled in per the guidelines document. However, for the metadata database, the programmers require that this variable have a response. Will you kindly provide a response for this cell?
- B166: This variable was not required to be filled in per the guidelines document. However, for the metadata database, the programmers require that this variable have a response. Will you kindly provide a response for this cell?
- B252, B255: Pakistani is not a language that is recognized by ISO. Do you know which of the languages spoken in Pakistan were used for the interviews? Please correct the cell accordingly.
- B262: I think I've been able to locate the questionnaire for Norway (<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/24987>).
- B271, B273, B275, B277, B281, B292, B294, B297, B301, B303, B305: These variables were not required to be filled in per the guidelines document. However, for the metadata database, the programmers require these variables to have a response. Will you kindly double check the responses that have been coded?
- B308: You mentioned that you were able to find some information about this survey in a chapter. Which document is this chapter in? Will you add this detail to this cell?

Appendix E: Sample Emails (Template Feedback for Final Quality Control)

The email template below can be used to send feedback for a template that underwent the **final quality control**.

Subject: CA16111 Metadata Template Feedback - Final (*Country Name*)

Dear (*Names of national representatives*),

Thank you for submitting the final version of your template. I have now fully reviewed the template and found it to be (*overall quality of the template*). I sincerely appreciate all the work and effort you have put in to capture as much detail as possible for each of the surveys coded.

As indicated above, your template is in (*1-2 word description*) shape, but I do have some questions and/or points of clarification for a handful of variables and/or surveys. To this email, I have attached the following documents to help pinpoint the template items that require your assistance:

- A Word document with detailed information about which variables and/or surveys have questions and/or points of clarification identified. Specific instructions have also been provided for each variable and/or survey that has been flagged.
- A copy of your template (as an Excel file), with the flagged variables and/or surveys highlighted in yellow. Please make sure to use this version of the Excel file when making the requested edits/corrections (as stated in the above Word document). [NOTE: This Excel file also has cells that are highlighted in a color other than yellow. Such cells require no action by you, but they have been marked in this way to be transparent about the template review and cleaning processing adopted by the Sciences Po team.]

After reviewing the 2 feedback documents described above, please send me (*Reviewer's email address*) an email, putting Ami Saji (amikatherine.saji@sciencespo.fr) and the Sciences Po team email (sshoc.project@sciencespo.fr) in copy, with an estimated turnaround time for sending back the revised Excel file that incorporates all of the corrections/changes requested.

In addition to reviewing and correcting the final version of your template, I would like your assistance in obtaining consent for the use of the survey metadata. Specifically, we would need to obtain explicit written consent from any data archive/repository that is housing a survey documented on the final version of your template. Below are the specific steps you will need to take:

- Review the final version of your template and identify all the surveys that are currently being managed and stored by a data archive/repository (see variable 9.3 of the template).

- NOTE: If none of the surveys are being managed and stored by a data archive/repository, please email Ami Saji (amikatherine.saji@sciencespo.fr) and the Sciences Po team email (sshoc.project@sciencespo.fr) to confirm this in writing.
- Contact all of the relevant data archives/repositories by using the following email template: [Consent Email Template - Use of Metadata for EMM Survey Registry](#).⁴¹
 - NOTE: This email template will ask the data archives/repositories to provide or decline their consent via email. Consent can be provided or declined in English or in the native language.
- Send all of the responses you have received from the data archives/repositories by emailing Ami Saji (amikatherine.saji@sciencespo.fr) and the Sciences Po team email (sshoc.project@sciencespo.fr). For any data archives/repositories that have not responded to your consent email, please follow up with them after 1-2 weeks.

Finally, once your template has completed the final quality control process, we will be adapting the Excel file, so that it can be read by and uploaded to the EMM Survey Registry. For any template that has been added to the registry, we plan to credit/acknowledge everyone who has contributed to the metadata template work (i.e. all the names listed on the Annotations sheet of the metadata template) by displaying their names on the registry in the following 2 ways:

1. A country profile will be created for each country participating in the metadata work and for which the survey metadata information has been added to the registry (see example of the country profile for [Croatia](#)). As you will see, the first piece of information provided on the country profile is the names of the individuals (as indicated on the Annotations sheet) who had filled out the metadata template.
2. For each survey record (see example, [Voter survey 2015](#)), we have created a new section called, "11.Information on this compilation of metadata". Variable 11.1 of this section lists the names of the individuals (again, as indicated on the Annotations sheet) who had filled out the metadata template and the survey record.

If you have **NO OBJECTIONS** to having your name displayed in the specific ways outlined above, **NO ACTION is needed by you**. However, if in the future you decide that you would like to have your name removed from the registry (you have the right to make this request at any time), please follow the instructions provided in the paragraph below.

If you **DO NOT want to have your name displayed in the specific ways outlined above**, please send an email to Ami Saji (amikatherine.saji@sciencespo.fr) and the Sciences Po team email (sshoc.project@sciencespo.fr), by **(insert a date that is at least 1 week from the date this email is sent)** stating that you do not want your name displayed on the registry. If no such email is received by **(date selected)**, we will assume you are providing us with consent (for the time being).

⁴¹ For security purposes, the link to this file has been removed in this published version of this guidelines document.

Thank you again for your contributions and please do not hesitate to reach out to me at *(Reviewer's email address)* and/or Ami Saji if you have any questions and/or concerns.

Warm regards,

(Reviewer's name)

Appendix F: Protocol for Coding Variables with “Missing” Responses

The table below shows for each variable, the response options that can be used when a cell is missing a response. Cells in the last column that have been highlighted in yellow denote variables that are drop-down menu style (i.e. closed response) and require the response options to be modified.

An Excel version of this table can be found here: [Appendix F - Final Metadata Template: Coding Missing Responses](#).⁴²

Variable (red = "compulsory")	Response type (closed, open)	Subnational only	Response option(s) when missing
1. General identification information about the survey			
1.0. Country	Closed	No	N/A - must be filled out
1.1. ID number (leave blank)	Open	No	N/A - to be filled out by the central team
1.2. Acronym	Open	No	-9. Don't know -99. Information not available -999. Not applicable
1.3. Survey Name Eng.	Open	No	N/A - must be filled out
1.4. Survey Name Nat.	Open	No	N/A - must be filled out
1.5. Scope of survey	Closed	No	-9. Don't know
1.6. Which subnational level	Open	Yes	-9. Don't know -99. Information not available
1.7. Name of region(s) Eng.	Open	Yes	-9. Don't know -99. Information not available
1.8. Name of region(s) Nat.	Open	Yes	-9. Don't know -99. Information not available
1.8a. If subnational, add the NUTS/LAU code of the specific cities/regions	Open	Yes	-9. Don't know -99. Information not available
1.8b. If subnational, type of subnational area	Closed	Yes	-9. Don't know -99. Information not available
1.9. Representative of the population	Closed	No	-9. Don't know -99. Information not available
1.10. Type of survey	Closed	No	-9. Don't know
1.10a. If "other" in 1.10. specify	Open	No	-9. Don't know -99. Information not available -999. Not applicable
1.11. Starting date of survey	Open	No	-99. Information not available
1.12. End date of survey	Open	No	-99. Information not available
1.12a. Survey in development/not yet completed	Closed	No	-9. Don't know
1.13. Main topic(s) in the survey			

⁴² For security purposes, the link to this file has been removed in this published version of this guidelines document.

1.13.1.Asylum seekers and refugee issues	Closed	No	-9.Don't know
1.13.2.Citizenship and naturalization	Closed	No	-9.Don't know
1.13.3.Consumption and/or leisure	Closed	No	-9.Don't know
1.13.4.Demographic characteristics/behaviours	Closed	No	-9.Don't know
1.13.5.Discrimination, racism and/or xenophobia	Closed	No	-9.Don't know
1.13.6.Educational attainment/trajectory, human capital, skills	Closed	No	-9.Don't know
1.13.7.Family reunification, marriage, family relations	Closed	No	-9.Don't know
1.13.8.Gender relations, gender identity, sexuality	Closed	No	-9.Don't know
1.13.9.Housing/housing access	Closed	No	-9.Don't know
1.13.10.Health/health access	Closed	No	-9.Don't know
1.13.11.Human smuggling and trafficking	Closed	No	-9.Don't know
1.13.12.Identity (ethnic, national, racial, religious) and belonging	Closed	No	-9.Don't know
1.13.13.Income-related (and/or poverty)	Closed	No	-9.Don't know
1.13.14.Interethnic contact & conflict	Closed	No	-9.Don't know
1.13.15.Labour market integration	Closed	No	-9.Don't know
1.13.16.Language skills/training	Closed	No	-9.Don't know
1.13.17.Leisure, sports, arts	Closed	No	-9.Don't know
1.13.18.Legal status/administrative situation	Closed	No	-9.Don't know
1.13.19.Migration trajectory (past/future)	Closed	No	-9.Don't know
1.13.20.Political inclusion and participation, and social/political attitudes	Closed	No	-9.Don't know
1.13.21.Public attitudes about migration and migrants	Closed	No	-9.Don't know
1.13.22.Reasons for migration/migration drivers	Closed	No	-9.Don't know
1.13.23.Return migration	Closed	No	-9.Don't know

1.13.24.Social cohesion and/or civic engagement and/or networks	Closed	No	-9.Don't know
1.13.25.Space use, spatial consequences	Closed	No	-9.Don't know
1.13.26.Time use	Closed	No	-9.Don't know
1.13.27.Transnational patterns (e.g. remittances, travel, engagement with 'home' country politics, etc.) and diasporas	Closed	No	-9.Don't know
1.13.28.Unaccompanied migrant minors	Closed	No	-9.Don't know
1.13.29.Religion	Closed	No	-9.Don't know
1.13.30.Other	Closed	No	-9.Don't know
1.13.30a. If "other" in 1.13, specify	Open	No	-9.Don't know -99.Information not available -999.Not applicable
1.14. Main purpose of the survey			
1.14.1.Research/academic	Closed	No	-9.Don't know
1.14.2.Public policy	Closed	No	-9.Don't know
1.14.3.Government statistics	Closed	No	-9.Don't know
1.14.4.NGO/ non-profit organisations	Closed	No	-9.Don't know
1.14.5.Commercial	Closed	No	-9.Don't know
1.14.6.Other	Closed	No	-9.Don't know
1.14a. If "other" in 1.14, specify	Open	No	-9.Don't know -99.Information not available -999.Not applicable
1.15. Coverage of the target population in terms of age	Closed	No	-9.Don't know -99.Information not available
1.16. Coverage of the target population in terms of the sex of the respondents	Closed	No	-9.Don't know -99.Information not available
1.17. Comments relevant to variables in section 1	Open	No	-999.Not applicable
2. Information about the inclusion of the survey in the larger study			
2.1. ID number larger study (leave blank)	Open	No	-999.Not applicable (only if the survey is not part of a larger study; otherwise, this variable should have a response because it will be filled in by a central team)
2.2. Study Acronym	Open	No	-9.Don't know -99.Information not available -999.Not applicable

2.3. Name of the larger study Eng.	Open	No	-999.Not applicable (only if the survey is not part of a larger study; otherwise, this variable should have a response)
2.4. Name of the larger study Nat.	Open	No	-999.Not applicable (only if the survey is not part of a larger study; otherwise, this variable should have a response)
2.5. Name of other countries/regions/cities Eng.	Open	No	-9.Don't know -99.Information not available -999.Not applicable
In case of repeated/longitudinal surveys:			
2.6. Date of the first survey	Open	No	-99.Information not available -999.Not applicable
2.7. Frequency of waves/panels	Open	No	-9.Don't know -99.Information not available -999.Not applicable
2.8. Wave number	Open	No	-9.Don't know -99.Information not available -999.Not applicable
In the case of surveys that are a part of an international survey programme			
2.9. Date when the survey first became a part of international survey programme	Open	No	-99.Information not available -999.Not applicable
2.10. Frequency of waves since then	Open	No	-9.Don't know -99.Information not available -999.Not applicable
If sample is pooled:			
2.11. How many surveys pooled	Open	No	-9.Don't know -99.Information not available -999.Not applicable
2.12. Which other surveys pooled	Open	No	-9.Don't know -99.Information not available -999.Not applicable
2.12a. Pools emigrants from more than 1 country	Closed	No	-9.Don't know -99.Information not available -999.Not applicable
2.13. Any qualitative studies linked to survey	Closed	No	-9.Don't know -99.Information not available -999.Not applicable

			Please use -999. Not applicable if the survey does not belong to a larger study instead of "0.No". "0.No" should be used only when the survey belongs to a larger study and there is no qualitative component as part of the larger study.
2.14. If "Yes", describe the studies	Open	No	-9. Don't know -99. Information not available -999. Not applicable
2.15. Additional comments to section 2	Open	No	-999. Not applicable
3. Ethnic and Migrant Target Population			
3.1. EMM target population: which minority group(s)	Open	No	N/A - must be filled out
3.1a. EMM target population with terms standardized	Open	No	N/A - must be filled out by the central team
3.2 Was the EMM target population	Closed	No	-9. Don't know -99. Information not available
3.2a. If "other" in 3.2., describe which	Open	No	-9. Don't know -99. Information not available -999. Not applicable
3.3. Operationalization of target population			
3.3.1. Country of birth of respondent	Closed	No	-9. Don't know
3.3.2. Country of birth of parents/grandparents	Closed	No	-9. Don't know
3.3.3. Citizenship/nationality of respondent (current)	Closed	No	-9. Don't know
3.3.4. Citizenship/nationality of respondent (at birth)	Closed	No	-9. Don't know
3.3.5. Citizenship/nationality of parents/grandparents (current)	Closed	No	-9. Don't know
3.3.6. Citizenship/nationality of parents/grandparents (at birth)	Closed	No	-9. Don't know
3.3.7. Ethnic self-identification of respondent (one response allowed)	Closed	No	-9. Don't know
3.3.8. Ethnic self-identification of respondent (multiple responses allowed)	Closed	No	-9. Don't know
3.3.9. Ethnic self-identification of parents/grandparents	Closed	No	-9. Don't know

3.3.10.Mother tongue/language related question	Closed	No	-9.Don't know
3.3.11.Classification by interviewer	Closed	No	-9.Don't know
3.3.12.Classification by third agent /by proxy (e.g. by a government authority, and NGO, a social/cultural mediator, etc.)	Closed	No	-9.Don't know
3.3.13.Classification by geographical location	Closed	No	-9.Don't know
3.3.14.Through other means/characteristics	Closed	No	-9.Don't know
3.3.15.Information not available on definition of target population	Closed	No	-9.Don't know
3.3a. If "other means" in 3.3., describe which	Open	No	-9.Don't know -99.Information not available -999.Not applicable
3.4. Migrant/minority related questions			
3.4.1.Country of birth of respondent	Closed	No	-9.Don't know
3.4.2.Country of birth of parents	Closed	No	-9.Don't know
3.4.3.Country of birth of grandparents	Closed	No	-9.Don't know
3.4.4.Nationality of respondent (current)	Closed	No	-9.Don't know
3.4.5.Nationality of respondent (at birth)	Closed	No	-9.Don't know
3.4.6.Nationality of parents (current)	Closed	No	-9.Don't know
3.4.7.Nationality of grandparents (current)	Closed	No	-9.Don't know
3.4.8.Nationality of parents (at birth)	Closed	No	-9.Don't know
3.4.9.Nationality of grandparents (at birth)	Closed	No	-9.Don't know
3.4.10.Ethnic self-identification of respondent (one response allowed)	Closed	No	-9.Don't know
3.4.11.Ethnic self-identification of respondent (multiple responses allowed)	Closed	No	-9.Don't know
3.4.12.Ethnic self-identification of parents	Closed	No	-9.Don't know
3.4.13.Ethnic self-identification of grandparents	Closed	No	-9.Don't know
3.4.14.Mother tongue/language related question	Closed	No	-9.Don't know

3.4.15.Classification by interviewer	Closed	No	-9.Don't know
3.4.16.Information not available on migrant/minority related questions	Closed	No	-9.Don't know
3.5. Size of the EMM target pop. as a whole	Open	No	-9.Don't know -99.Information not available
3.6. Survey includes subgroup of majority pop.	Closed	No	-9.Don't know -99.Information not available
3.6a. Survey designed as a general population survey	Closed	No	-9.Don't know -99.Information not available
3.7. Additional comments to section 3	Open	No	-999.Not applicable
4. Sampling Method			
4.1. Sampling strategy - closed	Closed	No	-9.Don't know -99.Information not available
4.2. Sampling strategy - open	Open	No	-9.Don't know -99.Information not available
4.3. Sample design - full information	Open	No	-9.Don't know -99.Information not available
4.4. Sampling frame(s)	Open	No	-9.Don't know -99.Information not available
4.5. Sampling units	Open	No	-9.Don't know -99.Information not available
4.6. Comments on sampling methods	Open	No	-999.Not applicable
5. Sample Size for the overall survey			
5.1. Total gross/issued sample	Open	No	-9.Don't know -99.Information not available
5.2. Total net/achieved sample	Open	No	-9.Don't know -99.Information not available
5.3. Overall response rate	Open	No	-9.Don't know -99.Information not available
5.4. Overall response rate calculated	Closed	No	-9.Don't know -99.Information not available
5.5. If "other" in 5.4., describe	Open	No	-9.Don't know -99.Information not available -999.Not applicable
5.6. Comment on any known issues	Open	No	-999.Not applicable

5.7. Are weights provided	Closed	No	-9. Don't know -99. Information not available
5.8. If weights: Please describe	Open	No	-9. Don't know -99. Information not available -999. Not applicable
5.9. Additional comments to section 5	Open	No	-999. Not applicable
6. Sample sizes for any subgroups in which the survey is partitioned			
6a1. SG1 Name of sub-group	Open	No	-999. Not applicable
6a2. SG1 Gross/issued sample	Open	No	-9. Don't know -99. Information not available -999. Not applicable
6a3. SG1 Net/achieved sample	Open	No	-9. Don't know -99. Information not available -999. Not applicable
6a4. SG1 Response rate	Open	No	-9. Don't know -99. Information not available -999. Not applicable
6a5. SG1 Overall response rate calculated	Closed	No	-9. Don't know -99. Information not available -999. Not applicable
6a6. SG1 If "other" in 6a5., describe	Open	No	-9. Don't know -99. Information not available -999. Not applicable
6a7. SG1 Comment on any known issues	Open	No	-999. Not applicable
6b1. SG2 Name of sub-group	Open	No	-999. Not applicable
6b2. SG2 Gross/issued sample	Open	No	-9. Don't know -99. Information not available -999. Not applicable
6b3. SG2 Net/achieved sample	Open	No	-9. Don't know -99. Information not available -999. Not applicable
6b4. SG2 Response rate	Open	No	-9. Don't know -99. Information not available -999. Not applicable
6b5. SG2 Overall response rate calculated	Closed	No	-9. Don't know -99. Information not available -999. Not applicable
6b6. SG2 If "other" in 6b5., describe	Open	No	-9. Don't know -99. Information not available -999. Not applicable
6b7. SG2 Comment on any known issues	Open	No	-999. Not applicable
6c1. SG3 Name of sub-group	Open	No	-999. Not applicable

6c2. SG3 Gross/issued sample	Open	No	-9.Don't know -99.Information not available -999.Not applicable
6c3. SG3 Net/achieved sample	Open	No	-9.Don't know -99.Information not available -999.Not applicable
6c4. SG3 Response rate	Open	No	-9.Don't know -99.Information not available -999.Not applicable
6c5. SG3 Overall response rate calculated	Closed	No	-9.Don't know -99.Information not available -999.Not applicable
6c6. SG3 If "other" in 6c5., describe	Open	No	-9.Don't know -99.Information not available -999.Not applicable
6c7. SG3 Comment on any known issues	Open	No	-999.Not applicable
6d1. SG4 Name of sub-group	Open	No	-999.Not applicable
6d2. SG4 Gross/issued sample	Open	No	-9.Don't know -99.Information not available -999.Not applicable
6d3. SG4 Net/achieved sample	Open	No	-9.Don't know -99.Information not available -999.Not applicable
6d4. SG4 Response rate	Open	No	-9.Don't know -99.Information not available -999.Not applicable
6d5. SG4 Overall response rate calculated	Closed	No	-9.Don't know -99.Information not available -999.Not applicable
6d6. SG4 If "other" in 6d5., describe	Open	No	-9.Don't know -99.Information not available -999.Not applicable
6d7. SG4 Comment on any known issues	Open	No	-999.Not applicable
6e1. SG5 Name of sub-group	Open	No	-999.Not applicable
6e2. SG5 Gross/issued sample	Open	No	-9.Don't know -99.Information not available -999.Not applicable
6e3. SG5 Net/achieved sample	Open	No	-9.Don't know -99.Information not available -999.Not applicable
6e4. SG5 Response rate	Open	No	-9.Don't know -99.Information not available -999.Not applicable

6e5. SG5 Overall response rate calculated	Closed	No	-9. Don't know -99. Information not available -999. Not applicable
6e6. SG5 If "other" in 6e5., describe	Open	No	-9. Don't know -99. Information not available -999. Not applicable
6e7. SG5 Comment on any known issues	Open	No	-999. Not applicable
6.8. Additional comments to section 6	Open	No	-999. Not applicable
7. Data collection information			
7.1. Name of person/institution/institute that undertook fieldwork	Open	No	-99. Information not available
7.2. Data collection mode			
7.2.1. Face to face (PAPI)	Closed	No	-9. Don't know
7.2.2. Face to face (CAPI)	Closed	No	-9. Don't know
7.2.3. Telephone	Closed	No	-9. Don't know
7.2.4. Web / email survey	Closed	No	-9. Don't know
7.2.5. Paper self-administered (collected)	Closed	No	-9. Don't know
7.2.6. Paper self-administered (postal)	Closed	No	-9. Don't know
7.2.7. Other	Closed	No	-9. Don't know
7.2.8. Information not available	Closed	No	-9. Don't know
For personal interviews:			
7.3. Who interviewed	Closed	No	-9. Don't know -99. Information not available -999. Not applicable
7.4. Interviewers spoke migrant languages	Closed	No	-9. Don't know -99. Information not available -999. Not applicable
7.5. If "yes", which	Open	No	-9. Don't know -99. Information not available -999. Not applicable
7.5a. Migrant languages in ISO code	Open	No	-999. Not applicable
7.6. Questionnaire in migrant language	Closed	No	-9. Don't know -99. Information not available -999. Not applicable

7.7. If "yes", which	Open	No	-9. Don't know -99. Information not available -999. Not applicable
7.7a. Questionnaire in migrant language in ISO code	Open	No	-999. Not applicable
7.8. Average duration/length of interview	Open	No	-9. Don't know -99. Information not available
7.9. Number of questions	Open	No	-9. Don't know -99. Information not available
7.10. Additional comments to section 7	Open	No	-999. Not applicable
8. Availability			
8.1. Availability of the survey	Closed	No	4. Unavailable 5. Unknown availability
8.2. If available, where is the dataset stored	Open	No	-9. Don't know -99. Information not available -999. Not applicable
8.3. ID number of archive where dataset is stored	Open	No	-9. Don't know -99. Information not available -999. Not applicable
8.4. DOI for the dataset	Open	No	-9. Don't know -99. Information not available -999. Not applicable
8.5. Access to complete dataset	Closed	No	-9. Don't know
8.6. Access to portions of dataset	Closed	No	-9. Don't know -999. Not applicable (full dataset accessible)
8.7. Access to aggregate data results	Closed	No	-9. Don't know
8.8. Restrictions for data access, describe which	Open	No	-9. Don't know -99. Information not available -999. Not applicable
8.9. Dataset language(s) available	Open	No	-9. Don't know -99. Information not available
8.9a. Dataset languages in ISO code	Open	No	-999. Not applicable
8.10. Availability of survey doc. & q'aire	Closed	No	4. Unavailable 5. Unknown availability
8.11. If available, where is the survey doc. stored	Open	No	-9. Don't know -99. Information not available -999. Not applicable
8.12. ID number of archive where survey doc. is stored	Open	No	-9. Don't know -99. Information not available -999. Not applicable
8.13. DOI for the documentation	Open	No	-9. Don't know -99. Information not available -999. Not applicable
8.14. If the technical documentation is available, standard used for the documentation	Closed	No	-9. Don't know -99. Information not available -999. Not applicable
8.15. Language(s) in which technical survey documentation is available	Open	No	-9. Don't know -99. Information not available

8.15a. Language(s) of technical survey documentation in ISO code	Open	No	-999.Not applicable
8.16. Availability of the survey questionnaire for individual research use	Closed	No	4.Unavailable 5.Unknown availability
8.17. If available, where is the survey q'aire stored	Open	No	-9.Don't know -99.Information not available -999.Not applicable
8.18. Identification number of the archive / repository where questionnaire is stored	Open	No	-9.Don't know -99.Information not available -999.Not applicable
8.19. DOI for the questionnaire	Open	No	-9.Don't know -99.Information not available -999.Not applicable
8.20. Language(s) in which survey questionnaire is available	Open	No	-9.Don't know -99.Information not available
8.20a.Language(s) of survey questionnaire in ISO code	Open	No	-999.Not applicable
8.21. Comments relevant to variables in section 8 (on data availability)	Open	No	-999.Not applicable
9. Data producers, owners, distributors and citations			
9.1. Institution/team responsible for data production	Open	No	-9.Don't know -99.Information not available
9.2. Institution/team that owns the data	Open	No	-9.Don't know -99.Information not available
9.3. Institution/team distributing the dataset	Open	No	-9.Don't know -99.Information not available
9.4. Contact details for queries/request	Open	No	-9.Don't know -99.Information not available
9.5. Citation for dataset	Open	No	-9.Don't know -99.Information not available
9.6. Citation for technical documentation	Open	No	-9.Don't know -99.Information not available
9.7. Citation(s) for any other publications	Open	No	-9.Don't know -99.Information not available -999.Not applicable
9.8. Comments:	Open	No	-999.Not applicable
10. Additional Information			
10.1 Data quality (for internal use only)	Closed	No	-9.Don't know
10.2. Add.info. Data quality	Open	No	-999.Not applicable
10.3. Sources of information	Open	No	N/A - must be filled out
10.4. Any other comments about the survey	Open	No	-999.Not applicable

Appendix G: Standardization Rules for Proper Nouns

Proper nouns need to be standardized within each template and across all of the templates. The table below provides specific instructions on how to handle each type of proper noun:

Type of proper noun	Format
Survey name	Full name of the survey, exactly as it appears. [Note: Specific instructions about how to code the survey name were provided in the guidelines document] Correct example: European Union Minorities and Discrimination Survey Incorrect example: EU Minorities and Discrimination Survey
Person's name	First name followed by the surname. Correct example: Jane Smith Incorrect example: Smith, Jane
Institutional name	Full name followed by the acronym (do not make one up) in parentheses. Correct example: United Nations High Commissioner for Refugees (UNHCR) Incorrect example: UNHCR
Geographical location (e.g. country, city, region)	Full legal name; no abbreviations. Correct example: France Incorrect example: FR