Compilers' Manual on Forced Displacement Statistics

Expert Group on Refugee, Internally Displaced Persons and Statelessness Statistics (EGRISS)

EGRISS
EXPERT GROUP ON REFUGEE, IDP AND STATELESSNESS STATISTICS

(COVER)
COLOMBIA.
Embera indigenous communities of Bogota return to Alto Andágueda in 2023.
©Unit for Comprehensive Care and Reparation for Victims of Colombia (UARIV)

SOUTH AFRICA.
2022 Population and Housing Census.
©Statistics South Africa (Stats SA)

TABLE OF CONTENTS

ACRONYMS

ADAPT	Advanced Data Planning Tool
ANC	Antenatal care
ARMSTAT	Statistical Committee of Armenia
B&H	Bosnia and Herzegovina
CAR	Central African Republic
CDR	Call Detail Records
CEAVE	Executive Commission for Attention to Victims of the State of Chihuahua
CENTROSTAD	Central American Statistical Commission
CM	Compilers' Manual
CODs	Common Operational Datasets
COVID-19	Coronavirus Disease 2019
CPR	Central Population Register of Norway
CRC	Capture-recapture
CRF	Central Register of Foreigners
DANE	Colombia's National Department of Statistics
DDPR	Database of Displaced Persons
Destatis	German Federal Statistical Office
DHS	Demographic Household Survey
DOB	Date Of Birth
DTM	Displacement Tracking Matrix
EAs	Enumeration Areas
EDAUUR	Working Group on Useful and Usable Data and Analysis
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
EGRISS	Expert Group on Refugee, Internally Displaced Persons and Statelessness Statistics
ENPOVE	Survey Directed to the Venezuelan Population Residing in Peru
ESCWA	United Nations Economic and Social Commission for West Asia
ESS	European Statistical System
ESS	Ethiopian Statistics Service
EU	European Union
Eurostat	Statistical Office of the European Union

FANTA	Food and Nutrition Technical Assistance
GCR	Global Compact on Refugees
GIS	Geographic Information System
GPS	Global Positioning System
HDX	Humanitarian Data Exchange
HLG-MOS	High Level Group for the Modernisation of Official Statistics
HoWStat	Household Welfare Statistics Survey
IAEG-SDGs	Inter-agency and Expert Group on SDG Indicators
IASC	Inter-Agency Standing Committee
IDMC	Internal Displacement Monitoring Centre
IDP	Internally Displaced Person
ILO	International Labour Organisation
IM	Information Management
IMF	International Monetary Fund
INEGI	National Institute of Statistics and Geography of Mexico
IOM	International Organization for Migration
IRIS	International Recommendations on IDP Statistics
IROSS	International Recommendations on Statelessness Statistics
IRRS	International Recommendations on Refugee Statistics
IVR	Interactive Voice Response
JDC	World Bank – UNHCR Joint Data Center on Forced Displacement
JIPS	Joint IDP Profiling Service
JMP	Joint Monitoring Programme for Water Supply, Sanitation and Hygiene
LFS	Labour Force Survey
MEDSTAT	Mediterranean Statistical Cooperation Program
MICS	Multiple Indicators Cluster Surveys
MOUs	Memorandum of Understanding
NCDC	National Center for disease Control and Public Health
NGO	Non-Governmental Organisation
NSDS	National Strategy for the Development of Statistics
NSI	National Statistical Institute
NSO	National Statistical Office
NSS	National Statistical System
OCHA	Organization of Humanitarian Affaires
OECD	Organisation for Economic Co-operation and Development
ONS	Office for National Statistics
OSM	OpenStreetMap

PES	Post Enumeration Survey
PIN	Personal Identification Number
PNV	Pneumococcal Vaccination coverage
PSU	Primary Sampling Units
QCRI	Qatar Computing Institute
RDD	Random Digit Dialing
RDF	Refugee Data Finder
RDS	Respondent Driven Sampling
RRS	Refugees and Returnees Service
SDG	Sustainable Development Goal
SESRE	Socio-Economic Survey of Refugees in Ethiopia
SESRIC	Statistical, economic and social research and training centre for Islamic countries
SIM	Subscriber Identity Module
SNBS	Somalia National Bureau of Statistics
SSC	Sector Statistics Committees
UDI	Norwegian Directorate of Immigration
UID	Unique Identifiers
UK	United Kingdom
US	Unites States of America
UN	United Nations
UNECA	United Nations Economic Commission for Africa
UNECE	United Nations Economic Commission for Europe
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNFPA	United Nations Population Fund
UNHCR	United Nations High Commissioner for Refugees
UNICEF	United Nations Children's Fund
UNRWA	United Nations Relief and Works Agency for Palestine Refugees in the Near East
UNSC	UN Statistical Commission
UNSD	United Nations Statistics Division
UrbAN-S	Urban Analysis Network Syria
WFP	World Food Programme
WHO	World Health Organization

ACKNOWLEDGEMENTS

The revised version of the Compilers' Manual (CM) has resulted from the collaborative efforts of the Expert Group on Refugee, IDP and Statelessness Statistics (EGRISS).

Since the group's establishment in 2016 by the United Nations (UN) Statistical Commission (E/2016/24), the development of a CM containing clear operational instructions on how to collect refugee statistics, complementing the international recommendations, has been highlighted as a priority. In 2020, a first draft of the CM was presented and formally welcomed by the UNSC (Decision 51/116). In 2022, members of the EGRISS Technical Subgroup on 'Methodology and Guidance' embarked on a process to revise and update the document.

Special thanks go to Jacob Wilcock for his technical editor role, as co-lead, to Felix Schmieding, Senior Statistician with the UNHCR-Joint Data Center on Forced Displacement (JDC), and Mary Strode for her work drafting the original version of the CM.

The following country, regional and international organizations contributed to the featured case studies: Bosnia and Herzegovina Agency for Statistics, Central Statistical Agency of Ethiopia, Colombia's National Department of Statistics (DANE) and Unit for the Attention of Victims, German Federal Statistical Office (Destatis), Institute of Social Studies and Analysis in Georgia, National Institute of Statistics and Geography of Mexico (INEGI), Somalia National Bureau of Statistics, Statistics Norway, Statistical Committee of the Republic of Armenia (Armstat), National Statistics Office of Georgia, Ugandan Bureau of Statistics and the Joint IDP Profiling Service (JIPS).

The following country, regional and international organizations supported the review process of the document as members: Armstat, Destatis, Hungarian Central Statistical Office, National Bureau of Statistics of Nigeria, INEGI, Palestinian Central Bureau of Statistics, Statistics Norway, International Organization for Migration (IOM), JIPS, Organization for Economic Cooperation and Development, Statistical Office of the European Union (Eurostat), United Nations Children's Fund and the World Bank.

The CM was developed in close cooperation with all EGRISS members and partners. For their guidance throughout the revision, special mention goes to the Steering Committee members: Kurdistan Region Statistics Office, DANE, Philippine Statistics Authority, Statistics Norway, Victims Unit, African Union, Eurostat, Internal Displacement Monitoring Centre, IOM, JIPS, the UN Refugee Agency, United Nations Statistics Division, World Bank, JDC.

The EGRISS Secretariat supported the process, coordinating exchanges and making the publication possible. For their work, Natalia Krynsky Baal, Filip Mitrovic, Charis Sijuwade and Julieta Mendive are recognized.

For their financial contributions, made available through the European Commission department for International Partnerships, thanks are extended to JDC.

Words of appreciation are also extended to Rec Design for their design work, specifically to Renata Figueiredo.

INTRODUCTION

Background

1. The numbers of refugees, asylum seekers and internally displaced persons (IDPs) have increased rapidly in recent years. Almost every country in the world is affected by forced displacement either as a source, point of transit, or host of refugees, asylum seekers or IDPs, making forced displacement a global phenomenon. Developing countries are disproportionately affected by forced displacement.
2. The scale of the problem is growing. By the end of 2021, the number of people forcibly displaced due to persecution, conflict, violence, human rights violations or events seriously disturbing public order had grown to 89.3 million – more than doubling over the decade[1]. An additional 5.9 million persons were internally displaced by disasters at the end of 2021[2].
3. Forced displacement has gained prominence on the international agenda. The New York Declaration for Refugees and Migrants, adopted by the United Nations General Assembly in 2016, recognizes the unprecedented level of human mobility and acknowledges the shared responsibility to manage large movements of refugees and migrants through international cooperation (United Nations, 2016). It also reaffirms the intention of Member States to realize the full potential of the 2030 Agenda for Sustainable Development for refugees and migrants. In 2018, the United Nations General Assembly affirmed the Global Compact on Refugees (GCR)[3] a blueprint for governments, international organizations, and other stakeholders to ensure that host communities get the support they need and that refugees can lead productive lives. The 2020 report of the UN Secretary General's High-Level Panel on Internal Displacement[4] and subsequent Action Agenda on Internal Displacement (2021)[5], articulated similar recommendations to enhance global responses to internal displacement.
4. With the growing prominence of forced displacement internationally, there is increasing interest at national and international levels in statistics on refugees, asylum seekers, IDPs and related populations. The GCR highlights the critical importance of reliable, comparable, and timely data on refugees and advocates for the inclusion of refugees and host communities within national data and statistical collection processes. However, often the data available on refugees and IDPs are not generated as official statistics produced by the statistical systems of national governments but rather arise from operational data

1 UNHCR – Global Trends, Forced Displacement in 2021
2 Internal Displacement Monitoring Centre – Global Internal Displacement Database
3 UNHCR - Global Compact on Refugees – Booklet
4 Shining a Light on Internal Displacement: A Vision for the Future (internaldisplacement-panel.org)
5 Secretary-General's Action Agenda on Internal Displacement (un.org)

that are collected by a variety of actors as part of their (primarily) humanitarian response. This is data collected for the purpose of supporting displaced people, rather than for statistical purposes and so there can be no expectation that they meet the standards set for official statistics.

5. Including displaced populations in official statistics makes them visible in policy discussions, both at a national and international levels. Official statistics can help monitor the stocks and flows of forcibly displaced populations and provide analysis of important socioeconomic indicators that illustrate how the rights of displaced populations are being met, to inform understanding of their well-being, needs, integration and inclusion – for example in education and health systems. Conversely, incomplete or inadequate statistics on displaced populations can undermine estimates of total population stocks (through gaps in understanding of births, deaths or migration) which in turn affects other statistics, including those measured in per capita terms. Producers of official statistics are committed to delivering the data necessary to monitor the Sustainable Development Goals (SDGs) and displaced populations need to be visible in the SDGs. At the 51st UN Statistical Commission in 2020, a specific indicator on refugees[6] was included among the SDG indicators and the Inter-Agency and Expert Group on SDG Indicators recommends disaggregation by forced displacement status for 12 priority SDG indicators[7].

6. To enable the production of accurate and comparable statistics on forcibly displaced populations, the UN Statistical Commission (UNSC) established the Expert Group on Refugee Internally Displaced Persons and Statelessness Statistics (EGRISS)[8]. As of December 2022, group comprises 54 national statistical and immigration authorities and 34 regional/international organizations[9]. EGRISS has produced two sets of statistical standards: International Recommendations on Refugee Statistics (IRRS)[10], endorsed by the UNSC in 2018, and the International Recommendations on IDP Statistics (IRIS)[11], endorsed

6 SDG indicator number 10.7.4. UNHCR is the custodian agency for the indicator, which it calculates from figures reported to it by countries.

7 See pages 18-19 of https://unstats.un.org/unsd/statcom/50th-session/documents/BG-Item3a-Data-Disaggregation-E.pdf

8 Membership - EGRISS (egrisstats.org)

9 Countries and territories include: Afghanistan, Austria, Azerbaijan, Bangladesh, Belgium, Bosnia and Herzegovina, Cambodia, Cameroon, Canada, Chad, Colombia, Côte d'Ivoire, Djibouti, Ecuador, Egypt, Ethiopia, Germany, Georgia, Greece, Hungary, Kurdistan region of Iraq, Jordan, Kenya, Kosovo (all references to Kosovo in this paper should be understood to be in the context of Security Council resolution 1244 (1999)), Lebanon, Libya, Malaysia, Mali, Mexico, Morocco, the Niger, Nigeria, Norway, Pakistan, Palestine, the Philippines, Somalia, South Africa, Thailand, Turkey, Uganda, the United Kingdom of Great Britain and Northern Ireland, Ukraine and the United States of America. International organizations include: African Development bank; Central American Statistical Commission (CENTROSTAD); the European Commission (Directorate-General for Migration and Home Affairs and Directorate-General for Employment, Social Affairs and Inclusion); the European Asylum Support Office; the Economic and Social Commission for Western Asia; the Statistical Office of the European Union (Eurostat); the Internal Displacement Monitoring Centre (IDMC); the International Organization for Migration (IOM); the Joint Internally Displaced Person Profiling Service (JIPS); the Organization of Humanitarian Affaires (OCHA); the Organization for Economic Cooperation and Development (OECD); the Household International Migration Surveys in the Mediterranean Countries; Euro-Mediterranean statistics cooperation project (MEDSTAT); Paris21; Statistical, economic and social research and training centre for Islamic countries (SESRIC); United Nations Economic Commission for Africa (UNECA); the United Nations Population Fund (UNFPA); the Office of the United Nations High Commissioner for Refugees (UNHCR); the United Nations Children's Fund (UNICEF); the United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA), UN SPIDP OHCHR; the Statistics Division, Department of Economic and Social Affair (UNSD)s; the World Bank; the World Food Programme (WFP) and Academia.

10 International Recommendations on Refugee Statistics (IRRS) - EGRISS (egrisstats.org)

11 International Recommendations on IDP Statistics (IRIS) - EGRISS (egrisstats.org)

by the UNSC in 2020. It has developed a similar set of recommendations on statelessness statistics (IROSS), endorsed by the UNSC in 2023[12].

7. This Compilers' Manual complements the IRRS and IRIS. In time, it will be updated to complement the International Recommendations on Statelessness Statistics (IROSS).

Purpose and target audience of this Compilers' Manual

8. This manual provides a practical and easy-to-use entry point to the content of IRRS and IRIS, facilitating their operationalization in regular statistical workflows. It is aimed primarily at technical personnel in National Statistical Systems. It complements the statistical frameworks in the two Recommendations, providing a set of Use Cases that summarise the relevant information contained in the Recommendations, with links to further reading. Case Study examples of good practice in the production of refugee and IDP statistics are interwoven throughout the manual.

9. With a view to being succinct and user-friendly, the Compilers Manual makes a deliberate effort to *not* provide comprehensive discussions of statistical production cycles and business processes, but rather focuses only on the elements that are specific to refugee and IDP contexts. For example, the Use Cases related to sample surveys discuss sampling and questionnaire design (both of which require specific action when surveying refugees and IDPs) but do not cover data processing in any detail because data processing in forced displacement settings is no different from best practice in other survey settings. This logic extends to the fact that the Manual does not provide generic discussions of collecting data in conflict-affected locations – which may be a reality of IDP and refugee data production in some countries, but it is not unique to this area of statistics and therefore not a focus of this Manual[13]. Similarly, the manual does not discuss the COVID-19 pandemic, which had a significant impact on official statistics data collections, although the pandemic may be the context for some case studies that illustrate how challenges specific to refugee or IDP data collection were overcome. Finally, the Compilers Manual will often refer to and summarize IRRS and IRIS content, but not reproduce it at length.

10. The Manual is intended to be a 'living document' which will be amended and extended as the body of expertise and knowledge develops worldwide.

Use Cases and structure of the Manual

11. This Manual is structured along the following Use Cases[14] which should cover most scenarios in which National Statistical Offices will consult the IRRS and IRIS. Further Use Cases may be added, or existing ones updated, in future versions of the Manual.

    A. Including refugees or IDPs in a population census

12 International Recommendations on Statelessness Statistics (IROSS) - EGRISS (egrisstats.org)

13 Other sources of guidance on data collection in fragile or conflict locations are available, for example *Data Collection in Fragile States* (World Bank, 2020).

14 Given the target audience of this Manual as stated in the preceding section, coordination of international partners is not a Use Case foreseen or discussed in this Manual.

B. Including refugees in a sample survey of the national population, or running a stand-alone survey of refugees

C. Including IDPs in a sample survey of the national population, or running a stand-alone survey of IDPs

D. Using Government administrative data

E. Sources of operational data from humanitarian organisations

F. Non-traditional data sources

G. Coordinating and planning refugee and IDP statistics in national statistical systems

The annexes provide in-depth case studies and additional tools and techniques which were not specifically developed for forced displacement contexts but which may nevertheless be relevant to producers of refugee and IDP statistics.

12. There is overlap between the Use Cases, in particular between Use Case B (including refugees in sample surveys) and Use Case C (including IDPs in sample surveys). Overlapping material has been repeated in each Use Case. This is intentional to enable the Use Cases to function as stand-alone resources so that most users will be able to access all the information for the particular scenario they are interested in by downloading a single Use Case, rather than reading across multiple Use Cases. Use Case A has not been split into separate Use Cases covering refugees and IDPs because the additional content and considerations recommended in the context of a census does not vary significantly depending on whether the user is interested in refugees or IDPs (or both).

13. For most users of the manual, it is expected that it will be self-explanatory which Use Case to refer to. For statisticians who are aiming to begin collecting data on refugees or IDPs but who don't have a data source in mind, Table G.1 (paragraph 228) may serve as a suitable starting point, providing a guide to the different categories of data source that can be used to produce statistics on displaced populations.

Key definitions and concepts

Definitions of refugees and IDPs

14. A refugee is defined in the 1951 *Convention Relating to the Status of Refugees and 1967 Protocol Relating to the Status of Refugees* as someone "who, owing to a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group, or political opinion, is outside the country of his nationality, and is unable or, owing to such fear, is unwilling to avail himself of the protection of that country". The IRRS distinguishes between refugees and two refugee related populations, for the purposes of statistical measurement. The status of these population groups is primarily classified in legal terms by their claim to refugee status, or by their descent from someone seeking international protection:

- Persons in need of international protection include individuals in a country other than their own who are seeking or who have received international protection:
  - » prospective and current asylum seekers;
  - » persons with determined protection status, including refugees;
  - » others in refugee-like situations.

- Persons with refugee background include persons who are not currently in need of international protection but who have a refugee background:
  - » naturalized former refugees;
  - » children born of refugee parents;
  - » reunified refugee family members from abroad;
  - » others with a refugee background.
- Persons returned from abroad after seeking international protection include persons who have returned to their home country after seeking international assistance abroad:
  - » repatriating refugees and asylum seekers;
  - » other persons returning from international protection abroad.

15. The definition of IDPs comes from the *Guiding Principles on Internal Displacement*, which states that IDPs are 'persons or groups of persons who have been forced or obliged to flee or to leave their homes or places of habitual residence, in particular as a result of or in order to avoid the effects of armed conflict, situations of generalised violence, violations of human rights, or natural or human-made disasters, and who have not crossed an internationally recognised state border'. The IRIS distinguishes between IDPs and IDP-related populations, and further categorises IDPs by the location where they reside. It is worth noting that once a person becomes an IDP, unless they die or emigrate (for at least 12 months) they remain an IDP even if they return to their place of habitual residence, until they have overcome their displacement related vulnerabilities – at which point they enter the IDP-related population:

- IDPs - Persons who have displacement-related protection needs:
  - » IDPs who remain in locations of displacement;
  - » IDPs who have returned to their place of habitual residence (IDPs in locations of return);
  - » IDPs who have settled elsewhere in the country (IDPs in other settlement locations).
- IDP-related populations:
  - » Children born after displacement to at least one IDP parent;
  - » Other non-displaced family members of IDPs;
  - » Those who have overcome key displacement-related vulnerabilities.

Stocks, flows and socio-economic indicators

16. A key aim of collecting data on refugees is to be able to analyse stocks and flows:

- Stocks - A stock refers to the total number of individuals who have a specific set of characteristics, at a specific moment in time. These characteristics, put together, define the status of these individuals. For example, counts or proportion of population who fall into the different categories of refugees or IDPs referenced in paragraphs 14 and 15.
- Flows - Flows affect this stock over time, either by making it increase ("in-flows") or decrease ("out-flows"). In-flows correspond to the number of individuals who acquire the characteristics of those in scope during a specified time period, including through birth. For out-flows, it is the number of people who cease to be in the stocks, e.g. by changes in status, death, or out-migration.

17. There can also be 'in-flows' and 'out-flows' between the sub-stocks, related to changes in legal status. For example, an asylum seeker who is granted refugee status is an outflow from the sub-stock of 'Asylum seekers', and an inflow into 'Persons with determined status' (see IRRS, Figure 3.1).

18. Official statistics on forcibly displaced populations also need to support analysis of their socio-economic characteristics. Socio-economic indicators are used to assess refugees' needs and integration into host communities, and IDPs progress towards overcoming key displacement-related vulnerabilities.

19. IRRS and IRIS provide more details on the definitions of refugees and IDPs and the statistical frameworks around their measurement.

Consideration of ethical principles

20. Collecting data on displaced populations presents sensitivities, as anyone who identifies themselves as such could face threats from doing so or be stigmatised. When collecting data on displaced populations, all the relevant laws and standards on data collection and data protection must be applied, including those set out in relevant national laws, in the *UN Fundamental Principles of Official Statistics and the UN Personal Data Protection Privacy Principles*. All enumerators, coordinators, data managers, and anyone involved in the development of the official statistics must be aware and follow such principles[15].

15 See for example UN Fundamental Principles of Official Statistics, UN Personal Data Protection and Privacy Principles, IOM data protection manual, Handbook on Data Protection in Humanitarian Action

USE CASE

A

INCLUDING REFUGEES OR IDPS IN A POPULATION CENSUS

REPUBLIC OF ARMENIA.
Population Census 2022 training.
©Statistical Committee of the Republic of Armenia (Armstat)

EGRISS

The census – general background

21. The population census is a key data source providing benchmarks of the population residing in a country at a certain point in time, so it is important that displaced people are included in census data collection. The census provides an opportunity to produce accurate statistics on stocks of refugees and IDPs, analysis of their characteristics and comparisons with the wider population. It can help understand flows, through collecting data about a person's previous locations. While the infrequency of censuses and the time lag before data become available can impact on the relevance of the data, the data on displaced people collected through a (usually decennial) census can be used to compare with and adjust other more frequent data sources. Good coverage of displaced people in the census increases its value as a sampling frame for subsequent household surveys.

22. The UN's *Principles and Recommendations for Population and Housing Censuses (Revision 3)* issues recommendations and good practice for census taking and is widely used by national statistical offices in countries throughout the world. The recommendations provide guidance on the main characteristics of population and housing censuses, general material on census methodology, planning, and operations, and more detailed guidance on the content of censuses, in terms of topics for population and housing. As shown in Box A.1, the recommendations very explicitly and unequivocally address the inclusion of refugees, asylum seekers and IDPs in census data collection.

Box A.1

Guidance from *Principles and Recommendations for Population and Housing Censuses* (Revision 3, UN 2017) on including displaced populations

The *Principles and Recommendations for Population and Housing Censuses* makes the following key recommendations concerning displaced populations

Concepts relating to the place of enumeration (paragraph 2.53)

> "There are various population groups for which some uncertainty may arise about their inclusion in the usual resident population. The following persons would generally be considered in the usually resident population:
>
> ... (e) Persons who may be illegal, irregular or undocumented migrants, as well as asylum seekers and persons who have applied for or been granted refugee status or similar types of international protections, provided that they meet the criteria for usual residence in the country"

Usual resident population count (paragraph 4.37)

> "Usual residents may or may not have citizenship of the country, and they may also include undocumented persons, applicants for asylum or refugees. Usual residents then may include foreigners who reside (legally or illegally), or intend to reside, in the country continuously for either most of the last 12 months or for 12 months or more, depending on the definition of place of usual residence that is adopted by the country."

Difficult to enumerate populations (paragraph 4.48)

> "The following difficult-to-enumerate groups are relevant to the production of any population count:
>
> ... (d) Refugees, asylum seekers and internally displaced persons. Refugee populations, asylum seekers and internally displaced persons (in and outside camps) should be enumerated and their numbers presented separately, allowing calculation of country population excluding such groups, when such a population count is required for non-demographic purposes."

Total population (core topic) (paragraphs 4.84 and 4.85)

> "The description should show clearly whether each group listed below was or was not included in the total. If the group was enumerated and identified as a separate group, its magnitude should be given; if it was not enumerated, an estimate of its size and the method of estimation should be given, if possible. If any group is not represented at all in the population, this fact should be stated and the magnitude of the group should be shown as "zero". This may occur particularly with groups (a), (b), (d) and (n) described below.
>
> The groups to be considered are: ...(n) refugees in camps."

The *Principles and Recommendations* also specifically list refugee camps and camps for internally displaced people in its definition of collective living quarters (paragraph 4.424)

23. The guidance presented in this Manual (and in IRRS and IRIS) is entirely consistent with *Principles and Recommendations for Population and Housing Censuses* but provides additional advice to ensure appropriate coverage and identification of refugees and / or IDPs. The UNECE also made recommendations for the 2020 Population Census round which covered many of the core variables required to screen the population for Refugees and / or IDPs, although several of the items are non-core topics which are at the discretion of countries to add.

24. This Use Case addresses the priorities for including displaced people in the census. The key things to consider are:

- Include questions to identify displaced people in the design of the questionnaire –and where space permits capture additional information, for example on date and location of displacement.
- Ensure that fieldwork is designed to include displaced people, by adapting to cover the locations where they are likely to be found.
- Ensure enumerator training covers refugees and / or IDPs, including the relevant definitions and concepts, practical considerations and data protection.

25. The guidance and advice in this Use Case is focussed on the traditional population census format, conducted through full enumeration of the population, rather than alternative formats such as those that focus on using administrative data. However, other Use Cases may contain information relevant for other formats – for example Use Case D on administrative data.

Questionnaire design

26. The priority in designing a census questionnaire that takes account of displaced populations is to include questions that enable identification of refugees and / or IDPs in the data. The sets of core topics recommended in the *Principles and Recommendations for Population and Housing Censuses* go some way towards achieving this – but on their own these topics are not sufficient to ensure accurate identification of all displaced people and it is necessary to include additional questions.

27. The precise design of the census questionnaire varies from country to country, drawing on the *Principles and Recommendations for Population and Housing Censuses* but adapting to reflect the relevant national context and the extent to which there are displaced populations. If it is important to capture the entire IDP population, for example to inform sampling frames for future surveys, it is recommended that a specific set of questions on forced displacement is asked (see Box A.3). Given space constraints on census forms this is only likely to be justifiable in cases where there have been major internal displacements.

28. The following boxes outline the core topics and additional questions that are recommended for identifying refugees (Box A.2) and then adapting this to cover IDPs (Box A.3). The questions should be asked of all individuals, not just adults. They should also be asked of those living in camps and other temporary accommodation: it is important not to assume that those living in camps are always refugees or IDPs.

29. There is benefit in considering how personal identifiers collected through the census could be used through record linking to match respondents to recognized refugee and / or IDP administrative records. This is covered in more detail in Use Case D: Using Administrative Data.

Identifying refugees

30. The questions in Box A.2 support identification of most populations in the scope of the IRRS but with some important limitations with the accuracy of the identification, reflecting the complexity of the topic:

Persons in need of international protection

This population is usually identified through the questions, but several caveats apply:

- No differentiation at the sub-stock level: With the recommended questions, it is not possible to differentiate between the different sub populations that fall under the *persons in need of international protection* population (prospective and current asylum seekers, persons with determined protection status, and others in refugee-like situations).
- There are scenarios that can lead to possible exclusion errors. For example:
  - » If an individual does not view their move into the country as 'forced' but still has refugee status, they won't be correctly identified as a person with determined protection status. Two specific examples that can lead to this scenario include group recognition (a country declaring that all individuals originating from another country in a period of time should be classified as refugees) and refugees "in situ"– those individuals who relocate through choice before something happens that makes it impossible for them to return.
  - » Individuals who are granted refugee status, then leave the country and return (not forced) will still hold refugee status in that country but the most recent displacement was not forced so they will not be identified through the questions.
- It is also possible for inclusion errors to occur:
  - » An individual born in one country but a citizen of another, with no refugee background; if they subsequently were living in the country of their birth and forced to flee back to the country of their citizenship, the questions would identify them as "person with a refugee background", because the reason for their most recent international migration was "forced displacement".

Persons with refugee background

This population is partially identified through the questions, but not wholly, so it is not possible to identify the total for this group. Naturalized former refugees are identified, as are children born of refugee parents if they live in the same household as their refugee parent, and assuming the refugee parent is identified (see limitations above).

- Possible exclusion errors include:
  - » Children born of refugee parents, where they do not live in the same household as their refugee parent.
- Possible inclusion errors:
  - » It is possible that some of those identified as naturalized former refugees may have never achieved determined protection status if they acquired citizenship through another route, or if they viewed their migration as forced but entered the country without being identified as a person in need of protection.

Persons returned from abroad after seeking international protection

This population is not identified through these questions.

31. The mapping in Box A.2 illustrates how responses to each category of question will yield identification.

32. If it is important in the country context (and there is space on the census forms) to accurately distinguish between all the different categories and sub populations then the questions would need to be adapted and / or additional questions introduced. For expanded question sets, readers are referred to the respective questionnaire sections of Use Case B and C on surveys.

Box A.2

Census questions for the identification of refugees only (not IDPs)

IRRS Recommended topics and questions for Census to support identification of refugees and related categories: the following question topics identify whether respondents are migrants based on the *Principles and Recommendations for Population and Housing Censuses, Revision 3* and are recommended as core topics

i. Country of birth
ii. Country of citizenship
iii. Acquisition of citizenship
iv. Year or period of arrival in the country

In addition, it is recommended to include as a core census question:

v. Reason for (international) migration, with response categories:
   a. Employment (including military service)
   b. Education and training
   c. Marriage, family reunification or family formation
   d. Forced displacement (refugees, asylum seekers, temporary protected status, others)
   e. Other

The reason should refer to the main reason for the most recent move.

Mapping of these questions to populations of the statistical framework:

Question topic	Answer responses	
	Persons in need of international protection	Persons with a refugee background, excluding children born of refugee parents
i. Country of birth	Foreign country	Foreign country
ii. Country of citizenship	Foreign country	Country of census
iii. Acquisition of citizenship	Any	Any
iv. Year or period of arrival in the country	Any	Any
v. Reason for (international) migration	Answer response (d)	Answer response (d)

Identifying IDPs

33. The standard migration questions asked in most population censuses, which focus on most recent movement across administrative areas, may be a first step in identifying some IDPs. However, the last move of an IDP is not always the one related to their forced displacement, hence the standard migration questions have limited value for identifying all IDPs. They will typically not capture:

- Place of habitual residence;
- Those who have returned to their habitual place of residence within the period specified by the census;
- Those who within the period specified by the census have moved one or several times after their initial displacement; and
- Those who have moved within the same administrative area used in census geographical classifications.

34. To be sure to capture the entire IDP population the set of questions recommended in IRIS should also be asked. This is particularly the case if (a) IDPs have started to return to their places of habitual residence and it is expected that they still have key displacement-related vulnerabilities; (b) IDPs are expected to have settled elsewhere after their initial displacement, but still have key displacement-related vulnerabilities; (c) displacements are expected to have happened within the smallest administrative area proposed by the census; or (d) migration might have occurred for a combination of reasons including forced displacement.

35. The questions recommended in Box A.3 will identify IDPs and go some way towards distinguishing between sub-populations (IDPs in locations of displacement, return and other settlement locations) but this is not always possible with certainty.

IDPs

The recommended questions, alongside 'place of usual residence', will usually enable identification of IDPs and the three subpopulations (IDPs in locations of displacement, IDPs in locations of return, and IDPs in other settlement locations).

- Possible categorisation errors:
  - » For individuals who have been recently displaced, their "place of usual residence" is likely to be defined as the location they have been displaced from, rather than the location where the census has found them. In this case they are likely to be falsely identified as IDPs in location of return.
- Possible exclusion errors:
  - » If individuals move abroad for a period of less than 12 months and return to the same country, they should still count as IDPs – rather than returning refugees, for example – but the questions do not elicit the time spent abroad so these individuals will not be identified as IDPs based on the mapping provided.
  - » IDPs "in situ" may not be identified – those individuals who relocate through choice before something happens that makes it impossible for them to return – because their initial move was not forced.

- Possible inclusion errors:
  - The questions will only identify people who have entered the stock of IDPs, they won't identify those who exit the stock i.e. overcome their displacement related needs and should therefore move into the "IDP related populations" category.
- Other limitations:
  - If individuals have been subject to successive forced displacements only one can be captured in the census form.

IDP-related populations

The recommended questions will identify children born after displacement to at least one IDP parent, if they live in the same household as their IDP parent, and assuming the IDP parent is identified (see limitations above). Other non-displaced family members of IDPs and those who have overcome key displacement-related vulnerabilities are not identified.

- Exclusion errors:
  - The questions do not allow identification of children of at least one IDP parent where they do not live in the same household as their IDP parent.

36. It may be unrealistic to expect that further questions for identification of IDP-related persons could be added to a census questionnaire, but if this is important in the country context and there is space, the questions can be adapted or expanded, for example by drawing on the format recommended for surveys, which is found in Use Case C[16].

37. The mapping in Box A.3 illustrates how responses to each category of question will yield likely (but not certain) identification. The questions will also yield identification of refugees following the same mapping as established in Box A.2, noting that the questions on forced displacement are not limited to movements within country.

16 It should be noted that there may be other drivers for internal displacement that are not explicitly captured in the response options, such as domestic violence or harassment. This is the subject of ongoing academic discussion and debate. Future iterations of this manual will reflect any emerging consensus.

Box A.3

Census questions for the identification of IDPs

IRIS Recommended topics and questions for Census to support identification of IDPs and related categories: the following question topics identify whether respondents are geographic / internal migrants based on the *Principles and Recommendations for Population and Housing Censuses, Revision 3* and are recommended as core topics

i. place of usual residence
ii. place where present at time of census
iii. place of birth
iv. duration of residence
v. place of previous residence
vi. place of residence at a specified date in the past

(Note - where there have been large displacements in a country, it can be preferable to select a date that coincides with a significant displacement event, as this not only helps people to recall where they were on the date chosen, improving accuracy, but also helps to identify IDPs)

In addition, it is recommended to include the following set of questions as core census questions:

- Has (NAME) ever been forced or obliged to flee?
- (If yes, Reason)
- If yes, when was this?
- If yes, where did you move from? / where did you reside before you were forcibly displaced?
- If yes, where did you move to? (within country borders/abroad)

Box A.3 continued

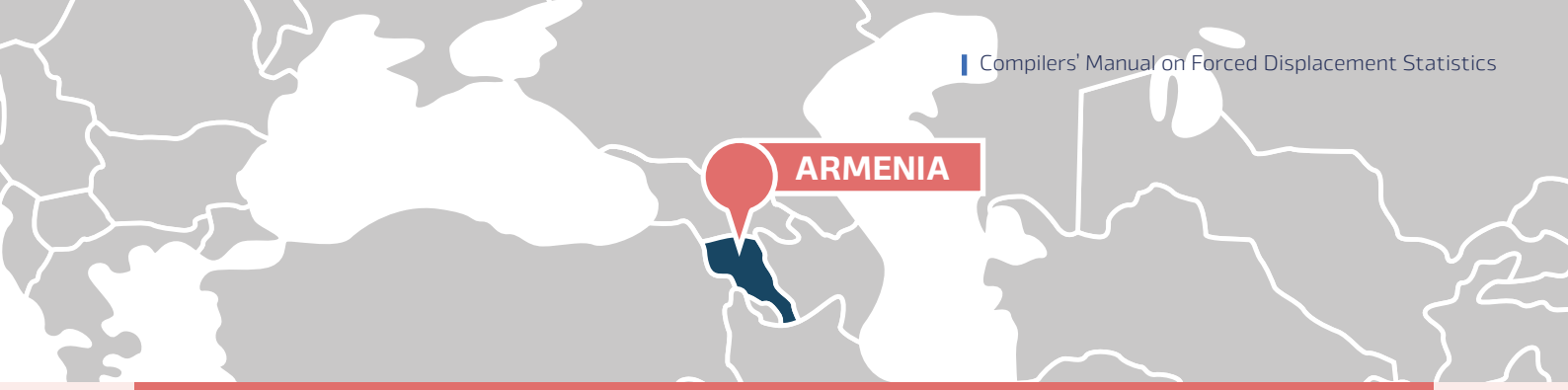
Mapping of these questions to populations of the statistical framework:

Filter: Those answering "yes" and that they moved within the country are identified as IDPs. The following table further categorizes such respondents into the three sub-categories.

Question topic	Answer responses		
	IDPs in locations of displacement	IDPs in locations of return	IDPs in other settlement locations
(i) Place of usual residence	Location	Location	Location
(ii) Place where present at time of census	Location matches (i)	Location matches (i)	Any location
(iii) Place of birth	Any	Any	Any
(iv) Duration of residence	Time period	Time period	Time period
(v) Place of previous residence	Any	Any	May match (xi)
(vii) Place of residence at a specified date in the past	Use this question to identify likely IDPs if there has been a significant displacement event.	Use this question to identify likely IDPs if there has been a significant displacement event.	Use this question to identify likely IDPs if there has been a significant displacement event.
(viii) Has (NAME) ever been forced or obliged to flee?	Yes	Yes	Yes
(ix) If yes, when was this?	Answer falls within time period in (iv)	Answer falls within time period in (iv)	Answer does not fall within time period in (iv)
(x) If yes, where did you move from? / where did you reside before you were forcibly displaced?	Location	Location matches (i)	Location, different from (i)
(xi) If yes, where did you move to? (within country borders/abroad)	Location matches (i)	Location within country.	Location within country, different from (i)

Other indicators for refugees and IDPs

38. The census will naturally collect a range of socioeconomic indicators that will be relevant to forcibly displaced people, but there are no recommended modifications to those questions with a view to adapting them to the forced displacement context.



CASE STUDY: INTEGRATING DATA ON DISPLACEMENT INTO THE REPUBLIC OF ARMENIA CENSUS

The Republic of Armenia undertook a census in October 2022, using a combined approach based on the population register and a 25% sample of addresses. The accurate identification of displaced communities was a priority for the census, in the context of recent conflicts, refugee influxes from neighbouring countries and challenges arising from the COVID-19 pandemic. The Statistical Committee (ARMSTAT) convened a roundtable to seek contributions from the relevant public institutions, migration service and non-government organisations. The final census form included a set of questions designed to identify place of birth, citizenship, refugee status and reason for relocation. The census was accompanied by extensive communications to reach out to all relevant communities.

Identification questions relevant to displaced populations, included in the Republic of Armenia census:

- Place of birth
  - » Country

in case of Armenia:

  - » Marz (Region), Community and Settlement

- Country of citizenship
  - » The second country of citizenship, for dual citizens
  - » The method of acquiring citizenship for the citizens of the Republic of Armenia
    a. By birth
    b. Granting citizenship
    c. Other
  - » For those who have not citizenship, indicate if he/she are:
    a. Recognized as a refugee, displaced from Azerbaijan
    b. Recognized as a refugee, displaced from another country
    c. Asylum seeker
    d. Undocumented stateless person
    e. Unknown

- Continuous residence of a person in the given place of residence since birth
  - » Yes/No

If "no"

  - » The year and the month, starting from which, he/she continuously resides in the given settlement
  - » Previous place of residence:
    - Country

In case of Armenia:

    - Marz (the region), Community, Settlement
  - » What is the type of settlement the person come from: *Urban/Rural*
  - » Main reason for changing the place of residence:
    a. Military operations in other countries
    b. A well-founded fear of persecution in other countries for race, national, religious affiliation, membership of any social group or political opinion
    c. Family circumstances
    d. Establishing Residence
    e. Education
    f. Employment
    g. Repatriation
    h. Temporary protection
    i. Climate change
    j. Other
  - » Has the person ever lived abroad for at least 6 months (except for "previous place of residence" above) and then he/she returned to Armenia? *Yes/No*

If "yes"

    a. The year and month of the last arrival in Armenia.
    b. What country did he/she come from?

Fieldwork planning

39. To ensure that refugees and / or IDPs are covered during a Census it is necessary to plan the fieldwork to overcome some specific challenges related to displaced people.

Importance of not treating refugee and IDP camps as institutional households

40. In many censuses there are several questionnaire forms, including one form for a private household and another for "institutional establishments". These institutions often comprise schools, hospitals, hotels and workers' camps and may also be used to cover camps for IDPs or refugees. Institutional questionnaires may not be suitable for camps for the displaced, particularly where these are more than short term arrangements. As the question content is often much shorter in the institutional questionnaire than in the household questionnaire, the information will be insufficient to provide for the requirements of users interested in displaced people. Wherever possible the camps should be enumerated using the questionnaires designed for private households.

Communications planning

41. Most censuses involve a detailed communication plan at a national and local level, to ensure that residents are aware the census is taking place and primed to take part at the correct time. It is important to ensure that displaced people are considered as part of this communication plan, if good coverage is to be achieved. There are no fixed prescriptions that will work everywhere, but experience from various countries should be observed and evaluated. The main messages to convey in census communications include:

    - That the Census will include displaced people, refugees and IDPs and that their responses are important.
    - How the data will be used (positive messages)
    - That confidentiality and data protection are assured
    - Any special arrangements that have been made to include displaced people (such as availability of interpreters).

42. To reach displaced people, communication at local and community level is often the most effective. It is important to ensure that communication is available in languages relevant to displaced populations, as refugees and IDPs may speak a different language than the local population. Examples of communication strategies include:

    - Leaflets or visits to relevant institutions (refugee camps, etc) in the lead up to Census, with interpreters available if language is likely to be a barrier.
    - Identification of and engagement with community leaders before the Census, to establish the most effective way of reaching displaced populations and to gain their buy-in and support.
    - Acknowledge the potentially sensitive nature of the question topics for displaced populations and explain the purpose of collecting the data and principles of data protection.

Including refugees and IDPs in pilot testing

43. All Censuses will include pilot testing, and as key hard-to-reach populations, refugees and / or IDPs should be included in this testing phase. As part of the pilot it is important to:

    - identify the challenges involved in reaching displaced populations and the best strategies to overcome those challenges;
    - test that the questions and question flow can be correctly interpreted, and language barriers overcome to enable correct identification of the populations (refugees and / or IDPs); and
    - review and mitigate any risk of causing non-response due to the sensitive nature of the topics.

Enumerator training

44. Successful coverage of displaced people in a census relies on effective enumeration. Displaced people are often harder to reach and can need support in order to take part in a census and this requires specific content in the enumerator training course.

45. Enumerators and their supervisors should be conversant with languages or dialects of the area in which they will be working, and in the case of displaced persons there should be provision for interviewers (if the census involves any face-to-face element) who are fluent in the languages and dialects from the displaced person's place of origin. One way of achieving this is to hire interviewers directly in the community speaking the dialect. If this is not taken account of then any field operations may be impacted by poor response rates, and inability to collect good quality data.

46. Of particular importance is training data collection staff on the definitions and concepts used in statistics about displacement, which may be unfamiliar to many experienced field staff. For example, refugee and IDP concepts such as habitual residence, usual residence, migration concepts, refugee categories, meaning of displacement, meaning of refugee and IDP, and reasons for displacement and migration.

47. Many displaced people are difficult to find, as their place of residence is often not fixed or is impermanent. They may be in shelters, camps, temporary accommodation or lodging with other households. Data collection staff must be carefully trained in recording all members of the household – including temporary members who may be displaced (in full alignment with the census recommendations on who should be enumerated in a household).

Fieldwork in practice

48. On the whole, provided fieldwork planning and enumerator training has taken full account of the need to cover displaced people, there are few issues specific to refugees or IDPs in terms of the actual fieldwork. However, it is worth noting that effective feedback from enumerators during the census can be crucial in identifying any emerging issues in reaching refugees and /or IDPs, which might be addressed by continued or improved communication and community engagement.

The post enumeration survey

49. Assessing the quality of population and housing census data is an integral part of the census operation. UNFPA advocates for countries to carry out some sort of quality assessment exercise. The Post Enumeration Survey (PES) of a Population Census is a sample survey that evaluates the results of the Census in terms of potential underestimation or overestimation of the size of the population as well as the characteristics of the population.

50. Any census Post Enumeration Survey should look at the coverage of displaced populations and the accuracy of data collected about them. To do this it is essential to carry out the PES in refugee camps, shelters, rented homes or any other institution or other places where refugees and IDPs may be staying. Special attention must be paid to the demographic characteristics and the refugee status of the persons concerned. This may need a special design component within the PES.

51. A specific subject that requires planning and attention is how to deal with any changes in the situation of the households, that could be as a result of forced displacement. This displacement may have occurred between the census and the PES and may impact on the household's composition. Specific questions have to be included in the PES that allow the detection of such changes, taking either the census or the PES as a base line. If the PES is taken as a base line, a code has to be included for each household member to indicate whether (s)he:

    i. Was also present in this household and dwelling at census time;

    ii. Was temporarily absent from the household and dwelling at census time;

    iii. Was born after the census; or

    iv. Resided in a different household and dwelling at the time of the census.

52. Note that iv) also includes the situation in which the entire household has moved from one physical dwelling to another. In cases ii) and iv), it may be wise to ask if the person was enumerated at the place where he or she spent census night. In addition, it must be verified if any residents at the census time have:

    i. Died since then; or

    ii. Become residents of a different household and dwelling since then.

53. Alongside PES, for refugees and asylum-seekers it should be possible to analyze coverage by comparing census figures with figures from the relevant refugee status determination authorities.

Data protection / confidentiality / disclosure

54. Data protection and confidentiality is of key importance in all census activities, but collecting data on displaced populations presents heightened sensitivities, as anyone who identifies themselves as such could face threats from doing so or be stigmatised. When collecting data, if the census involves any face-to-face element the displaced population has special needs or conditions of interview, and the 'no-harm' principles and UN ethical standards on data collection and data protection must be applied alongside relevant national laws and the standards set out in the UN Fundamental Principles of Official Statistics[17].

Analysis and dissemination

55. The process of analysing and presenting census data collected on displaced populations is fundamentally no different from that for any other data, but it is important that the data are included in publication plans. In countries where there are significant populations of refugees and / or IDPs, published outputs should address two key requirements:

- The key disaggregated census data sets (including the main results report and sub regional reports) should be published and include a standard disaggregation by refugee and / or IDPs, alongside the disaggregation by other variables.
- Refugees and / or IDPs should be included as one of the published thematic reports produced using census data, following the census.

17 UN Fundamental Principles of Official Statistics, UN Personal Data Protection and Privacy Principles.

USE CASE

B

INCLUDING REFUGEES IN A SAMPLE SURVEY OF THE NATIONAL POPULATION, OR RUNNING A STAND-ALONE SURVEY OF REFUGEES

ETHIOPIA.
2023 Socio-Economic Survey of Refugees in Ethiopia (SESRE) data collection.
©Ethiopian Statistical Service

EGRISS

Sample surveys – general background

56. Sample surveys are an important data source for measuring the characteristics and living conditions of refugees and refugee related populations. When planning a representative survey of refugees, NSOs have the option of including refugees in a broader survey of the national population (e.g. LFS, MICS, DHS, integrated living conditions survey, etc.), or to design a specialized survey for refugee populations. Both approaches have their merit: a specially designed refugee survey has the potential to collect broader and deeper information, whereas inclusion in a national multi-topic survey allows refugees to be compared with other groups in the general population. The two options are not mutually exclusive, and countries may decide to implement a complementary mix of both options. (A detailed discussion of pros and cons of the two approaches, and of sample surveys more broadly, is provided in IRRS paras 156-159 and 176-179).

57. The priorities when considering including refugees in a sample survey are:

- In the design of the questionnaire, include recommended questions to identify refugees.
- Identify which indicators of needs and integration are priority for data collection and analysis. If expanding an existing survey, establish whether any are already included in the data collection and add in question sets for any additional indicators.
- Establish a sampling frame that includes refugees. Where necessary expand the existing sampling frame or find an alternative sampling frame that includes refugees, or the locations where refugees reside.
- Undertake field visits to test sampling strategies, identify potential issues or concerns and to build relations between the parties involved in fieldwork implementation.
- Ensure that enumerator training allows sufficient time to cover refugees, including the relevant definitions and concepts, practical survey considerations and data protection.

58. Other than these special considerations during sample design and adjustments of the questionnaire and analysis, there are no further major technical implications from the statistical framework set out in IRRS for surveys of refugees. All the usual principles and best practices for running sample surveys for official statistics apply - including on fieldwork security risk assessment and mitigation - and are not discussed in this manual.

59. Many of the issues discussed in this Use Case are similar to the considerations when including IDPs in a sample survey (Use Case C) and so are repeated in Use Case C to ensure both Use Cases serve as stand-alone resources.

Questionnaire design

60. Designing an effective questionnaire is fundamental to the success of any survey. When considering including refugees in an existing or new survey the key factors are to ensure appropriate identification and classification questions are included and the question flow in the sections covering places of residence and migration status works accurately given the additional complexity. Typically, questions on refugee status should be introduced in the flow of questions on migration status.

61. Wherever possible, when including refugees, survey development timescales should allow additional time for questionnaire development, to provide an opportunity to bring together relevant stakeholders and seek their input to the survey design. Investing time to engage

stakeholders (for example, relevant government agencies, key communities, organisations supporting refugees, UNHCR) early in the project can help ensure the success of the survey by bringing together relevant knowledge and expertise and appropriately accounting for any sensitivities. It also helps raise awareness of the data that will be collected and promotes its use when the survey is complete.

Refugee identification and classificatory questions

62. The IRRS recommends questions for identifying refugees and refugee related populations in household surveys (IRRS paragraph 183) that are the same as those discussed in Use Case A for censuses:

    a. Country of birth

    b. Country of citizenship

    c. Acquisition of citizenship

    d. Year or period of arrival in the country

    e. Reason for migration, with response categories:

        - Employment (including military service)
        - Education and training
        - Marriage, family reunification or family formation
        - Forced displacement (refugees, asylum seekers, temporary protection status, others)
        - Other

63. In addition to these core identification questions, IRRS recommends the following classificatory variables in collecting and compiling refugee statistics. Survey practitioners will need to assess which of these are relevant and can realistically be obtained through a survey interview in a given context:

    a. Country of previous or last residence

    b. Date of first displacement/leaving previous country of habitual residence

    c. Parents' refugee statuses

    d. If an unaccompanied child (under the age of 18 years and separated from both parents or legal guardian)

    e. Legal residential/international protection status as applicable to the national context:

        - Persons in need of international protection
            - prospective asylum seeker
            - asylum seeker
            - individual refugee status
            - prima facie refugee status
            - subsidiary or complementary protection status
            - temporary protection status
            - refugee resettled in a third country
            - refugee-like (state form of protection granted)

- Persons with a refugee background
  - naturalised former refugee
  - child born of refugee parents without own refugee status
  - reunified family member
  - others with refugee background
- Persons returned from abroad after seeking international protection
  - repatriated asylum seeker
  - returning from other form of international protection
  - returning from other international protection

64. EGRISS has published a methodological paper, *Towards a standardized approach to identify IDPs, refugees and related populations in household surveys*[18] that expands on the IRRS recommendation and suggests survey question flows for use in household surveys. The paper discusses the questions that need to be included in order to address each element of identification of refugees and related populations; the limitations and potential sources of inclusion and exclusion error; and the practical considerations around administering the questions within a survey. It also provides question flows for inclusion in MICS and DHS.

Capturing indicators of needs and integration, and priority SDG indicators

65. Beyond basic socio-demographic variables (age, sex) and the identification and classificatory questions discussed in the previous section, IRRS also proposes a common set of indicators that can measure the satisfaction of refugees' needs and the progress towards their integration on a global basis. Most of these indicators can be obtained through surveys and should be considered for inclusion in a refugee survey as appropriate and feasible. It is also important to consider collecting GIS data where possible, to enable locational analysis.

66. In addition to the indicators of needs and integration recommended by IRRS, the IAEG-SDGs in collaboration with EGRISS has highlighted 12 SDG indicators that are of particular relevance to refugees. For survey practitioners, this implies that these indicators should be captured in any survey including refugees (whether this is a national survey that includes refugees or a refugee-specific survey) if applicable and feasible within the scope of the planned survey. The 12 indicators overlap with IRRS's indicators of needs and integration. Measurement of these SDG indicators for refugees should be aligned wherever possible with the national statistical practice for capturing these indicators among the wider national population, to ensure comparability.

67. Table B.1 outlines the IRRS indicators and the 12 SDG indicators (highlighted orange). EGRISS does not provide detailed definitions of the indicators but rather refers to the relevant international statistical standards for definitions and questionnaire specifications, such as the SDG metadata sheets[19]. If including refugees in an existing survey, it is likely that some of these indicators are already included and the scope for adding additional indicators may be limited. If launching a new survey of refugees, Table B.1 will help inform which indicators to measure.

---

18 EGRISS Methodological Paper Series (2023) Towards a standardized approach to identify IDPs, refugees and related populations in household surveys

19 See IRRS paras 320-445 for a detailed discussion of the indicators, other than SDG indicators.

68. It will not be possible to cover all the topics listed in a single survey and the actual content of any survey should be determined by local priorities and users' needs. The identification of an appropriate comparator population is necessary, and this is discussed further in paragraph 78.

Table B.1: Priority indicators of satisfaction of immediate and ongoing needs and of integration (excluding classificatory variables). The 12 SDG indicators highlighted by IAEG-SDGs as particularly relevant to refugees are highlighted (red).

Topic	Indicators	Relevant references
Demographic and migration		
Family and household situation	Marital status, Composition of the family, Information on absent family members.	
Civil		
Legal access to labour market		
Right to own property	SDG indicator 1.4.2: Proportion of total adult population with secure tenure rights to land, with legally recognized documentation and who perceive their rights to land as secure, by sex and by type of tenure. See following table.	SDG indicator metadata sheet
Eligibility to state benefits	SDG indicator 1.3.1 Proportion of population covered by social protection floors/systems, by sex, distinguishing children, unemployed persons, older persons, persons with disabilities, pregnant women, newborns, work-injury victims and the poor and the vulnerable	SDG indicator metadata sheet
Access to justice	SDG indicator 16.3.1 Proportion of victims of violence in the previous 12 months who reported their victimization to competent authorities or other officially recognized conflict resolution mechanisms	SDG indicator metadata sheet This can be a relatively small number compared to the overall sample of the surveyed individuals, and difficult to report if below a defined threshold of unweighted cases.
	16.9.1 Proportion of children under 5 years of age whose births have been registered with a civil authority, by age	SDG indicator metadata sheet

Education		
Educational attainment	SDG indicator 4.1.1: Proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex. See following table.	SDG indicator metadata sheet Collecting data for this indicator requires detailed learning assessments at the three levels specified in the indicator, which in most countries are administered through participation in various national, regional and/or international school-based assessments (as opposed to household-based surveys, which often will not sample sufficient numbers of children at a required age and education level). Inclusion of forcibly displaced children in these learning assessments is the preferred data source for this indicator, although if such an approach is likely to yield a small number of refugee children then an alternative approach should be considered. It will be important to ensure that reading proficiency tests are available in the language the child feels most comfortable reading. Countries wishing to include a learning assessment in a household survey covering forcibly displaced children should align the assessment method in the survey with the relevant school-based assessments that a country carries out. If not applicable, the MICS (Link, see "Foundational Learning: module") and EGRA&EGMA may also provide useful survey-based approaches relevant to the lower schooling levels.
Participation in education	SDG Indicator 4.1.2: Completion rate (primary education, lower secondary education, upper secondary education)	SDG indicator metadata sheet This indicator is often calculated by including education modules in household questionnaires.
	SDG Indicator 4.5.1: Parity indices (female/male, rural/urban, bottom/top wealth quintile and others such as disability status, indigenous peoples and conflict-affected, as data become available) for all education indicators on this list that can be disaggregated	SDG indicator metadata sheet
Participation in preschool education and access to childcare	SDG Indicator 4.2.2: Participation rate in organized learning (one year before the official primary entry age), by sex	SDG indicator metadata sheet
(Host country) language proficiency		

Economic		
Employment status	SDG indicator 8.5.2 Unemployment rate, by sex, age and persons with disabilities	SDG indicator metadata sheet. See ILO's model questionnaires (Link) for labour force surveys which include the required questions for capturing unemployment, as well as UNHCR's guidance to labour measurement in forced displacement contexts (Link) for a more specific discussion.
Informal sector employment	SDG indicator 8.3.1 Proportion of informal employment in total employment, by sex	SDG indicator metadata sheet. See ILO's model questionnaires (Link) for labour force surveys which include the required questions for capturing informality, as well as UNHCR's guidance to labour measurement in forced displacement contexts (Link) for a more specific discussion. Informal employment refers to people that in their main or secondary job were in one of the following categories: Own-account workers, employers, members of producers' cooperatives employed in their own informal sector enterprise Own-account workers engaged in goods production for their own final use Contributing family workers (no explicit written contract of employment, not subject to labour legislation or social security) Employees holding informal jobs

Economic - continued		
Income and consumption	SDG indicator 1.2.1 Proportion of population living below the national poverty line, by sex and age	SDG indicator metadata sheet. Poverty measurement is a complex undertaking. No internationally standardized questionnaire module for capturing consumption expenditure exists (since food and non-food items consumed vary from country to country). When comparing consumption aggregates against a national poverty line, it is crucial to capture consumption in a way that aligns with the primary poverty survey instrument used in the country from which the poverty line was derived. This paper documents best practices around survey-based consumption and poverty measurement. Compute a comprehensive estimate of total household consumption or income (with consumption being the preferred welfare indicator) and construct a correctly weighted aggregate of consumption or income per person. Compare this aggregate with the national poverty line. To measure consumption, cover questions on food and non-food expenditure as well as food consumed from own production (either through recall questions using lists of consumption items or through diaries) An alternative approach may be to use the Wealth Index constructed from data on asset ownership collected in DHS, or similar through MICS
Youth not in education/training/ employment	SDG indicator 8.6.1 Proportion of youth (aged 15-24 years) not in education, employment or training.	SDG indicator metadata sheet
Access to land for growing food	See "Right to own property" above.	

Social inclusion		
Poverty	See "income and consumption" above.	
Material deprivation. This is similar to poverty, but while poverty is generally measured using information on income and consumption, material deprivation is measured by the lack of access to basic necessities. These can include food, water, sanitation, and housing, as well as some durable items.	Indicator 1.2.2: Proportion of men, women and children of all ages living in poverty in all its dimensions according to national definitions	SDG indicator metadata sheet An alternative approach may be to use the Wealth Index constructed from data on asset ownership collected in DHS, or similar through MICS
Housing conditions	SDG indicator 11.1.1: Proportion of urban population living in slums, informal settlements or inadequate housing	SDG indicator metadata sheet
	SDG indicator 7.1.1 Proportion of the population with access to electricity	SDG indicator metadata sheet. A single survey question on primary source of lighting is sufficient to capture this indicator. See question HL2 in the broader set of energy-related questions recommended by WHO and the Word Bank (Link). Sources of electricity such as generators, candles, batteries, etc., are not considered due to their limited working capacities and their function as backup sources for lighting.
	SDG indicator 16.1.4 Proportion of the population that feel safe walking alone around the area they live	SDG indicator metadata sheet. Survey questionnaire specification provided directly in the "methodology" section of the metadata sheet. This is also available in the victimization module in MICS for both men and women 15-49 years of age.
Overcrowding	See "housing conditions" above.	

Health		
Self-reported health status		
Essential health services	SDG indicator 3.8.1: Coverage of essential health services	SDG indicator metadata sheet The indicator is an index reported on a unitless scale of 0 to 100, which is computed as the geometric mean of 14 tracer indicators of health service coverage. Household surveys collect about half of the indicators that go into this composite index. Some of the indictors listed are Level 2 indicators in IRRS, so they could be collected even if the composite indicator is not calculated MICS collects Use of tobacco (3.a.1), ANC coverage (3.8.1) need for family planning (3.8.1) DTP coverage (3,8.1) PNV coverage (3.b.1) measles (3.b.1) and care seeking for pneumonia. This means different indicators need to be collected with long questionnaires to estimate this composite one, and they include a variety of indicators from tobacco use, to vaccination for children up to 3 years of age and care seeking for pneumonia, which usually has very low denominator as only children that have had a disease episode within two weeks before the survey are taken into consideration.
	6.1.1 Proportion of population using safely managed drinking water services	SDG indicator metadata sheet. For survey questionnaire specification, see table 1 and "core questions" W1-W5 of the JMP recommendations (Link) The indicator also requires W6 from the JMP core questions, which includes E.coli testing for each household. This requires procurement of equipment and training[20].
	2.2.1 Prevalence of stunting among children under 5 years of age	SDG indicator metadata sheet. In terms of survey questionnaire design, only fields for age/DOB and height are required. However, it must be noted that height measurement is non-trivial. See guidance on anthropometric data collection from DHS (Link), MICS (Link) and the FANTA Project (Link).

20 Basic information on the E.coli testing process is available through the UNICEF website.

	3.1.2 Proportion of births attended by skilled health personnel	SDG indicator metadata sheet The relevant question in DHS and MICS reads: 'Who assisted with delivery of (NAME)?' and answer choices include both health personnel and other persons. See figure 1 of this article for a direct comparison and discussion. MICS enquires about the last birth during the past 2 years, whereas DHS enquires about all births during the past 5 years.

69. Examples of questionnaires from other refugee surveys are available in the UNHCR microdata library[21].

Sampling considerations when surveying refugees

70. In theory, sampling in forced displacement contexts is no different from sampling elsewhere, and the same general principles and methods apply[22]. However, in practice the selection from and application of these methods is often faced with specific challenges when it comes to refugees (IRRS paras 163-175).

71. Forcibly displaced persons may be hard to reach, either because they are difficult to identify, not willing to be identified, on the move or live in areas difficult to access. Moreover, displaced people often constitute a small proportion of the wider population and are not evenly distributed around a country or region – making it difficult or impossible to rely on sampling frames which don't allow for the direct identification of forcibly displaced households, or at least of small geographical areas where they are predominantly located. Refugees can live either in dedicated camps or they can live in houses among the general population, either as part of another household or as an independent household.

Identifying a sampling frame

72. The first question that guides the decision on sampling from refugee populations is how to construct a frame from which to sample refugees in a targeted manner. In the simplest case, this can either be a list of the wider population which allows identifying forcibly displaced, or a list solely describing the refugee population. Such a list can often come from administrative records, either from relevant government agencies or operational data from international agencies such as UNHCR. In order to gain access to such sources it is likely that a data sharing agreement will be required, which can take some time to establish and should set out how data privacy will be protected - for example, by basing sample selection on anonymised data, so that only the personal details of those invited to take part in the survey are shared.

73. If one does not have an existing list of possible respondents, an area-based frame needs to be created. The following methods are commonly used:

21 UNHCR Microdata Library

22 Useful reading on survey sampling in general can be found in UNSD's "Designing Household Survey Samples: Practical Guidelines", "Household Sample Surveys in Developing and Transition Countries" , and "Sampling Rare and Elusive Population".

CASE STUDY: ADAPTIVE CLUSTER SAMPLING IN UGANDA AND ETHIOPIA

In 2022, Fafo Institute for Labour and Social Research undertook a survey to understand economic development and integration of refugee communities in urban settings in Uganda and Ethiopia, commissioned by the World Bank. The aim of the survey was to obtain detailed information that would help explore labour market outcomes for refugees and host communities living side by side and engaging in a shared labour market settings. The surveys were carried out with support from the NSI, who provided the initial list of PSUs.

Due to the difficulty of surveying refugees in urban settings where they tend to be in smaller proportions compared to the host population, the survey used an adaptive cluster sampling approach to obtain a sufficiently large, representative sample of refugees. The procedure begins by constructing the list of households in the initial randomly selected PSUs. The listing process asked only for the name of the head of household (for identification purposes), the number of household members and nationality. Nationality proved to be an effective proxy for refugees. A threshold of 10% of non-national households in a PSU was set for the listing exercise to be extended to all nearest neighbour PSUs. Theoretically the listing could continue to all neighbours so long as the threshold criteria is fulfilled. In the Fafo survey, a single neighbour listing was conducted for cost reasons but resulted in sufficient number of refugee households.

a. If the locations in which refugees reside are known, and the National Statistics Office maintains a list of Primary Sampling Units (PSUs) for the country, the two can be used to identify those PSUs where refugees reside. The first stage of sampling can then include either only those PSUs, or over-sampling from those PSUs. Relevant PSUs are selected randomly or systematically, and then a household listing is carried out or updated in the selected PSUs, capturing household's refugee status[23].

23 Thematic report: Data disaggregation of SDG indicators by forced displacement, by Carlos Barahona-Zamora with the support of Emily Nevitt, Alex Thomson, and Ciara McHugh at Stats4SD, and Andrea Pellandra at UNHCR November 2020, Data disaggregation of SDG indicators by forced displacement.

CASE STUDY: SAMPLING IN A CHALLENGING SECURITY SITUATION IN SOMALIA

In a survey in Somalia, it was not feasible to conduct a full listing of all households in an enumeration area, as this was too time-intensive and may have raised suspicion. Instead, a micro-listing approach was used, which required enumeration areas to be segmented into smaller enumeration blocks using satellite imagery. Enumeration blocks were small enough for enumerators to list and select households immediately before conducting the interview.

Source: https://openknowledge.worldbank.org/bitstream/handle/10986/31267/WPS8735.pdf

b. A similar approach can be applied if the locations of refugees are unknown, by drawing a random selection of PSUs and undertaking a listing exercise, using a very short set of questions, for example with nationality as a proxy for refugee status. In those PSUs with higher concentrations of refugees – above a pre-determined threshold such as 10% - the household listing is extended to the neighbouring PSUs (or one nearest neighbour PSU). This process continues extending until PSUs no longer cross the threshold. In this way, PSUs with the highest concentration of refugees are identified and the listing can be used as the sampling frame. The effectiveness of this adaptive cluster sampling approach[24] relies on the settlement distribution of refugee households in urban settings. When refugees are settled in clusters with enough concentration, adaptive cluster sampling is a useful method to efficiently and effectively sample refugee populations.

c. If a list of PSUs of reasonable quality does not exist, alternative ways to identify smaller geographical areas to sample from exist. Satellite, aerial or drone photos of areas in which refugees are known to reside can be used to divide geographical areas into sampling units to be sampled and the households existing there listed to be sampled in a second stage.

d. If refugees' locations are (partially) unknown, Mobile Phone Tracing may be an option. Digital trace data generated by refugees' mobile phones can help identify the areas in which refugees are located and thus the geographical areas to list/sample from. This is not the same as identifying refugees' phones and then sampling them for a phone survey (see section (e) for comparison). Approaches such as these naturally require careful consideration around data privacy. Further discussion of non-traditional data sources is included in Use Case F.

---

24 Thompson, Steven K. "Adaptive Cluster Sampling." Journal of the American Statistical Association 85, no. 412 (1990): 1050–59. Adaptive Cluster Sampling.

CASE STUDY: SURVEY OF NEWLY ARRIVED VENEZUELANS IN PERU

National authorities in Peru, supported by the UN inter-agency coordination platform, initiated ENPOVE - a household survey on the living conditions and experiences of newly arrived Venezuelan migrants and refugees in 2018. To overcome the challenge of not having a dedicated sampling frame for the survey, the Peruvian National Statistical office identified areas with high concentrations of Venezuelans based on the findings of the national population and housing census conducted in late 2017. This helped narrow down the geographical scope of the survey to the national capital and five provinces that were home to an estimated 85% of the Venezuelan population. Those areas were then oversampled in a conventional two-stage design.

In 2019, with a view to expanding the sampling options of the ENPOVE in light of a high and spacially dispersed influx of Venezuelans since the 2017 Census, a Mobile Phone Tracing exercise was carried out to establish alternative ways of identifying areas with high concentration of Venezuelans. Based on 15 million anonymized call detail records obtained from a major telecommunication company, mobile phones were tagged as likely to belong to a Venezuelan, if they were (a) registered under the name of a Venezuelan national, or (b) were used at least 30 times over the past month to make/receive a voice call or text message to/from Venezuela; or accessed a website of interest to Venezuelans. Using georeferencing, these phones were then associated with census PSUs – providing an additional indication of Venezuelans present in these PSUs and thereby complementing the census data from 2017.

Source: https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1666/libro.pdf and https://openknowledge.worldbank.org/bitstream/handle/10986/34175/Big-Data-for-Sampling-Design-The-Venezuelan-Migration-Crisis-in-Ecuador.pdf

e. Moving away from area-based sampling approaches, it may be possible to undertake a phone survey, assembling a list frame of phone numbers of the refugee population, by Mobile Phone Tracing or via Random Digit Dialling[25], or to use phone numbers collected through previous surveys. If launching a face-to-face refugee survey, it may also be worth collecting phone numbers and consent for future phone-based panel surveys at the same time.

25 The usual risks of sample undercoverage resulting from phone ownership and network connectivity rates among the target population need to be taken into account.

CASE STUDY: CREATING A SAMPLE BASED ON CALLING RANDOM MOBILE PHONE NUMBERS IN AFGHANISTAN

Over two million Afghan refugees had returned to their homes between 2014 and 2017 as a result of the gradually stabilizing situation in their country. Limited information was available on the conditions of those returnees and the hazardous security situation in many parts of the country made it difficult to send enumerators to the field to collect data for a household survey. This prompted the World Bank to explore the feasibility of conducting remote interviews through phone surveys given the rising trend in mobile phone ownership rates within the Afghan population. The sampling frame was constructed through a Random Digit Dialling (RDD) process to ensure that all mobile phone owning households had an equal probability of inclusion in the sample. Once a household was contacted, a short automated Interactive Voice Response (IVR) program would screen the respondent for basic demographic information and displacement status to confirm their eligibility for the survey. Some 3,254 returnees were identified through the process who were then contacted by a live enumerator to complete the main survey questionnaire. Validation of the sample design and weight calibration was done by comparing sample composition and characteristics to data from a census of Afghan refugees in Pakistan and records maintained by humanitarian agencies involved in facilitating the return of Afghans from Pakistan.

Source: http://documents1.worldbank.org/curated/en/298881533562809348/pdf/Afghanistan-World-Bank-Phone-Survey-of-Afghan-Returnees-Methodology-and-Representativeness.pdf

74. Sometimes a combination of different approaches is required. For example, if expanding a national survey to include refugees living in camps and outside camps, it is possible that PSUs will be available but that they exclude the camps. In this case, it may be necessary to use registration lists of refugees in the camps as a complementary sampling frame.

Drawing a sample

75. Constructing a sampling frame of refugees allows a survey to apply probability-based sampling methods. This ensures that every unit in the population has a known, non-zero chance of being selected into the sample, and its probability of selection can be accurately determined. This makes it possible to produce unbiased estimates of population totals, by weighting sampled units according to their probability of selection. It also allows for estimates of uncertainty, like confidence intervals. Probability-based sampling is considered best practice when sampling for surveys and ought to be the first choice - whenever it is possible to construct a sampling frame, i.e. a list of units to draw a sample from.

76. Including refugees in a general household survey is likely to require a boost to the sample ratio for refugees to ensure that a sufficient number are reached through the survey – because refugees tend to be a small proportion of the overall population. The size of a boost will depend on the analysis requirements – for example the disaggregations and level of precision that are expected.

CASE STUDY: BOOSTING A SAMPLE SIZE IN UGANDA'S DEMOGRAPHIC AND HEALTH SURVEY

A boost was applied during the 2018 Uganda DHS survey to ensure that sufficient data on refugees would be collected. A boosted sample for host communities was also needed, to ensure the sample was still representative at the regional level (the regular DHS sample is not representative at the district level but only at the regional level). This was done in a way that was consistent with the selection protocols of the primary sample while enabling the production of robust estimates at the level of the host community. The envisaged sample size for refugees and the host community booster sample for the Northern stratum is shown below:

	Northern stratum		
	# of sample households in regular DHS	# of sample households added via boost	Total # of households in domain
Refugees	0	1,200	1,200
Host communities	690	510	1,200

77. If sampling from highly concentrated population such as refugees living in camps, it is generally not necessary to consider clustering the sample. Clustering is a technique that one uses to bring down sample dispersion and make travel during fieldwork manageable, but it comes at the cost of losing precision and driving up design effects. When sampling from a highly concentrated population there is no real cost saving from clustering but there is still a loss of precision, so it is preferable to obtain a registration list for the camp (if it exists and is sufficiently up to date) and use it as a complementary sampling frame from which to draw a systematic and unclustered selection of dwellings.

78. If carrying out a dedicated survey of refugees, it is also important to identify an appropriate "host community" to include in the sample as a comparator population. There is no standard definition of host community, so at present this will need to be locally determined. Examples include the wider population that lives within a certain radius of a refugee camp, or households within the same administrative area.

79. It should also be noted that responses to questions on attitudes, intentions to move, and decision-making power in the household may differ substantially by gender. When feasible and appropriate, using a random selection protocol of female and male respondents within a sampled household, or interviewing more than one member of each household, are ways to ensure more reliable data. The decision on eligible household members to be included in the survey will also need to reflect which variables are to be collected, since the eligible respondents vary across different SDG indicators, for example.

Non-probability sampling methods

80. When constructing a sampling frame is not feasible, reverting to non-probability-based sampling is sometimes considered. Commonly used non-probability sampling techniques are respondent driven sampling (RDS) or snowball sampling. If refugees are difficult to find, either because they are few in numbers, are hiding, are homeless, on the move or live in an area difficult to access, such methods may be considered as an option of last resort. They should however be used with caution, as they do not allow for statistically representative, unbiased inference from the sample to the population. Application of non-probability sampling methods in official statistics production is uncommon.

Limitations of registration systems as sampling frames

81. If a potential sampling frame exists from registration records, a frequent challenge is that many registration systems group individuals into "cases" or "registration groups", which are different from a household as defined in standard survey methodology (which is built around the concepts of shared dwelling and housekeeping arrangements). A detailed discussion of the issue as well as possible solutions is provided in the UNHCR guidance note titled "Resolving the 'case versus household' issue when conducting a Socioeconomic Assessment"[26].

82. In a dedicated survey of refugees or IDPs that includes those both inside and outside camps, consideration needs to be given to the relative accuracy of sampling frame sources. It is possible that registration systems for those living in camps will be more accurate and complete than for those living outside camps, and respondents will be easier to reach inside camps than outside – introducing a risk of non-response bias. Conversely, another frequent challenge with sampling frames derived from refugee registration systems is that address/contact information can be outdated.

Box B.1

Guidance material on sampling in forced displacement contexts

Two dedicated resources on sampling in forced displacement contexts have recently been published and may be considered for further details and examples:

- Stephanie Eckman and Kristen Himelein (2022): Innovative Sample Designs for Studies of Refugees and Internally Displaced Persons
- Joint IDP Profiling Service (JIPS, 2020): Sampling Guide for Displacement Situations and Practical Examples

26 UNHCR proGres Database: Handling differences between cases and households

Fieldwork planning

83. Before any survey fieldwork commences, it is advisable to conduct one or more field visits to fully understand and plan for the locations and practicalities involved in reaching refugees. This can be important in overcoming the additional challenges that surveying refugees can entail. Ideally one or more of the most challenging locations / scenarios should be visited, with the following objectives:

- Observe the situation on the ground;
- Test whether the sampled households can in practice be located and accessed and make contingency plans if not, for example through engaging local community leaders, local authorities or landlords for access;

CASE STUDY: FIELD VISITS FOR THE SOCIO-ECONOMIC SURVEY OF REFUGEES IN ETHIOPIA

In 2022 Ethiopia launched a national refugee survey—the Socio-Economic Survey of Refugees in Ethiopia (SESRE) which was funded by the Joint Data Center—using the same methodology and content as its national household poverty survey, the Household Welfare Statistics Survey (HoWStat). The objective of SESRE was to sample refugees in most refugee camps across Ethiopia and implement it, for the first time, through the Ethiopian Statistics Service (ESS). In order to prepare for the survey, a team from ESS undertook field visits to 3 camps with the following objectives:

a. Better understand the situation in the camps with respect to how camps are set up;
b. Better understand the challenges around identifying sampled households in the camps;
c. Discuss the objectives of SESRE with camp management and refugee community leaders;
d. Engage refugee community leaders to start information campaigns on SESRE in the camps; and
e. Link ESS field teams with the field office teams of the Refugees and Returnees Service who is responsible for managing refugee camps in Ethiopia.

UNHCR and RRS supplied the sampling frame based on the most up-to-date registration lists, but thanks to the field visits, it quickly became apparent that the registration lists were outdated and did not contain sufficient information to locate respondents. Based on information from the field visits, a new approach to locate respondents was defined which relied on the knowledge and contacts of refugee community leaders inside the camps to locate respondents rather than location information from registration lists. Moreover, relationships between different government entities (ESS and RRS) at the survey locations could be established early to ensure full collaboration across institutions. Refugee community leaders received information and an understanding of SESRE before implementation was started which was helpful in receiving information which improved survey implementation. The knowledge from field visits was further built into the training of enumerators and ensured a more efficient and effective way to successfully roll out the survey.

- Build relations between the different agencies and communities involved in implementing the survey – trust and rapport can be essential to success of the survey;
- Identify any language barriers and how best to address them (questionnaire translations, availability of interpreters).
- Test the questionnaire, both in terms of content and flow. Do the questions lead to correct identification and classification of refugees; are there any issues with translations; are there any sections of the questionnaire that are seen as more sensitive and potentially might result in non-response.

84. It is advisable to allow additional budget for overcoming language barriers, including translation of questionnaires; hiring enumerators with relevant languages; and increased availability of interpreters. There may also be additional costs for extended duration of fieldwork compared with a standard survey, reflecting the challenges of identifying refugees, and potentially greater use of supervisors to train, monitor and support enumerators.

85. Surveys may involve a communication plan at a national or local level, to ensure that residents are aware the survey is taking place and primed to take part at the correct time. It is important to ensure that displaced people are considered as part of the communication plan, if good coverage is to be achieved. There are no fixed prescriptions that will work everywhere, but experience from various countries should be observed and evaluated. The main messages to convey include:

- That the survey will include displaced people and refugees and that their responses are important.
- How the data will be used (positive messages) and clarity that it is not linked to allocating assistance.
- That confidentiality and data security are assured
- Any special arrangements that have been made to include displaced people (such as availability of interpreters).
- What opportunities there will be for respondents to hear about the survey results following data collection. Depending on the scale of the survey, it may be appropriate to offer some dissemination aimed specifically at the target population.

The communication plan may also consider raising awareness of any existing referral mechanisms available to the community, for feedback or protection concerns, according to the principles of "do no harm".

86. To reach displaced people, communication at local and community level is often the most effective. Examples of communication strategies include:

- Leaflets or visits to relevant institutions (refugee camps, etc) in the lead up to Census, with interpreters available if language is likely to be a barrier.
- Identification of and engagement with community leaders before the survey, to establish the most effective way of reaching displaced populations and to gain their buy-in and support.
- Acknowledge the potentially sensitive nature of the question topics for displaced populations and explain the purpose of collecting the data and principles of data protection.

87. If it is possible to include refugees within the survey workforce, for example as enumerators, there can be many benefits, including greater success reaching and effectively communicating with the refugee community.

Enumerator training

88. Displaced people are often harder to reach and can need support in order to take part in a survey and this requires specific content in the enumerator training course. Allowing sufficient time to address these issues in the enumerator training is key to delivering a successful survey. It may take between one and two additional days to cover the appropriate material.

89. Topics to cover in enumerator training include:

- The definitions and concepts used in statistics about displacement, which may be unfamiliar to many experienced field staff. For example, refugee concepts such as habitual residence, usual residence, migration concepts, refugee categories, meaning of displacement, meaning of refugee and IDP, and reasons for displacement and migration.
- Accurate recording of all members of the household, including temporary members who may be displaced. This is important as displaced people may fail the interviewers standard understanding of 'usual members of the household', leading to them being inaccurately treated as visitors and excluded from household rosters.
- The sensitivities around interviewing refugees and the data protection implications. When conducting surveys, the displaced population has special needs or conditions of interview, and the 'no-harm' principles and UN ethical standards on data collection and data protection must be applied. Enumerators must be aware of the existing referral mechanisms for community feedback or protection concerns, should any issues present themselves during the course of the data collection.
- Building trust and rapport, and skills in interviewing vulnerable people, such as children and disabled people. People who have been displaced have suffered trauma and any interview may ask people to relive painful experiences.
- Strategies for managing personal well-being for the enumerators, if they are likely to be exposed to traumatic testimonies during the data collection – which can become emotionally exhausting.

ETHIOPIA.
2023 Socio-Economic Survey of Refugees in Ethiopia (SESRE) data collection.
©Ethiopian Statistical Service (ESS)

CASE STUDY: TRAINING ENUMERATORS FOR THE SOCIO-ECONOMIC SURVEY OF REFUGEES IN ETHIOPIA

The 2022 Socio-Economic Survey of Refugees in Ethiopia (SESRE) was implemented, for the first time, by the Ethiopia Statistics Service (ESS). Though ESS has much experience carrying out national household surveys, it was the first time, they surveyed refugee populations. To ensure sensitivities and special protection needs are taken into consideration, the two week training of enumerators included one day specifically focused on the intricacies of collecting data from refugee populations. This content was taught by representatives from the government's Refugees and Returnees Service (RRS) and UNHCR, and aimed at providing:

a. Definitions and concepts that are specific to refugees;
b. Overview of the situation of refugees in the country as well as how camps are organized and managed;
c. Practicalities of enumerating refugees in large camps;
d. Advice on locating respondents;
e. Advice on effectively communicating the purpose of the survey to avoid creating expectations by respondents;
f. Information on special protection needs of respondents.

The feedback received from enumerators and ESS staff showed that the time invested in these specific aspects of the training was instrumental in providing enumerators with the right context and sensitivities to successfully implement the survey.

Fieldwork in practice

90. On the whole, provided fieldwork planning and enumerator training has taken full account of the need to cover displaced people, there are few issues specific to refugees in terms of the actual fieldwork. However, it is worth noting that effective feedback from enumerators during any survey can be crucial in identifying any emerging issues in reaching refugees, which might be addressed by continued or improved communication and community engagement.

91. If budget allows, it is advisable to undertake more monitoring of fieldwork than for a standard survey, reflecting the challenges associated with identifying refugees and potential issues overcoming sensitivities around data collection. Increased availability of supervisors can help to monitor the quality and consistency of data collection and play a facilitating role, communicating the purpose of the survey with more authority, to gain trust and avoid non-response.

Data quality

92. There are few data quality issues that are specific to surveys including refugees, over and above other surveys in challenging locations, and so data quality assurance should follow standard processes. The only additional data quality risk is around the accuracy of survey responses if respondents consider that their responses will influence their refugee status or outcomes in some way. This should be mitigated through accurate communication and good enumerator training.

93. Often it is possible to triangulate survey data with data from other sources, such as administrative data from other government agencies or operational data produced during the regular activities of humanitarian organisations, to make sure that high level trends match between the various available datasets without any major deviation. For example, if the variation between two stocks is not significant, but administrative data identifies a high number of smaller, shorter term movements (e.g. in the case of preventative evacuations leading to quick returns, or mass movements leading to cross border returns) then this should help to verify or correct official statistics on flows or alert the need for better or more frequent data collection. Similarly, through comparing with cross border data, if movements lead to significant cross-border flight or returns, the datasets on both sides of the border should correspond with each other to ensure validity. Further discussion of administrative data, operational data and non-traditional data sources is included in Use Cases D, E and F respectively.

Data protection / confidentiality / disclosure

94. Due consideration should be given to privacy in accordance with national and international laws. The UN Fundamental Principles of Official Statistics states that "individual data collected by statistical agencies for statistical compilation ... are to be strictly confidential and used exclusively for statistical purposes"[27]. Disclosure should also include considerations of interoperability: how datasets on different population types talk to each other, their ability to share information on those populations, or even on specific individuals, their characteristics and their needs.

27 United Nations Principles and Recommendations for Population and Housing Censuses Revision 3.

Analysis and dissemination

95. The process of analysing and presenting the survey data collected on refugees is no different from that for any other data and should be planned to meet the needs of users. This section briefly describes the outputs that are commonly produced to meet users' needs, with references for further information if required.

96. Data collected on identification and classification of refugees is used to produce analysis of stocks and flows for the different categories of refugees and refugee related populations, presented as counts and proportions cross-tabulated by sex and age group. Details are in IRRS paragraphs 114 – 125.

97. Analysis of stocks is recommended for all three categories of refugee and refugee related populations.

- **a.** Counts of each type of refugee and refugee related populations, by sex and age, or age group. If age group is used, then it is important to distinguish children from adults aged 18 years and over, and for some purposes it may also be helpful to identify those under 16 years of age.
- **b.** Proportions of each type of refugee and proportions who are female. IRRS elaborates on these, drawing out a recommended set of indicators – see Box B.2.

98. Flows are inherently more difficult to collect through surveys and in practice the recommended indicators are likely to need to come from administrative data. Where it is possible to collect data, flows are recommended only for those in need of international protection and those who have returned to their country of habitual residence after having sought international protection abroad, not for those with a refugee background.

- **a.** Counts of inflows during a specified reference period, for each sub-category of refugees, for those in need of international protection or who have returned to their country of habitual residence after having sought international protection abroad, by sex and age.
- **b.** Proportions - IRRS elaborates on these, drawing out a recommended set of indicators – see Box B.2.

SUDAN.
Action planning workshop in Tawila Locality held in May 2021.
©JIPS

Box B.2

Refugee stock and flow indicators (from IRRS, 2018)

i. Stock of refugees in need of international protection (IRRS, 2018 paragraph 115):
   a. Percentage of the total population who are in a country for international protection.
   b. Proportion females among persons in a country for international protection.
   c. Percentage of persons present in the country for international protection who are asylum seekers.
   d. Percentage of persons present in the country for international protection who have determined status.
   e. Total number of unaccompanied and separated children under 18 in a country for international protection, by sex and age.
   f. Proportion of persons in a country for international protection who have remained for up to 5 years.
   g. Proportion of persons in a country for international protection who have remained for up to 10 years.
   h. Proportion of persons in a country for international protection who have remained for over 10 years.

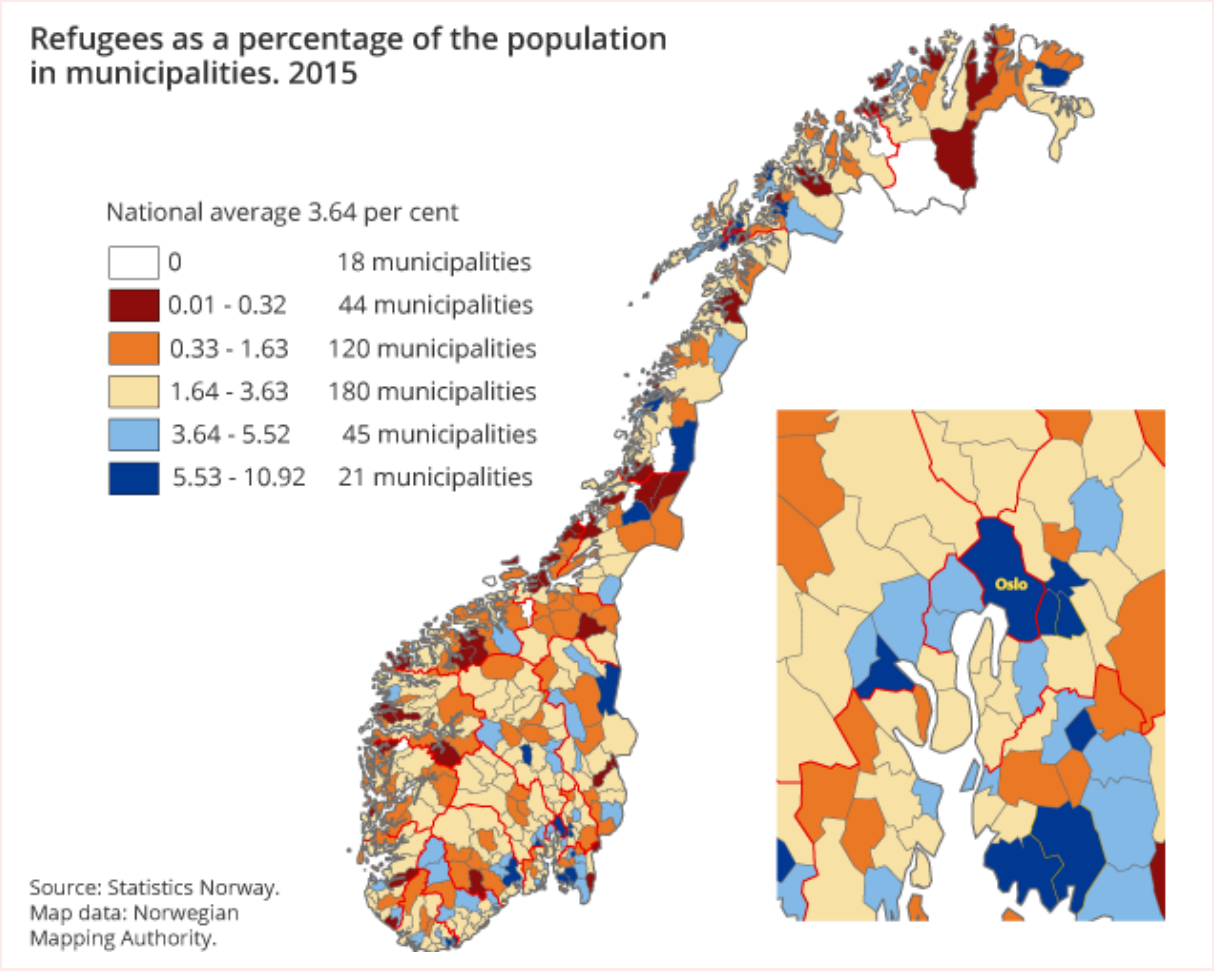
ii. Stocks of persons with a refugee background (IRRS, 2018 paragraph 120):
   a. Percentage of the total population who have a refugee background.

iii. Stocks of persons returned after having sought international protection abroad (IRRS, 2018 paragraph 123):
   a. Percentage of persons returned from having sought international protection abroad among all returned citizens.
   b. Proportion females among persons returned from having sought international protection abroad.

iv. Flows of refugees in need of international protection (IRRS, 2018 paragraph 119):
   a. Total number of asylum seekers who received a decision during a period of time, by sex and age.
   b. Percentage of asylum applications decisions that are positive (or negative) during a period of time, by sex and age.
   c. Percentage of determined refugee status granted during a period of time, by type (Prima facie, individual, derivative, complementary/subsidiary, temporary protection), by sex and age.
   d. Percentage of rejected asylum seekers who left the country during a period of time, by sex and age.
   e. Recognition rate during a period of time [the added value of the use of longitudinal information in more adequate calculation of the recognition rate should be acknowledged].
   f. Percentage of persons entered for international protection during a period of time who were resettled elsewhere.

CASE STUDY: STATISTICS NORWAY'S GEOSPATIAL ANALYSIS

Persons with refugee background as a percent of municipal population, 1 January 2015.



The best examples of data dissemination are often through online resources, which enable dynamic access to the data and can tell the data story in an engaging and interactive way. This Case Study captures one image from an online resource, to illustrate and inspire. *The full details are available in the following source material: Refugees in Norway* https://www.ssb.no/en/befolkning/artikler-og-publikasjoner/refugees-in-norway

99. Geospatial analysis is also recommended, to visualise the stocks and flows.

100. There are no specific recommendations relevant to refugees on how other data on socioeconomic indicators for measuring refugees' needs and integration (including SDGs) should be analysed and presented. Typical tabulations could include disaggregation by category of refugee, sex and age group. More complex analysis and commentary might address the integration of displaced people into host communities, the progress made towards overcoming key displacement-related vulnerabilities, and other key policy issues.

101. In planning how the survey analysis will be disseminated, if part of a wider survey it is preferable for refugee analysis to be disseminated in the main survey report as a standard disaggregation but it may also be impactful to produce a separate report focussing just on refugees. There is no right answer here, it will depend how users' needs can best be met in the specific national context. It is also important to consider how survey findings can be shared with the refugee communities themselves. This feedback loop is important: as well as meeting the needs of a key user community it demonstrates what the survey is for and promotes the importance of the survey, encouraging future engagement and participation.

MEXICO.
August 2022 presentation by the National Institute of Statistics and Geography (INEGI) of the quantitative results of the Chihuahua profiling survey.
©Antonio Mendoza

USE CASE

C

INCLUDING IDPS IN A SAMPLE SURVEY OF THE NATIONAL POPULATION, OR RUNNING A STAND-ALONE SURVEY OF IDPS

CENTRAL AFRICAN REPUBLIC.
2021 Harmonized Household Living Conditions Survey.
©JDC/Central African Institute of Statistics, Economic and Social Studies (ICASEES)

EGRISS

Sample surveys – general background

102. Sample surveys are an important data source for measuring the characteristics and living conditions of IDPs and IDP related populations. When planning a representative survey of IDPs, NSOs have the option of including IDPs in a broader survey of the national population (e.g. LFS, MICS, DHS, integrated living conditions survey, etc.), or to design a specialized survey for IDP populations. Both approaches have their merit: a specially designed IDP survey has the potential to collect broader and deeper information, whereas inclusion in a national multi-topic survey allows IDPs to be compared with other groups in the general population. The two options of course are not mutually exclusive, and countries may decide to implement a complementary mix of both options. (A detailed discussion of pros and cons of the two approaches, and of sample survey more broadly, is provided in IRIS paras 220-225 and 238-244).

103. The priorities when considering including IDPs in a sample survey are:

- In the design of the questionnaire, include recommended questions to identify IDPs.
- Identify which indicators of progress towards durable solutions are priority for data collection and analysis. If expanding an existing survey, establish whether any are already included in the data collection and add in question sets for any additional indicators.
- Establish a sampling frame. If possible, expand an existing sampling frame or develop an alternative sampling frame that includes IDPs, or the locations where IDPs reside, including camps and temporary settlements.
- Undertake field visits to test sampling strategies, identify potential issues or concerns and to build relations between the parties involved in fieldwork implementation.
- Ensure that enumerator training allows sufficient time to cover IDPs, including the relevant definitions and concepts, practical survey considerations and data protection.

104. Other than these special considerations during sample design and adjustments of the questionnaire and analysis, there are no further major technical implications from the statistical framework set out in IRIS for surveys of IDPs. All the usual principles and best practices for running sample surveys for official statistics apply - including on fieldwork security risk assessment and mitigation - and are not discussed here.

105. Many of the issues discussed in this Use Case are similar to the considerations when including refugees in a sample survey (Use Case B) and so are repeated here to ensure both Use Cases serve as stand-alone resources.

Questionnaire design

106. Designing an effective questionnaire is fundamental to the success of any survey. When considering including IDPs in an existing or new survey the key factors are to ensure appropriate identification and classification questions are included and the question flow in the sections covering places of residence and migration status works accurately given the additional complexity. Typically, questions on IDP status should be introduced in the flow of questions on migration status.

107. Wherever possible, when including IDPs, survey development timescales should allow additional time for questionnaire development, to provide an opportunity to bring together relevant stakeholders and seek their input to the survey design. Investing time to engage stakeholders (for example, relevant government agencies, key communities, organisations supporting IDPs, UNHCR) early in the project can help ensure the success of the survey by bringing together relevant knowledge and appropriately accounting for any sensitivities.

CASE STUDY: COORDINATING WITH STAKEHOLDERS ON QUESTIONNAIRE DESIGN IN SUDAN

In 2020, a large-scale analysis covering eight localities across Darfur's five states was initiated to provide a shared evidence-base to support peacebuilding and durable solutions under the UN Peacebuilding Fund (PBF). JIPS was requested to provide technical support. The project targeted both displaced and non-displaced population groups to measure the progress towards durable solutions among IDPs.

The design of the household survey tool was a consultative process that involved all stakeholders in the project. This meant presenting and sharing the progress at each step and collecting input, questions and comments. It ensured that the final questionnaire had wide buy-in and support and could underpin a successful survey. JIPS led the design of methodology and tools and conducted the analysis and reporting. The process was guided by UNHCR and the other PBF agencies (UNDP, UNICEF, IOM, UN-Habitat and FAO), with technical guidance from the Durable Solutions Working Group in Sudan (DSWG). IOM collected the survey data and the Sudanese Development Initiative (SUDIA) undertook the qualitative area-level data collection.

The large-scale sample-based survey was combined with extensive in-depth qualitative data and community consultation sessions. These were led by the Sudanese Development Initiative (SUDIA), a Darfur-based non-governmental organization, with support from UNHCR and funding from UNDP. SUDIA's well-established expertise in conflict and displacement dynamics in the region, paired with a solid network of local authorities and stakeholders, were key in this context. Piloted in Tawila, the approach was then rolled out to the remaining seven localities and all population groups in focus of the PBF exercise. Consultations were pivotal not only to ground-truth the analysis, but also to increase transparency on how the data collected was going to be utilised, and to enable communities' co-ownership of their locality action plan.

IDP identification and classificatory questions

108. IDP statistics can be separated into two distinct population groups: IDPs themselves, and IDP-related populations that are potentially of interest. Identification questions may focus just on IDPs or also cover some or all IDP-related populations.

- IDPs - Persons who have displacement-related protection needs:
  - IDPs in locations of displacement
  - IDPs in location of return
  - IDPs in other settlement locations;
- IDP-related populations:
  - Children born after displacement to at least one IDP parent;
  - Other non-displaced family members of IDPs;
  - Those who have overcome key displacement-related vulnerabilities;

  (See IRIS Figure 3.1 page 28).

109. IRIS recommends adding a set of screening questions to national household surveys similar to those recommended for population census (IRIS paragraph 241) where the sample design is suitable for obtaining reliable estimates. It also advises that camps, reception centres, informal settlements and other institutional accommodation should be included in the sampling frame.

110. While self-identification of IDPs is relatively widespread in current national survey practice, IRIS acknowledges self-identification / self-declaration merely as an exceptional last resort[28]. This is largely because of the risk of over or under reporting, due to differences in how individuals interpret or define the concept of forced displacement and the potential for individuals to misreport in cases where being identified as an IDP is linked to assistance provision, or where it evokes fear of social stigma, discrimination or persecution. It should therefore not be considered best practice, and – if at all – be used only to complement a more objective measurement practice.

111. The IDP identification question set recommended for censuses (Use Case A, paragraph 37) constitutes a "bare minimum" approach aimed primarily at census contexts where the potential for expanding the questionnaire is minimal. While this set attempts to capture all of the necessary criteria for being an IDP, it does not cover the full extent of complexity: for example, the place of usual residence before displacement may not be properly established and no information is collected on the causing event or multiple displacements, so it is not possible to accurately identify the 3 sub-populations of IDPs. The census questions also do not allow identification of IDP-related populations, i.e. the children of at least one IDP parent, where these children do not live in the same household as their IDP parent.

28 See IRIS, paragraph 107.

112. EGRISS has published a methodological paper, *Towards a standardized approach to identify IDPs, refugees and related populations in household surveys*[29] that expands on the IRIS recommendation and suggests survey question flows for use in household surveys. The paper discusses the questions that need to be included in order to address each element of identification of IDPs and related populations; the limitations and potential sources of inclusion and exclusion error; and the practical considerations around administering the questions within a survey. It also provides question flows for inclusion in MICS and DHS.

Capturing indicators of progress towards durable solutions, and priority SDG indicators

113. Beyond basic socio-demographic variables (age, sex) and the identification and classificatory questions discussed in the previous section, IRIS also proposes a composite measure for assessing if IDPs have overcome their key displacement-related vulnerabilities. The *Criteria and sub-criteria included in the Composite Measure of Overcoming Displacement-Related Vulnerabilities* is explained in the IRIS Chapter 4 and is illustrated in Table C.1 slightly expanded set of criteria and sub-criteria is also recommended in IRIS to comprehensively measure progress towards solutions.

Table C.1: Criteria and sub-criteria included in the composite measure of overcoming displacement-related vulnerabilities

Criteria	Sub-criteria
1. Safety and security	1.1 Victims of violence (protection from)
	1.2 Freedom of movement
2. Adequate standard of living	2.1 Food security
	2.2 Shelter and housing
	2.3 Medical services
	2.4 Education
3. Access to livelihoods	3.1 Employment and livelihoods
	3.2 Economic security
4. Restoration of housing, land and property	4.1 Property restitution and compensation
5. Access to documentation	5.1 Documentation

Source: IRIS, 2018 Table 4.3

114. There is currently no agreed list of indicators for measuring the sub-criteria, but EGRISS is developing a prioritised list of relevant indicators that can be obtained through surveys.

29 EGRISS Methodological Paper Series (2023) Towards a standardized approach to identify IDPs, refugees and related populations in household surveys

115. In the meantime, for each sub-criterion, there are many different indicators which can be chosen. The indicators selected in each national context should, as far as possible, be aligned with already tested and standardized indicators. The Durable Solutions Indicator Library provides a thorough overview of potential options (see IRIS, 2020 paragraphs 127-128). Indicators selected should be carefully chosen to ensure they are relevant for all three sub-stocks of the displaced population: IDPs in locations of displacement, IDPs in locations of return and IDPs in other settlement locations. For a full discussion of this topic refer to IRIS, 2020 paragraphs 135 – 137.

116. In addition to the list of indicators being developed by EGRISS, the IAEG-SDGs in collaboration with EGRISS has highlighted 12 SDG indicators that are of particular relevance to displaced people. For survey practitioners, this implies that these indicators should be captured in any survey including IDPs (whether this is a national survey that includes IDPs or an IDP-specific survey) if applicable and feasible within the scope of the planned survey. EGRISS does not itself promulgate definitions of these SDG indicators but rather refers to the relevant SDG metadata sheets and statistical standards for definitions and questionnaire specifications. Table C.2 provides references to such relevant international standards and specifications. Measurement of these SDG indicators for IDPs should be aligned wherever possible with the national statistical practice for capturing these indicators among the wider national population, to ensure comparability.

117. To make the assessment of the progress of the IDP and IDP-related population a comparator population is needed, this may be the general population or sub-national population in the vicinity of the initial displacement who have not themselves suffered displacement. These variables should also be available in the comparator population where relevant to enable comparisons to be made.

118. It is also important to consider collecting GIS data where possible, to enable locational analysis.

Table C.2: SDG priority indicators on forced displacement

SDG Priority indicator (with links to metadata)	Details	Relevant references
2.2.1 Prevalence of stunting among children under 5 years of age	Measure height of children of age 0-5 years; stunting indicated by height-for-age < -2 standard deviation from the median of the WHO Child Growth Standards	SDG indicator metadata sheet. In terms of survey questionnaire design, only fields for age/DOB and height are required. However, it must be noted that height measurement is non-trivial. See guidance on anthropometric data collection from DHS (Link), MICS (Link) and the FANTA Project (Link).
3.1.2 Proportion of births attended by skilled health personnel	Measure attendance of births by skilled health personnel (e.g. doctors, nurses or midwives) during delivery	SDG indicator metadata sheet The relevant question in DHS and MICS reads: 'Who assisted with delivery of (NAME)?' and answer choices include both health personnel and other persons. See figure 1 of this article for a direct comparison and discussion. MICS enquires about the last birth during the past 2 years, whereas DHS enquires about all births during the past 5 years.

6.1.1 Proportion of population using safely managed drinking water services	Measure improved basic drinking water source which is located on premises, available when needed and free of faecal and priority chemical contamination; where *Improved drinking water sources* mean piped water into dwelling, yard or plot; public taps or standpipes; boreholes or tubewells; protected dug wells; protected springs; packaged water; delivered water and rainwater *Located on* premises means if the point of collection is within the dwelling, yard, or plot *Available when needed* means households are able to access sufficient quantities of water when needed *Free from faecal and priority chemical contamination* means water complies with relevant national or local standards (or follow the WHO Guidelines for Drinking Water Quality)	SDG indicator metadata sheet. For survey questionnaire specification, see table 1 and "core questions" W1-W5 of the JMP recommendations (Link)
11.1.1 Proportion of urban population living in slums, informal settlements or inadequate housing	Measure the urban population living in slums or informal settlements that are defined by a lack of at least one of the following: *Access to improved water source (*see definition under indicator 6.1.1.) *Access to improved sanitation facilities* (access flush/pour-flush toilets or latrines connected to a sewer, septic tank or pit; ventilated improved pit latrine, pit latrine with a slab or platform and composting toilets/latrines) *Sufficient living area* (not more than three people share the same habitable room) *Housing durability* (permanent building materials for walls, roofs, and floor; not in need of major repair or in a dilapidated state; complies with local building standards; location is not near toxic waste, in a flood plain, on a steep slope, or close to rail, highway, airport or power lines) *Security of tenure* (evidence of documentation to prove secure tenure status, de facto or perceived protection from forced evictions) Measure inadequate housing, where inadequate housing is defined as housing that is not affordable to the household (i.e. the net monthly expenditure on its cost exceeds 30% of the total monthly income of the household)	SDG indicator metadata sheet

1.2.1 Proportion of population living below the national poverty line, by sex and age	Compute a comprehensive estimate of total household consumption or income (with consumption being the preferred welfare indicator) and construct a correctly weighted aggregate of consumption or income per person. Compare this aggregate with the national poverty line. To measure consumption, cover questions on food and non-food expenditure as well as food consumed from own production (either through recall questions using lists of consumption items or through diaries)	SDG indicator metadata sheet. Poverty measurement is a complex undertaking. No internationally standardized questionnaire module for capturing consumption expenditure exists (since food and non-food items consumed vary from country to country). When comparing consumption aggregates against a national poverty line, it is crucial to capture consumption in a way that aligns with the primary poverty survey instrument used in the country from which the poverty line was derived. This paper documents best practices around survey-based consumption and poverty measurement.
4.1.1 Proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex	Measure the minimum proficiency level in reading and mathematics for a. Grade 2/3 b. End of primary education c. End of lower secondary education Proficiency levels are established via the Minimum Proficiency Levels (MPLs)	SDG indicator metadata sheet. Collecting data for this indicator requires detailed learning assessments at the three levels specified in the indicator, which in most countries are administered through participation in various national, regional and/or international school-based assessments (as opposed to household-based surveys, which often will not sample sufficient numbers of children at a required age and education level). Hence, inclusion of forcibly displaced children in these learning assessments is the preferred data source for this indicator. Countries wishing to nevertheless include a learning assessment in a household survey covering forcibly displaced children should align the assessment method in the survey with the relevant school-based assessments that a country carries out. If not applicable, the MICS (Link, see Foundational Learning: module) and EGRA&EGMA may also provide useful survey-based approaches relevant to the lower schooling levels.
7.1.1 Proportion of the population with access to electricity	Measure access to electricity where access is only considered if the primary source of lighting is the local electricity provider, solar systems, mini-grids and stand-alone systems. Sources such as generators, candles, batteries, etc., are not considered due to their limited working capacities and their function as backup sources for lighting.	SDG indicator metadata sheet. A single survey question on primary source of lighting is sufficient to capture this indicator. See question HL2 in the broader set of energy-related questions recommended by WHO and the Word Bank (Link).

8.3.1 Proportion of informal employment in total employment, by sex	Measure all people of working age who, during a short reference period (one week) are engaged in any activity to produce goods or provide services for pay/profit. Measure informal employment which refers to people that in their main or secondary job were in one of the following categories: a. Own-account workers, employers, members of producers' cooperatives employed in their own informal sector enterprise b. Own-account workers engaged in goods production for their own final use c. Contributing family workers (no explicit written contract of employment, not subject to labour legislation or social security) d. Employees holding informal jobs	SDG indicator metadata sheet. See ILO's model questionnaires (Link) for labour force surveys which include the required questions for capturing informality, as well as UNHCR's guidance to labour measurement in forced displacement contexts (Link) for a more specific discussion.
8.5.2 Unemployment rate, by sex, age and persons with disabilities	Measure unemployed people, which are defined as those of working age who were not in employment (in a short reference period), carried out activities to seek employment and were currently available to take up employment given a job opportunity.	SDG indicator metadata sheet See ILO's model questionnaires (Link) for labour force surveys which include the required questions for capturing unemployment, as well as UNHCR's guidance to labour measurement in forced displacement contexts (Link) for a more specific discussion.
1.4.2 Proportion of total adult population with secure tenure rights to land, with legally recognized documentation and who perceive their rights to land as secure, by sex and by type of tenure	Measure secure tenure rights to all types of land use (residential, commercial, agricultural, forestry, grazing, wetlands), where secure tenure rights are compromised of two sub-components: a. Legally recognized documentation (the landholder reports having the right to bequeath the land) b. Perceived security of tenure (the landholder does not report fear of involuntary loss of land through disputed land ownership, inability to use land, fear of eviction or dispossession)	SDG indicator metadata sheet
16.1.4 Proportion of the population that feel safe walking alone around the area they live	Measure the feeling of fear of crime in a context outside the house; suggested formulation is "How safe do you feel walking alone in your area/neighbourhood? (Very safe/fairly safe/bit unsafe/very unsafe/ I never walk alone/don't know). Feeling safe refers to those respondents who feel "very safe" and "fairly safe".	SDG indicator metadata sheet. Survey questionnaire specification provided directly in the "methodology" section of the metadata sheet.

16.9.1 Proportion of children under 5 years of age whose births have been registered with a civil authority, by age	Capture whether the *birth of children* under 5 has been registered with a civil authority, where birth registration is understood as a universal recording of the birth, a *birth certificate* is issued as a record that documents the circumstances of the birth and a *civil authority* is an officially authorized body for registration of births.	SDG indicator metadata sheet

119. Examples of questionnaires from other IDP surveys are available in the UNHCR microdata library[30].

Sampling considerations when surveying IDPs

120. In theory, sampling in forced displacement contexts is no different from sampling elsewhere, and the same general principles and methods apply[31]. However, in practice the selection from and application of these methods is often faced with specific challenges when it comes to IDPs (IRIS paras 226-231).

121. Forcibly displaced persons may be hard to reach, either because they are difficult to identify, not willing to be identified, on the move or live in areas difficult to access. Moreover, displaced people often constitute a small proportion of the wider population and are not evenly distributed around a country or region – making it difficult or impossible to rely on sampling frames which don't allow for the direct identification of forcibly displaced households, or at least of small geographical areas where they are predominantly located. IDPs can live either in dedicated camps or they can live in houses among the general population, either as part of another household or as an independent household.

Identifying a sampling frame

122. The first question that guides the decision on sampling from IDP populations is how one can construct a frame from which to sample IDPs in a targeted manner. In the simplest case, this can either be a list of the wider population which allows identifying forcibly displaced, or a list solely describing the IDP population. Such a list can often come from administrative records, either from relevant government agencies or operational data from international agencies such as UNHCR or the International Organisation for Migration. In order to gain access to such sources it is likely that a data sharing agreement will be required, which can take some time to establish and should set out how data privacy will be protected - for example, by basing sample selection on anonymised data, so that only the personal details of those invited to take part in the survey are shared.

30 UNHCR Microdata Library

31 Useful reading on survey sampling in general can be found in UNSD's "Designing Household Survey Samples: Practical Guidelines", "Household Sample Surveys in Developing and Transition Countries", and "Sampling Rare and Elusive Population".

CASE STUDY: PROFILING OF INTERNAL DISPLACEMENT DUE TO VIOLENCE IN EL SALVADOR

The Government of El Salvador, with the technical leadership of the General Directorate of Statistics and Census and the Latin American Faculty of Social Sciences, and the technical support of UNHCR and JIPS, conducted a profiling of IDPs between 2016 and 2018. The objectives of the exercise were to identify the number of families displaced by violence in El Salvador in recent years, and to better understand the living conditions and specific needs faced by the victims of displacement.

There was no register of internally displaced persons or households available and it was not possible to complete a full enumeration of the whole country, so a sampling approach was needed that would allow estimates of the entire IDP population on a country level and also ensure that the sample was representative of the target population.

Geographic areas that were receiving flows of IDPs were identified through a data review and mapping exercise using interviews and workshops with key informants. This produced a list of 202 priority study areas containing a total of 648 census segments. For the displaced target population, a stratified random cluster sampling approach was chosen. For the non-displaced target population, a two-stage sampling approach was chosen. The process of selecting the two samples can be summed up in the following way:

Source: https://www.jips.org/uploads/2018/03/El-Salvador-profiling-report-EN.pdf

123. If one does not have an existing list of possible respondents, it needs to be created. The following methods are commonly used:

a. If the locations in which IDPs reside are known, or can be determined, and the National Statistics Office maintains a list of Primary Sampling Units (PSUs) for the country, the two can be used to identify those PSUs where IDPs reside. The first stage of sampling can then include either only those PSUs, or over-sampling from those PSUs. Relevant PSUs are selected randomly or systematically, and then a household listing is carried out or updated in the selected PSUs, capturing household's IDP status.

CASE STUDY: SAMPLING FOR THE 2018 MULTIPLE INDICATOR CLUSTER SURVEY IN GEORGIA

The 2018 Georgia Multiple Indicator Cluster Survey (MICS) was carried out by the National Statistics Office of Georgia in collaboration with United Nations Children`s Fund (UNICEF) and National Center for disease Control and Public Health (NCDC), as part of the Global MICS Programme. The primary objective of the sample design was to produce statistically reliable estimates of most indicators, at the national level, for urban and rural areas, and for the ten regions of the country and country-level estimates for the IDP population.

A multi-stage, stratified cluster sampling approach was used for the selection of the survey sample. The sampling frame was based on the 2014 General Population Census of Georgia. The primary sampling units (PSUs) selected at the first stage were the enumeration areas (EAs) defined for the census enumeration. A listing of households was conducted in each sample EA, and a sample of households was selected at the second stage.

The urban and rural areas in each of the ten regions were defined as the sampling strata. Each major stratum (Region by Urban/Rural) was divided into IDP and Non-IDP sub-strata. Since some of the PSUs have no IDPs, or the proportion of IDPs was very low, only the PSUs where the IDP population is more than 48 percent of the PSU population were included in the IDP strata.

Source: https://mics-surveys-prod.s3.amazonaws.com/MICS6/Europe%20and%20Central%20Asia/Georgia/2018/Survey%20findings/Georgia%20MICS%202018-SFR_English.pdf

b. A similar approach can be applied if the locations of IDPs are unknown, by drawing a random selection of PSUs and undertaking a listing exercise, using a short set of questions to elicit IDP status. In those PSUs with higher concentrations of IDPs – above a pre-determined threshold such as 10% - the household listing is extended to the neighbouring PSUs (or one nearest neighbour PSU). This process continues extending until PSUs no longer cross the threshold. In this way, PSUs with the highest concentration of IDPs are identified and the listing can be used as the sampling frame. This adaptive cluster sampling approach has been used successfully for refugees in urban areas with high concentrations of refugees and should be applicable to IDPs.

c. If a list of PSUs of reasonable quality does not exist, alternative ways to identify smaller geographical areas to sample from exist. Satellite, aerial or drone photos of areas in which IDPs are known to reside can be used to divide geographical areas into sampling units to be sampled and the households existing there listed to be sampled in a second stage.

d. If IDP locations are (partially) unknown, Mobile Phone Tracing may be an option. Digital trace data generated by IDP's mobile phones can help identify the areas in which IDPs are located and thus where to list/sample. This is not the same as identifying IDP's phones and then sampling them for a phone survey (see section (e) for comparison). Approaches such as these naturally require careful consideration around data privacy. Further discussion of non-traditional data sources is included in Use Case F.

e. Moving away from area-based sampling approaches, it may be possible to undertake a phone survey, assembling a list frame of phone numbers of the IDP population, by Mobile Phone Tracing or via Random Digit Dialling[32]. If launching a face-to-face IDP survey it may also be worth collecting phone numbers and consent for future phone-based panel surveys at the same time.

124. Sometimes a combination of different approaches is required. For example, if expanding a national survey to include IDPs living in camps and outside camps, it is possible that PSUs will be available for out-of-camp populations, but that the PSU frame exclude the camps. In this case, it may be necessary to use registration lists of IDPs in the camps as a complementary sampling frame.

Drawing a sample

125. Constructing a sampling frame of IDPs allows a survey to apply probability-based sampling methods. This ensures that every unit in the population has a known, non-zero chance of being selected into the sample, and its probability of selection can be accurately determined. This makes it possible to produce unbiased estimates of population totals, by weighting sampled units according to their probability of selection. It also allows for estimates of uncertainty, like confidence intervals. Probability-based sampling is considered best practice when sampling for surveys and ought to be the first choice - whenever it is possible to construct a sampling frame, i.e. a list of units to draw a sample from.

126. Including IDPs in a general household survey is likely to require a boost to the sample ratio for IDPs to ensure that a sufficient number are reached through the survey – because IDPs tend to be a small proportion of the overall population. The size of a boost will depend on the analysis requirements – for example the disaggregations and level of precision that are expected.

32 The usual risks of sample under-coverage resulting from phone ownership and network connectivity rates among the target population need to be taken into account.

CASE STUDY: BOOSTING THE HARMONIZED HOUSEHOLD LIVING CONDITIONS SURVEY IN CENTRAL AFRICAN REPUBLIC

The National Statistical Institute of CAR carries out the Harmonized Household Living Conditions Survey (Enquête Harmonisé sur les Conditions de Vie des Ménages). The 2021 survey will cover a total of 6,000 households nationally. Nearly 15% of the total population in CAR are IDPs (an estimated 702,000 as of February 2020) but detailed data on this vulnerable population is lacking. Alongside the IDPs identified through the regular sampling strategy, the sample size in IDP camps will be increased by 600 households. The expanded sample in IDP camps will allow more detailed analysis of the welfare and living conditions of IDPs in both IDP camps and in the general population.

Source: https://www.jointdatacenter.org/central-african-republic-car-idp-survey-as-part-of-the-2021-harmonized-household-living-conditions-survey/

127. If sampling from a highly concentrated population such as IDPs living in camps, it is generally not necessary to consider clustering the sample. Clustering is a technique that one uses to bring down sample dispersion and make travel during fieldwork manageable, but it comes at the cost of losing precision and driving up design effects. When sampling from a highly concentrated population there is no real cost saving from clustering but there is still a loss of precision, so it is preferable to obtain a registration list for the camp (if it exists and is sufficiently up to date) and use it as a complementary sampling frame from which to draw a systematic and unclustered selection of dwellings.

128. If carrying out a dedicated survey of IDPs, it is also important to identify an appropriate "host community" to include in the sample as a comparator population. There is no standard definition of host community, so at present this will need to be locally determined. Examples include the wider population that lives within a certain radius of an IDP camp, or households within the same administrative area.

129. It should be noted that responses to questions on attitudes, intentions to move, and decision-making power in the household may differ substantially by gender. When feasible and appropriate, using a random selection protocol of female and male respondents within a sampled household, or interviewing more than one member of each household, are ways to ensure more reliable data.

Non-probability sampling methods

130. When constructing a sampling frame is not feasible, reverting to non-probability-based sampling is sometimes considered. Commonly used non-probability sampling techniques are respondent driven sampling (RDS) or snowball sampling. If IDPs are difficult to find, either because they are few in numbers, are hiding, are homeless, on the move or live in an area difficult to access, such methods may be considered as an option of last resort. They should however be used with caution, as they do not allow for statistically representative, unbiased inference from the sample to the population. Application of non-probability sampling methods in official statistics production is uncommon.

CASE STUDY: NON-PROBABILISTIC APPROACH IN IDP PROFILING SURVEY IN THE STATE OF CHIHUAHUA, MEXICO

Residents in the state of Chihuahua in northern Mexico have suffered from violence-related displacement due to organised crime in the state, with populations displaced to cities or border regions. In 2022, the National Institute of Statistics and Geography (INEGI) worked with the Joint IDP Profiling Service and UNHCR to plan a survey to understand more about the displaced populations. This is a population that is difficult to locate for probabilistic survey purposes, so a non-probabilistic approach was identified as the most appropriate sampling strategy.

INEGI identified three routes to reach IDPs in Chihuahua.

- An existing crime survey had asked about the measures respondents had taken to protect themselves from crimes, including a response option for relocation. Those respondents were targeted for the IDP survey.
- The Executive Commission for Attention to Victims of the State of Chihuahua (CEAVE) provides shelters and facilities for victims of crime. Interviewers visited CEAVE locations to conduct interviews.
- A snowball approach was attempted to identify other potential displaced people; however, this was not as successful as expected due to the sensitivity of the topic (the underlying reason for displacement being organised crime).

In total, 1,255 interviews were completed out of a total of 1,388 planned, approximately 70% at CEAVE locations and 30% from the crime survey. By focusing on the population that received attention from CEAVE, there is a bias in the sample frame (because it is not representative of all displaced persons) which needs to be taken into consideration during the analysis of the results.

Limitations of registration systems as sampling frames

131. A frequent challenge with sampling frames derived from IDP registration systems is that the registers are only focused on a sub-set of IDPs receiving assistance; or can be incomplete, for example if areas of the country are inaccessible to registration systems, or the address/contact information can be outdated.

Box C.1

Guidance material on sampling in forced displacement contexts

Two dedicated resources on sampling in forced displacement contexts have recently been published and may be considered for further details and examples:

- Stephanie Eckman and Kristen Himelein (forthcoming): Innovative Sample Designs for Studies of Refugees and Internally Displaced Persons
- JIPS (2020): Sampling Guide for Displacement Situations and Practical Examples

Fieldwork planning

132. Before any survey fieldwork commences, it is advisable to conduct one or more field visits to fully understand and plan for the locations and practicalities involved in reaching IDPs. This can be important in overcoming the additional challenges that surveying IDPs can entail. Ideally one or more of the most challenging locations / scenarios should be visited, with the following objectives:

- Observe the situation on the ground;
- Test whether the sampled households can in practice be located and accessed and make contingency plans if not, for example through engaging local community leaders, local authorities or landlords for access;
- Build relations between the different agencies and communities involved in implementing the survey – trust and rapport can be essential to success of the survey;
- Identify any language barriers and how best to address them (questionnaire translations, availability of interpreters).
- Test the questionnaire, both in terms of content and flow. Do the questions lead to correct identification and classification of refugees; are there any issues with translations; are there any sections of the questionnaire that are seen as more sensitive and potentially might result in non-response.

133. There may be additional costs for extended duration of fieldwork compared with a standard survey, reflecting the challenges of identifying IDPs, and potentially greater use of supervisors to train, monitor and support enumerators.

CASE STUDY: PILOT VISIT DURING IDP PROFILING SURVEY IN THE STATE OF CHIHUAHUA, MEXICO

As part of its 2022 survey of IDPs in Chihuahua, the National Institute of Statistics and Geography (INEGI) planned to conduct interviews at the office locations of the Executive Commission for Attention to Victims of the State of Chihuahua (CEAVE), which provides shelters and facilities for victims of crime. Residents in Chihuahua have suffered from violence-related displacement due to organised crime in the state, with populations displaced to cities or border regions.

To ensure the success of this element of the survey, INEGI organised a pilot visit aiming to address the following key issues:

- Logistics of interviewing at CEAVE offices – finding a suitable location (privacy concerns)
- Communicating the importance of the survey, to engage CEAVE officials and potential respondents.
- Testing the questionnaire through 20-25 test interviews

As a result of the pilot visit, enumerators were provided with additional training to improve their success in obtaining complete interviews, given the challenging topics covered. In total, the IDP survey achieved 1,255 complete interviews, of which 70% were at CEAVE locations.

134. Surveys may involve a communication plan at a national or local level, to ensure that residents are aware the survey is taking place and primed to take part at the correct time. It is important to ensure that displaced people are considered as part of the communication plan, if good coverage is to be achieved. There are no fixed prescriptions that will work everywhere, but experience from various countries should be observed and evaluated. The main messages to convey include:

- That the survey will include displaced people, refugees and IDPs and that their responses are important.
- How the data will be used (positive messages) and clarity that it is not linked to allocating assistance.
- That confidentiality and data security are assured
- Any special arrangements that have been made to include displaced people (such as availability of interpreters).
- What opportunities there will be for respondents to hear about the survey results following data collection. Depending on the scale of the survey, it may be appropriate to offer some dissemination aimed specifically at the target population.

The communication plan may also consider raising awareness of any existing referral mechanisms available to the community, for feedback or protection concerns, according to the principles of "do no harm".

135. To reach displaced people, communication at local and community level is often the most effective. Examples of communication strategies include:

- Leaflets or visits to relevant institutions (refugee camps, etc) in the lead up to Census, with interpreters available if language is likely to be a barrier.
- Identification of and engagement with community leaders before the survey, to establish the most effective way of reaching displaced populations and to gain their buy-in and support.
- Acknowledge the potentially sensitive nature of the question topics for displaced populations and explain the purpose of collecting the data and principles of data protection.

CASE STUDY: COMMUNITY ENGAGEMENT IN DATA PROCESSES IN SYRIA

The Urban Analysis Network Syria (UrbAN-S) project, implemented in 2018-2019, produced area-based urban profiles that provided a snapshot of the physical and social conditions from a conflict and displacement context. The project was funded by the European Union and implemented through a collaborative approach, led by iMMAP with JIPS, Mercy Corps' Humanitarian Action Team, and the European Commission Joint Research Centre.

Due to restrictions on international organisations' work in Syria in 2018-2019, it was difficult to access communities to conduct sample-based household surveys. The project therefore used an alternative approach that relied on data collection from 'community focal points'. Transparent criteria were defined to identify which community members would be included in the data collection, based on chain referral (snow-balling) and assessing the individual's connections, reputation, influence, motives and potential biases, length of time living in the area and aspects of their understanding of the situation of the wider community and geographical location.

The engagement with community members during data collection and the efforts put in to identify representative members to speak on behalf of the wider communities made space for a two-way information flow between the research teams and the communities.

Source: Community engagement in Data Processes in Displacement Contexts. JIPS, 2022. https://www.jips.org/jips-publication/community-engagement-in-data-processes-in-displacement-contexts-jips-2022/

Enumerator training

136. Displaced people are often harder to reach and can need support in order to take part in a survey and this requires specific content in the enumerator training course. Allowing sufficient time to address these issues in the enumerator training is key to delivering a successful survey. It may take between one and two days to cover the appropriate material.

137. Topics to cover in enumerator training include:

- The definitions and concepts used in statistics about displacement, which may be unfamiliar to many experienced field staff. For example, IDP concepts such as habitual residence, usual residence, migration concepts, refugee categories, meaning of displacement, meaning of refugee and IDP, and reasons for displacement and migration.
- Accurate recording of all members of the household, including temporary members who may be displaced. This is important as displaced people may fail the interviewers standard understanding of 'usual members of the household', leading to them being inaccurately treated as visitors and excluded from household rosters.
- The sensitivities around interviewing IDPs and the data protection implications. When conducting surveys, the displaced population has special needs or conditions of interview, and the 'no-harm' principles and UN ethical standards on data collection and data protection must be applied. Enumerators must be aware of the existing referral mechanisms for community feedback or protection concerns, should any issues present themselves during the course of the data collection.
- Building trust and rapport, and skills in interviewing vulnerable people, such as children and disabled people. People who have been displaced have suffered trauma and any interview may ask people to relive painful experiences.

Strategies for managing personal well-being for the enumerators, if they are likely to be exposed to traumatic testimonies during the data collection – which can become emotionally exhausting.

CASE STUDY: INTERVIEWING VULNERABLE IDPS IN GEORGIA

In 2014, the Institute of Social Studies and Analysis in Georgia undertook a survey of IDPs to improve understanding of progress towards durable solutions. In preparation for the survey, enumerators received a day's training on the concepts and definitions involved and the sensitivities around surveying IDPs, including how to reassure IDPs about the nature of the survey and confidentiality of responses.

Over 2000 IDPs took part in the survey, sharing their aspirations, thoughts and feelings regarding their future. For many of them, this was an important step in their own reflection process on returning home, integration, and relocation in displacement.

While conducting a survey sounds like a very technical exercise, for the IDPs involved it was an emotional journey as interviewers steered women, men, girls, and boys through the 63 questions of the six-page questionnaire. Tears were shed in memory of the past and as hopes for a better future were expressed. IDPs trusted the survey process and opened their doors, hearts and minds to ensure that the survey results will help to shape their future destiny in a way that reflects their desires.

Fieldwork in practice

138. On the whole, provided fieldwork planning and enumerator training has taken full account of the need to cover displaced people, there are few issues specific to IDPs in terms of the actual fieldwork. However, it is worth noting that effective feedback from enumerators during any survey can be crucial in identifying any emerging issues in reaching IDPs, which might be addressed by continued or improved communication and community engagement.

139. If budget allows, it is advisable to undertake more monitoring of fieldwork than for a standard survey, reflecting the challenges associated with identifying IDPs and potential issues overcoming sensitivities around data collection. Increased availability of supervisors can help to monitor the quality and consistency of data collection and play a facilitating role, communicating the purpose of the survey with more authority, to gain trust and avoid non-response.

Data quality

140. There are few data quality issues that are specific to surveys including IDPs, and so data quality assurance should follow standard processes, such as monitoring response rates and monitoring the completeness and distributions of key indicators. The only specific concern of note is the risk that survey respondents may believe that their responses will influence the provision of humanitarian aid and will thus misreport consumption in an attempt to influence its distribution.

141. Often it is possible to triangulate survey data with data from other sources, such as administrative data from other government agencies or operational data produced during the regular activities of humanitarian organisations, to make sure that high level trends match between the various available datasets without any major deviation. For example, if the variation between two stocks is not significant, but administrative data identifies a high number of smaller, shorter term movements (e.g. in the case of preventative evacuations leading to quick returns, or mass movements leading to cross border returns) then this should help to verify or correct official statistics on flows or alert the need for better or more frequent data collection. Similarly, through comparing with cross border data, if movements lead to significant cross-border flight or returns, the datasets on both sides of the border should correspond with each other to ensure validity. Further discussion of administrative data, operational data and non-traditional data sources is included in Use Cases D, E and F respectively.

142. Even if variations in operational flow data from incidents do not match variations in stocks, the differences between the different stock types should be proportional. However, an increase in the number of displacements does not mean that a similar increase will appear in stocks, as those displacements could relate to the same people being displaced on multiple occasions. For example, if the numbers of IDPs in an area decreases, the numbers corresponding to that decrease should be reflected in another stock, e.g. there are 10,000 less IDPs in Borno State in Nigeria, because there are 10,000 more Nigerian refugees in Cameroon.

Data protection / confidentiality / disclosure

143. Due consideration should be given to privacy in accordance with national and international laws. The UN Fundamental Principles of Official Statistics states that "individual data collected by statistical agencies for statistical compilation … are to be strictly confidential and used exclusively for statistical purposes"[33]. Disclosure should also include considerations of interoperability: how datasets on different population types talk to each other, their ability to share information on those populations, or even on specific individuals, their characteristics and their needs.

Analysis and dissemination

144. The process of analysing and presenting the survey data collected on IDPs is no different from that for any other data and should be planned to meet the needs of users. This section briefly describes the outputs that are commonly produced to meet users' needs, with references for further information if required.

145. Data collected on identification and classification of IDPs is used to produce analysis of stocks and flows for the different categories of IDPs and IDP related populations, presented as counts and proportions cross-tabulated by sex and age group, details are in IRIS paras 114 – 125.

33 UN Fundamental Principles of Official Statistics Implementation Guidelines

146. Some analysis will be required to identify the stocks and flows of those who have overcome key displacement-related vulnerabilities, in particular to derive the variable in the databases which enables the analyst to identify those who have overcome those vulnerabilities, and to remove them statistically from stock. It should be noted that the full methodology for determining IDPs to be removed from stock is in its early stages. A more developed methodology will be added to future editions of this manual.

Stocks

147. Analysis of stocks is recommended for all categories of IDPs and IDP-related populations where possible.

   a. Counts of each type of IDP and IDP related populations, by sex and age, or age group, place of usual residence, place of habitual residence, and years since initial displacement. If age group is used, then it is important to distinguish children from adults aged 18 years and over and analysis should include the total number of IDP-related unaccompanied and separated children under 18, by sex and age. For some purposes it may also be helpful to identify those under 16 years of age.

   b. Proportions of IDPs by the length of time the IDP was displaced; the location of the IDPs and if they have returned to their habitual places of residence, been resettled or are in locations of displacement; and proportions of women and children among the IDPs. IRIS elaborates on these, drawing out a recommended set of indicators – see Box C.2.

Box C.2

Basic indicators for stock of IDPs (from IRIS, 2020)

i. Percentage of the total population who are IDPs.*#
ii. Proportion of women among all IDPs.* #
iii. Proportion of children among all IDPs.
iv. Proportion of IDPs who were first displaced in the last 5 years.
v. Proportion of IDPs who were first displaced in the last 10 years.
vi. Proportion of IDPs who were first displaced over 10 years ago.
vii. Proportion of IDPs who were displaced more than once, whose first displacement was during the last 5 years.
viii. Proportion of IDPs who have returned to their habitual place of residence and have not yet overcome key displacement-related vulnerabilities (IDPs in locations of return)#
ix. Proportion of IDPs who have settled elsewhere in the country and have not yet overcome key displacement-related vulnerabilities (IDPs in other settlement locations)#
x. Proportion of IDPs in locations of displacement and who have not yet overcome key displacement-related vulnerabilities (IDPs in locations of displacement).#

*Key: # suitable for IDP-related stocks; * suitable for stocks of those who have overcome key displacement related vulnerabilities*

Flows

148. Flows are inherently more difficult to collect through surveys and in practice the recommended indicators for in-flows are likely to need to come from administrative data, while measuring the outflows (also included in Box C.3) is a demanding statistical exercise for which the advice is still a work in progress.

149. Where analysis of flows is possible, it is recommended for all categories of IDPs and IDP-related populations, with in-flows and out-flows cross-tabulated by sex, age, reason for displacement, and place of usual / habitual residence. The full list of recommended tabulations from IRIS is replicated in Box C.3.

Box C.3

Basic IDP flow tabulations (from IRIS, 2020)

Basic inflow statistics for IDPs

a. Total number of IDPs in a country who were forcibly displaced for the first time, during a specified period of time, by sex and age
b. Total number of persons in a country who were forcibly displaced for the first time, during a specified period of time, by sex and reason for displacement
c. Total number of persons in a country who were forcibly displaced for the first time, during a specified period of time, by sex and by current place of current usual residence (province/district)
d. Total number of persons in a country who were forcibly displaced for the first time, during a specified period of time, by sex and by place of habitual residence (province/district)

Basic inflow statistics for IDP-related persons

a. Total number of children born to at least one IDP-parent after the parents' last displacement, during a specified period of time, by sex and age
b. Total number of children born to at least one IDP-parent after the parents' last displacement, during a specified period of time, by sex and parents' place of habitual residence (province/district)
c. Total number of children born to at least one IDP-parent after the parents' last displacement, during a specified period of time, by sex and current place of usual residence (province/district)

Basic outflow statistics for IDPs

a. Total number of IDPs who have died or emigrated during a specified period of time, by sex and age.
b. Total number of IDPs who have overcome all key displacement vulnerabilities during a specified period of time, by sex, age and by current place of usual residence.
c. Total number of IDPs who have overcome all key displacement-related vulnerabilities during a specified period of time by habitual place of residence and current place of usual residence and main reason for initial displacement.

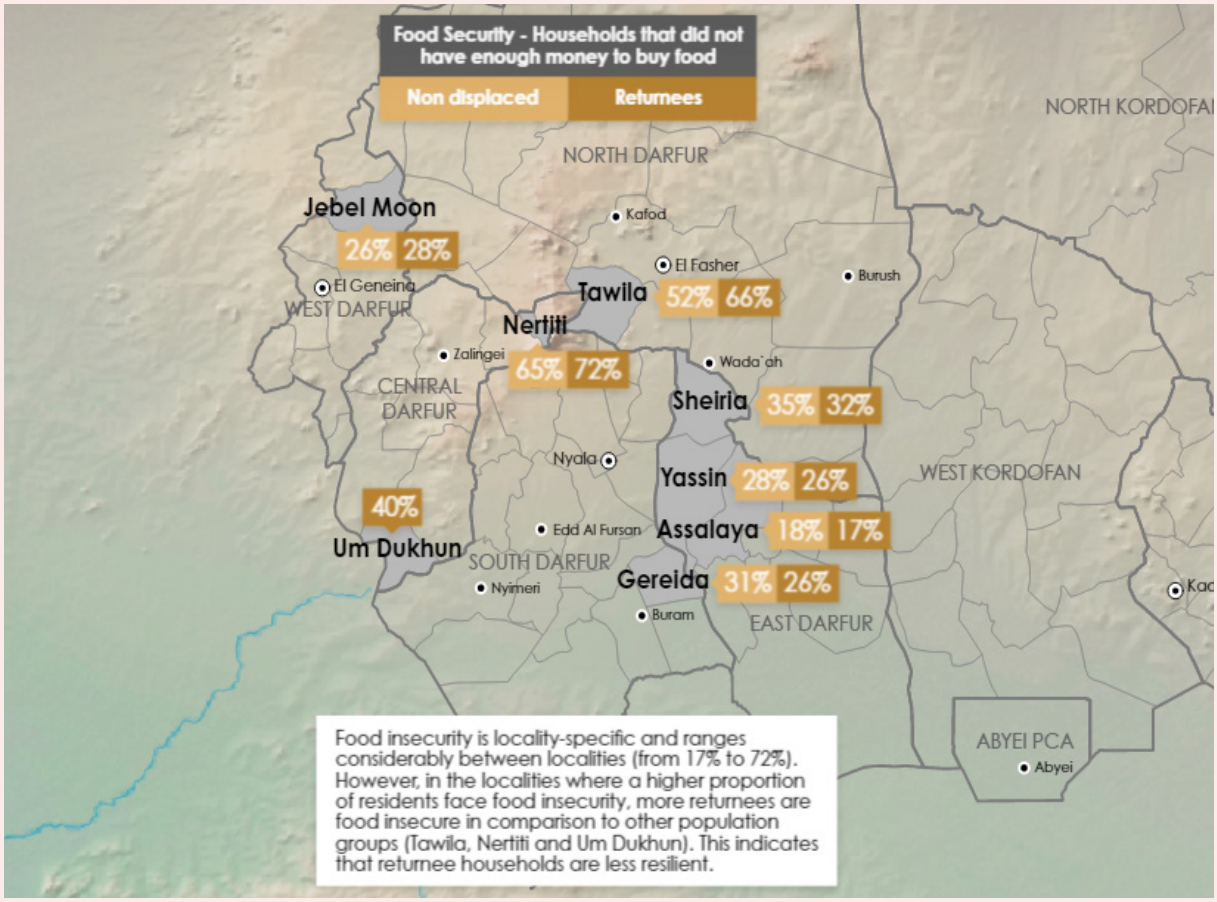
Basic outflow statistics for IDP-related persons

a. Total number of children born to at least one IDP-parent after the parents' last displacement, who have died or emigrated during a specified period of time, by sex and age.
b. Total number of children born to at least one IDP-parent after the parents' last displacement, who have overcome all key displacement vulnerabilities during a specified period of time, by sex, age and by current place of usual residence.
c. Total number of children born to at least one IDP-parent after the parents' last displacement, who have overcome all key displacement-related vulnerabilities during a specified period of time by parents' habitual place of residence and current place of usual residence and main reason for parents' initial displacement.

150. Geospatial analysis is also recommended, to visualise the stocks and flows.

CASE STUDY: IDP GEOSPATIAL ANALYSIS, PEACEBUILDING WITH DURABLE SOLUTIONS FOR DARFUR'S DISPLACED AT THE CORE, INSIGHTS FROM EIGHT LOCALITIES (EXCERPT)

In 2020, a large-scale analysis covering eight localities across Darfur's five states was initiated to provide a shared evidence-base to support peacebuilding and durable solutions under the UN Peacebuilding Fund. The key insights from the resulting locality studies have been condensed into five thematic briefs.



The best examples of data dissemination are often through online resources, which enable dynamic access to the data and can tell the data story in an engaging and interactive way. This Case Study captures one image from an online resource, to illustrate and inspire. The full details are available in the following source material: http://dswgsudan.org/pbfdarfur.

Socio-economic indicators and basic progress tabulations

151. Statistics on the progress made by IDPs in overcoming their displacement related vulnerabilities should focus on the total number of IDPs who have achieved the defined target for each of the different durable solutions criteria and related sub-criteria during a specified period of time, by current place of usual residence (district/province). This should be measured against a reference population, either the national or regional average, or the host community.

152. The two measures on progress which can be used to produce official statistics on internal displacement are:

- The durable solutions progress measure can be used to show the change in the share of IDPs who have overcome vulnerabilities linked to the criteria of the IASC Framework on Durable Solutions for IDPs over time; and
- The composite measure can be used to specify whether all key displacement-related vulnerabilities have been overcome and thus, whether or not persons can be taken out of the total IDP stock.

153. Over time, as new data are collected and more testing conducted in different national contexts, the recommended progress and composite measures can be further refined, including a list of agreed indicators and other relevant guidance. This will be included in future versions of this manual.

154. There are no further specific recommendations relevant to IDPs on how these or other data on socioeconomic indicators (including SDGs) should be analysed and presented. Typical tabulations could include cross-tabulations by category of IDP, sex and age group.

155. In planning how the survey analysis will be disseminated, if part of a wider survey it is preferable for IDP analysis to be disseminated in the main survey report as a standard disaggregation but it may also be impactful to produce a separate report focussing just on IDPs. There is no right answer here, it will depend how users' needs can best be met in the specific national context. It is also important to consider how survey findings can be shared with the IDP communities themselves. This feedback loop is important: as well as meeting the needs of a key user community it demonstrates what the survey is for and promotes the importance of the survey, encouraging future engagement and participation.

CASE STUDY: IDP DATA DISSEMINATION, COVID-19 IMPACT MONITORING – IDPS IN BURKINA FASO (PUBLISHED BY JOINT DATA CENTER, EXCERPT)

This is an excerpt from a statistical brief that presents the results from the third round of High Frequency Phone Surveys on Internally Displaced People (IDPs) in Burkina Faso, conducted between June 28 and July 20, 2021. The survey was designed to assess the socioeconomic experience of IDP households during the Covid-19 pandemic.

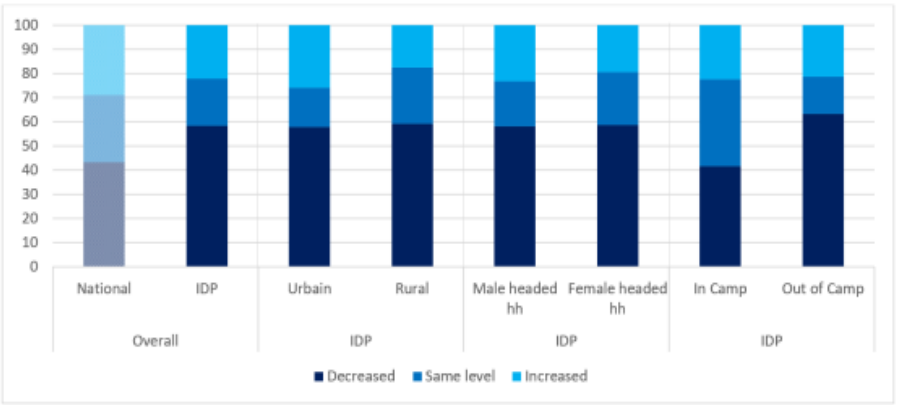
Burkina Faso • COVID-19 IMPACT MONITORING – IDP

SOURCE OF INCOME

While agriculture is the most important sector of employment for Burkinabè households generally, this sector is less important for internally displaced Burkinabè. There are three probable reasons: first, a lesser share of IDPs are engaged in agriculture (82 percent of the national rural sample were employed in agriculture, compared to just 57 percent of employed rural IDPs – not shown here); second, food produced by agriculture in Burkina Faso is mainly used for own consumption (see Figure 4); and third, because the July round was conducted towards the end of planting activities, the survey did not capture revenue from the sale of agricultural products.

Even so, the economic slowdown resulting from Covid-19 caused a reduction in total income for most of IDP households (Figure 3). Nearly six in ten displaced households reported that their income decreased over the year prior to the survey (43 percent for non-IDP households), and only 22 percent experienced an increase of total income (compared to 29 percent for non-displaced households). Importantly, income reduction was more common for IDPs living out of camps (63 percent, compared to 42 percent for IDPs living in camps).

Figure 3: Total household income compared to the same period in the previous year



Source: https://www.jointdatacenter.org/wp-content/uploads/2022/08/Burkina-Faso_brief-no3.pdf

USE CASE

D

USING GOVERNMENT ADMINISTRATIVE DATA

PHILIPPINES.
2021 birth registration in Kasanyangan, Zamboanga City.
©UNHCR/Martin San Diego

EGRISS

Administrative data – general background

156. There are many potential sources of data on displaced people other than surveys and censuses. This section focusses on administrative data sources and covers the issues relevant to the production of statistics on both refugees and IDPs. Eurostat defines administrative data as "the set of units and data derived from an Administrative source. This record keeping can be done by institutions belonging to the government sector or by private organisations"[34].

157. Administrative data are a useful source of information for measuring both the stock of displaced populations and the flows, showing the direction of displacement trends, and potentially for developing sampling frames. If they include variables allowing the identification of the target population, or if the specific administrative source can be linked at the individual level to another data source such as another register, a survey or a census, then they have potential to be part of a rich source of regular official statistics, including socio-economic indicators.

158. The priorities when considering using administrative data as a source of information on refugees and IDPs are:

- Identify the different available administrative sources and understand how they cover displaced people.
- Establish what the access requirements are, including the legal framework for accessing any sources that are potentially useful.
- Investigate the options for linking data sources to provide greater insights.
- Investigate and understand the limitations of any sources that will be used and build links with data owners to address the limitations wherever possible.

159. Administrative data have the advantage of being readily available at no additional cost and often more timely (up to date) than survey data. This can be particularly important for refugees and IDPs, as populations can change rapidly. However, because administrative data are not collected with statistical purposes in mind it is important to spend some time understanding the data processes to be aware of any limitations, such as definitional differences or data quality issues. Any limitations relating to the use of the statistics should be conveyed at the point of dissemination.

160. Statisticians should seek to build and maintain strong links with the organisations that own administrative registers, to raise awareness of the statistical uses of the data and foster collaboration: it may be possible to enhance the statistical value of administrative data through small changes to the underlying register.

34 Eurostat ESS Vision 2020 ADMIN (Administrative data sources).

Potential sources of administrative data

161. There are many administrative data sources that can provide relevant statistics on refugees and IDPs. These include:

- **General population registers** maintained by governments, covering the total resident population of a country.
- **Other government administrative registers for the wider population** including residence permit registers, tax registers, social security registers, register of border crossings, work permits, health registers, education registers, housing registers and land ownership registers[35].
- **Specific refugee / IDP registers maintained by government administrations** or organisations responsible for refugees, asylum seekers and IDPs, including registration of displaced people in camps and/or elsewhere, asylum seekers registers (applications), registers on decisions granting international protection, resettled persons register.

162. International agencies' databases (for example, registration systems, administration of cash assistance, etc.) are also administrative registers by their statistical nature, but because they are not government-owned and come with very different access requirements they are discussed separately in Use Case E and referred to in the Compilers' Manual as "operational data".

General population registers

163. A population register is a regularly updated administrative database of uniquely identifiable individuals who are usually resident in a country. In countries that maintain a population register, the register is the basis for population statistics, particularly on stocks and flows and can support a wide range of other statistics, through data linking which is facilitated by the unique identifiers in the dataset.

164. Refugees and related populations may be included and identifiable in a population register, but it is more common that asylum seekers are not treated as part of the usually resident population and are therefore omitted. In this case, identification would have to be achieved through data matching with other sources, for example administrative records of asylum applications or residence permit registers. See further information on data linking (paragraph 174). IDPs would usually be included in a population register but may not be identifiable as IDPs without linking to another source. Access requirements vary, but often access is permitted for statistical purposes, subject to data confidentiality agreements.

35 The United Nations' *Recommendations on Statistics on International Migration Revision 1* (1998) provides useful guidance for utilising administrative data sources for migration statistics in general, and refugee statistics in particular. These recommendations can be adapted for IDP registers, however IDPs are often not included.

CASE STUDY: USING THE CENTRAL POPULATION REGISTER TO PRODUCE REFUGEE STATISTICS IN NORWAY

Statistics on those with a refugee background in Norway are produced from the Central Population Register (CPR). The CPR of Norway was established in 1964 based on the 1960 Population Census. A unique 11-digit personal identification number (PIN) was introduced at the same time. The CPR includes all persons who have ever been a (legal) resident of Norway since 1960, regardless of their citizenship. It provides information on their name, address, residence status, place or country of birth, citizenship, reason for immigration, country of immigration or emigration, marital status and PIN numbers of spouse, mother and father. Persons who die or emigrate are not deleted from the register, but a code for their status is changed. All vital events and migrations and address changes are registered in the CPR.

Data on asylum seekers is transferred to the CPR from the Norwegian Directorate of Immigration (UDI). All asylum seekers must register with the Norwegian immigration authorities, and their case data and personal data are stored in a database that contains data on UN convention refugees as well as refugees who enter Norway on their own and apply for asylum.

Data are transferred regularly from the CPR to Statistics Norway, where it is linked with processed data from UDI enabling Statistics Norway to produce statistics on stocks and flows of people with a refugee background.

Government administrative registers for the wider population

165. There are many types of government administrative registers, other than population registers, that have potential to be a source of data on refugees and IDPs. The strength of these additional sources is often as a source of statistics on sub-categories of refugees and IDPs, rather than contributing to estimates of stocks and flows for the wider refugee / IDP population. For example, tax registers may provide information on refugees who pay tax; health registers may be a source of statistics on those who have engaged with the health system; and education registers may enable comparisons between children of refugees or IDPs and the wider population.

166. The extent to which refugees and IDPs are identifiable in government administrative registers will vary considerably[36] and will often rely on being able to link to other data sources, for example on asylum applications. As with population registers, access requirements vary, but often access is permitted for statistical purposes, subject to data confidentiality agreements.

167. A key limitation of government administrative registers can be the ability to identify unique individuals, rather than events. For example, registers of border crossings may include multiple records for people who have been returned to their country of origin and subsequently re-crossed the border; similarly, registers of work permits may include multiple records for individuals if people are required to renew permits. If the data source includes unique identifiers for individuals then such issues can be avoided.

Specific refugee / IDP registers maintained by government administrations

168. Many countries use administrative systems to co-ordinate the asylum application and decision-making process or to register displaced people in camps or other locations. Where this is done by government administrations – or on their behalf by other organisations – the data can be an important source of statistics on refugee / IDP stocks and flows.

169. Administrative registers are the main data source for statistics on asylum applicants and decisions taken during the asylum application process in regions such as Europe and Northern America. In the EU and in some additional neighbouring countries, they are also the main source of statistics on the number of refugees living in the country, through the residence permits statistics (which include categories such as refugees legal status, subsidiary protection, humanitarian reasons).

170. The use of "registers of asylum seekers" is specifically recommended for estimating the stock of asylum seekers (by length of stay in the country) and the number of new asylum applications within a given year (United Nations, 1998, p. 19), and for producing statistics on the outcome of the asylum procedure.

171. By their nature, these administrative systems will have good coverage of refugees or IDPs within their scope – but (particularly for camp-based registers) this may not be the entire stock, for example it may not include refugees or IDPs outside camps. If contributing to a full picture of stocks and flows for the entire country, care should be taken to understand any potential overlap or duplication of coverage between administrative registers.

36 For example, this UNESCO paper discusses the challenges identifying refugees in Education Management Information Systems.

CASE STUDY: PRODUCING STATISTICS ON FORCIBLY DISPLACED PEOPLE IN BOSNIA AND HERZEGOVINA

Statistics on IDPs and refugees in B&H are produced by a member of the National Statistical System, the Ministry of Human Rights and Refugees. This is a state level organisation responsible for unifying the data on displaced persons which is collected by municipalities in all three administrative units of B&H, and for producing statistical reports.

Producing statistics on refugees and IDPs was initiated by the government at the beginning of the war in the 1990s. When the war ended the first comprehensive official registration of persons in B&H was carried out in late 2000 by Ministries and local authorities. The registration and estimation of the number and the status of displaced persons was first completed in 2005.

The administrative records in the Database of Displaced Persons (DDPR) are now the main source of statistics on refugees and IDPs. This key data source establishes someone's status as IDP or refugee. The database is supplemented by census data. The questionnaire from the census in 2013 asks if a person had been a refugee after 1991, or if the person has returned from 'refuge'. It asks for details of the settlement that the person was displaced from; whether the person has returned; and if the person is still legally considered to be a displaced person.

Summary of key features of administrative systems and their limitations

172. Table D.1 illustrates the types of data and analysis that different sources of administrative data may support, if they identify refugees and IDPs or they can be linked with a source that enables such identification. As well as stocks and flows, administrative data have potential to support analysis of refugees needs and integration and IDPs progress toward durable solutions, as described by the indicators referenced in Use Cases B and C.

173. It is important to note that the identification of refugees and IDPs in administrative sources (and other sources which may be considered for data linking) may not be equivalent to identification following a set of survey questions, for example. There can be important definitional differences which need to be considered, particularly if using non-traditional sources as a source of statistics for comparison or to augment survey results: the two may not be covering the same population.

Table D.1: Analysis supported by administrative sources and their key limitations

Source	Stocks and flows analysis supported	Socio-economic analysis supported	Completeness and other limitations
Population register; register of foreigners	- Key source of statistics on stocks and flows - Basic characteristics, for example age, gender - Geospatial analysis, usual residence and previous residence	n/a	- May not routinely identify refugees or IDPs. - Need to ensure statistical laws allow access - Frequency of update.
Refugee / IDP register	- Key source of statistics on stocks and flows - Basic characteristics, for example age, gender - Geospatial analysis, usual residence and previous residence		- Should be complete, but consider usual quality concerns, presence of duplicates, time lag in registering, update frequency.
Asylum application register	- Key source of statistics on stocks and flows - Basic characteristics, for example age, gender - Geospatial analysis, usual residence and previous residence	n/a	- Covers asylum seekers only, not IDPs. - Won't cover refugee-related populations. - Refugees granted asylum before establishment of register won't be identified.
Residence permit register	- Key source of statistics on stocks and flows - Basic characteristics, for example age, gender - Geospatial analysis, usual residence and previous residence	n/a	- May not accurately identify sub categories of those requiring residence permits. - Possibility of duplication
Work permit register	- Stocks of refugees / IDPs with work permits	- Legal access to labour market	- Only captures those of working age. - Only captures those within the formal working environment.
Tax register, Labour Market Information System	- Stocks of refugees / IDPs registered for tax or employment.	- Tax details - Analysis of employment status and income status, occupation, and comparisons with wider population	- Only captures subset of refugees / IDPs who pay tax

Social security register	- Stock of refugees / IDPs in receipt of social security, as a proportion of total stock of refugees / IDPs and comparison with wider population	- Poverty and deprivation	
Border crossings register	- Flows of refugees and asylum seekers. - Contributes to estimates of total stocks.		- Duplication from individuals crossing borders multiple times. In flows likely to be more reliable than out flows. - May not identify refugee / IDP status of individuals, or accuracy of identification may be limited.
Health register	n/a	- Prevalence of health issues amongst refugee / IDP population. - Immunisation / vaccine coverage - Potential contribution to wider indicators of health status of refugees / IDPs	- Only covers those who engage with official health system. - Identification of refugee / IDP status within register is likely to be a limiting factor.
Education register	- Stocks of refugee / IDP school age children	- Educational attainment, literacy - Participation in education - Participation in preschool education and access to childcare	- Only covers those of school age, missing those who do not attend school. May not cover those who attend a non-government school.
Other sources, for example electricity subscribers	n/a	Source-specific data to link across sources to generate new insights. For example, the extent of electricity subscriptions in locations predominantly refugee / IDP might contribute to understanding of poverty and housing conditions.	

Data linking

174. There is no single source of data that can draw a complete statistical picture of refugees and IDPs. Because of the variety and the varying coverage of the possible sources, data linking (also called data integration) is a promising solution for improving the quality of refugee and IDPs statistics. Recent projects like the UNECE High Level Group for the Modernisation of Official Statistics (HLG-MOS) or the Eurostat European Statistical System 2020 vision project in the area of data integration illustrate the growing interest in data linking.

175. Data linking can address several challenges when compiling official statistics on refugees and IDPs by:

- Improving the coverage of target groups;
- Improving data availability, for example by adding variables especially socio-economic variables related to refugee and IDP populations;
- Improving timeliness, by linking survey data with regularly updated administrative data;
- Improving data quality and accuracy, for example by cross-referencing different sources in order to assess refugee and IDP status, or by deriving a more complete sampling frame.

176. Data linking is not limited to linking across administrative systems. The census, surveys of the general population or refugees / IDPs, and non-traditional data sources all have potential to add value through data linking with administrative data (or indeed with each other) for example through probabilistic data matching. This section provides some high-level key messages on data linking which can help and guide practitioners in NSOs interested in applying data linking to refugee and IDP statistics. The technical details of data linking are not specific to refugees and IDPs, so are not included here, but the use of data linking is illustrated through case studies of good practices in NSOs.

Unique identifiers

177. The key to successfully linking data sources (and the biggest challenge) is the availability of – or possibility of building – unique identifiers (UID) for statistical units, common across all the data sources involved. Examples of sources that may yield UIDs for refugees and IDPs include:

- Population register number.
- ID-Card or passport number for IDPs.
- Residence permit number for refugees.
- Tax or social security numbers.
- Operational data: UNHCR asylum seeker and refugee certificates registration number.

178. The following case studies illustrate different approaches to data linking, starting with a population register that identifies refugees / IDPs, compared with using a specific IDP register as a starting point.

CASE STUDY: DATA LINKING BASED ON A POPULATION REGISTER THAT IDENTIFIES REFUGEES IN NORWAY

Statistics on those with a refugee background in Norway are produced from the Central Population Register (CPR). Persons who have come to Norway as refugees or asylum seekers (as well as their family members) are included in the CPR once they have been granted permission to be residents – which can be temporary or permanent. Refugees are given a PIN on arrival, whereas asylum seekers are assigned another ID number, called the D-number, when they apply for asylum. It is only when a person is granted asylum that he or she is assigned a PIN number.

Statistics Norway receives regular updates on vital events and other changes in the CPR and has access to many other administrative and statistical data sources that use the same PIN. This makes it possible to link information from different administrative sources to the CPR, to check for possible errors and other shortcomings, such as missing data, and to generate new data, such as country of birth (if this is not in the CPR). The contents, coverage and quality of the administrative registers have become so good that it is no longer necessary to conduct traditional population and housing censuses. Statistics Norway publishes regular analysis, including on persons with a refugee background and on refugees in the labour market.

COLOMBIA.
Embera indigenous communities of Bogota return to Alto Andágueda in 2023.
©Unit for Comprehensive Care and Reparation for Victims of Colombia (UARIV)

CASE STUDY: USING AN IDP ADMINISTRATIVE REGISTER AS A STARTING POINT FOR DATA LINKING IN COLOMBIA

The official source to count the stock of IDPs in Colombia is the Single Victims' Registry managed by the Victims' Unit, a government agency responsible for providing assistance and reparation to victims of the armed conflict. The Single Victims' Registry has been counting IDPs since 1985. As at October 2019, 7.5 million IDPs were registered. IDP households who wish to be included in the Single Victims' Registry present a declaration before the Public Ministry (Ombudsman's Regional Office, Procurator's Municipal Office). According to the national legislation an IDP must declare within two years of the displacement event. Once an IDP is granted the status of victim, they are included in the Single Victims' Registry. Once registered they are entitled to access relevant assistance and reparation.

An individual assessment of vulnerability is run twice a year for IDPs registered in the Single Victims' Registry. The data used to run the assessment comes from both primary and secondary sources. Primary data are collected through a survey conducted by phone on a continuous basis. Secondary data, which is the main source of data, involves the exchange of official administrative records with other government agencies at both national and local level: for example the official education enrolment system, the social security system, the registers' office, and the housing and agricultural subsidies registers, among others. IDPs update their location directly to the Victims' Unit through services points located throughout the country.

There is a Protocol for Information Exchange between the Victims' Unit and all the government agencies which form part of the National System for the Assistance and Reparation of Victims. The process of information exchange is formalised through a Memorandum of Understanding, in compliance with the relevant legislation. The data themselves are linked through unique identifiers based on the national ID numbers allocated to all citizens.

Various statistics can be calculated out of the Single Victims' Registry: these are disaggregation by sex, age, ethnicity, type of human right violation (displacement, forced disappearance, homicide, threat, among others), geographic location or number of victims per year.

179. In practice, finding sources with UIDs can be very challenging, as most administrative systems evolve to serve a specific purpose without any thought to future data linking. For example, administrative registers may treat whole families as a single unit. In the absence of a suitable UID, it can be possible to build a synthetic UID through data matching, including imputation or probabilistic data matching. These methods can enable insights from combining survey sources with administrative data – see Case Study *Data Matching in Germany*. A key first step in data linking is to review the widest range of possible data sources, including surveys and non-traditional sources, giving consideration to the availability and quality of the sources and defining the possible outputs that could be derived through linking.

180. Estimating the unknown size of a target population when using multiple partially overlapping sources might also be attempted using the capture-recapture method. Application of this method to the refugee / IDP context is included in Annex 2.

CASE STUDY: DATA MATCHING IN GERMANY

Germany uses administrative data on residence permits from the Central Register of Foreigners (CRF) to identify refugees and foreigners in refugee like situations. The CRF does not provide information on their socioeconomic characteristics or housing and living conditions. Germany's Labour Force Survey (LFS) provides a detailed socioeconomic picture of respondents but does not confidently identify refugees and foreigners in refugee-like situations.

Using imputation-based data matching, the information from the administrative data set (donor) is used in order to impute refugee status in the LFS survey data set (recipient). This imputation is based on a set of common variables on citizenship, date of entry, age and gender. The final aim is the analysis of the unobserved joint distributions of refugee status and socioeconomic characteristics.

More details on this Case Study are included in Annex 1.

Access requirements

181. Some countries provide a legal basis for the NSO to access data and metadata for all available administrative data sources for evaluating their potential use in official statistics. If, based on the metadata, the data source is judged to be useful to build new or improve existing official statistics, then test data can be requested to further assess data quality and to produce prototype statistics. For other (non-administrative) data sources the access requirements are different. Table D.2 illustrates the likely access requirements for key data sources that may yield improved refugee / IDP statistics through data linking.

Table D.2: Likely access requirements for different data sources

Sources/ Access requirements	Legal environment	Availability	Sustainability	Statistical requirements	Required resources (cost, IT environment, technical skills)
Administrative registers	MOUs Statistical law	Micro/macro data	Changes in collected data: Consultation of statistical authorities by data owners. Cooperation between stakeholders	Confidentiality Statistical unit Statistical concept vs administrative	Compatibility between IT systems Data modelling
Censuses	Statistical law	Micro/macro data	Depends on statistical authorities	Low Frequency UID	Costly
General population surveys	Statistical law	Micro/macro data	Depends on statistical authorities	Frequency UID Size of the sample of the population of interest	Compatibility between IT stems Data modelling
Specific surveys[37] (Refugees, IDPs)	Refugee law Data protection policy & consent	Micro/macro data		Event driven Coverage	Data modelling
Operational data	Data protection policy & consent	Micro/macro data	Depends on data collector	Event driven Coverage	Data matching
New digital sources (social media, satellite images, mobile phones, GIS)	Memorandum of Understanding (MOUs) Commercial agreement Statistical law	Often unstructured data e.g. Tweets, Google	Problem of market entries and exits of data providers	Problem of biases in usage of digital devices and media; frame for grossing-up required to remove biases (e.g. census data) or other data selectivity models	Data scientists; Problem of competition between NSOs and businesses for skilled employees

Standardizing refugee and IDP identification in administrative systems

182. It would be a significant benefit for refugee / IDP statistics if there were some recommended standards for including refugees and IDPs in administrative systems. EGRISS will continue to work towards standards and examples of good practice for identifying refugees in administrative data systems to include in future updates of this manual.

37 It should be noted that the legal environment will be significantly different depending on whether the specific surveys are managed by the NSI or by other ministries, academic institutions or other Non Government Organisations.

Data quality

183. Administrative data are usually not collected with statistical purposes in mind so it is important to spend some time understanding the data processes to be aware of limitations and potential data quality issues which can affect their use. Data quality can vary significantly between administrative registers and can be as fundamental as differences in the definitions or coverage of refugees and IDPs. Removal of obsolete records is another example of a significant data quality issue common across administrative data. However, in general these data quality concerns are not unique to refugee and IDP data sources and the same processes should be followed as for any other statistical use of administrative data.

184. Statisticians should seek to build and maintain strong links with the organisations that own administrative registers, to raise awareness of the statistical uses of the data and foster collaboration: it may be possible to address data quality concerns or enhance the statistical value of administrative data through small changes to the underlying register.

185. Reviewing data quality after a mass influx of asylum seekers requires special attention. Data quality issues in administrative data occur in particular during periods of mass influx when administrative procedures are overburdened with the large numbers of incoming asylum-seekers. When the increase in asylum applications is sudden and unexpected, it is difficult to scale-up existing administrative capacities for registration and refugee status determination. As a result, data quality in administrative registers is often poorest when the demand for data on refugees is highest. Reviewing and validating the available data becomes a crucial task for statisticians in those times, since they must judge whether data quality is sufficient for the production of official statistics. When using administrative data in immediate response to a mass influx, statisticians should look out for:

- **Delayed registrations:** In response to humanitarian crises, organizing shelter and covering basic needs is a host country's first priority. Refugee status determination comes afterwards. During periods of mass influx, asylum seekers may face prolonged waiting periods for officially lodging their asylum applications. Where asylum seekers are properly registered only after having lodged their asylum application, administrative registers will cover the newly arrived only with considerable time lag.
- **Incomplete and false registrations:** Intending to assist overburdened immigration authorities, governments improvise and deviate from traditional workflows during periods of mass influx. Registration is then often no longer carried out by trained personnel in immigration offices but mobile registration units with the support of the armed forces and the police register asylum seekers in temporary accommodations or even in improvised reception facilities on the road. Considering that untrained staff registers displaced people who often lack official documents, incomplete and false registrations are hardly surprising.
- **Multiple registrations:** When asylum seekers are distributed within the host country for accommodation, small differences in spelling or translations often cause double registration. Language barriers are especially problematic when translations between different alphabets are involved e.g. between the Arabic, Cyrillic and Latin alphabet. While taking and comparing fingerprint scans would solve most of the problem, the infrastructure is not always in place. Once entered in the register, duplicate entries are difficult to identify with certainty and can thus be expected to bias results for a prolonged period.

USE CASE

E

SOURCES OF OPERATIONAL DATA FROM HUMANITARIAN ORGANISATIONS

POLAND.
Refugees from Ukraine queue outside the national stadium in Warsaw to receive a Polish social security number in 2022.
© UNHCR/Maciej Moskwa

EGRISS

Operational data – general background

186. This section provides a brief overview of major data sources maintained by humanitarian organizations, which may be useful to inform and complement government official statistics on refugees and IDPs.

187. Operational data are produced during regular activities of humanitarian agencies and are often intended primarily for internal or inter-agency use. In some contexts, operational data can be the most reliable and timely statistics available on refugees or IDPs, while in others they can be a valuable source for data linking or for comparison and triangulation with official statistics. It is therefore recommended that government statistical producers keep a good overview of the operational data available in their country.

188. Operational data are often not collected with statistical purposes in mind, so there can be no default expectation that operational data meet the standards set for official statistics. Before any use within the realm of official statistics, they should be carefully assessed using a national or international statistical data assessment tool, such as the UN Statistics Quality Assurance Framework[38], focusing on standardisation, coverage, accuracy and confidentiality. Ultimately, the statistical authority in charge (e.g. NSO, line ministry, specialised statistical unit within NSS) will need to determine the usefulness of operational data for official statistical production. If the integration of operational data into government official statistics is deemed appropriate and feasible then data sharing agreements with humanitarian agencies will be needed.

189. The factors to consider around the use of operational data and possible data linking are identical to those for administrative sources in Use Case D and so are not repeated here. This reflects the fact that many operational data sources are administrative registers in nature. The statistical analysis would necessarily be limited to the geographical and population scope of the relevant humanitarian agency, or could contribute to a fuller picture of stocks and flows provided care is taken to understand and address potential overlap between sources.

190. The following tables provide a brief overview of the different types of operational data sources available, and how they can be accessed.

38 United Nations National Quality Assurance Frameworks Manual for Official Statistics.

Table E.1: Primary data sources – Administrative registers of individuals or households

Organisation	Data source	Data content	How to access
United Nations High Commissioner for Refugees (UNHCR)	proGress	In many countries where governments have limited capacity to register refugees, the UNHCR collects administrative data on asylum seekers and refugees on behalf of governments. The data can be a key source of both population data (e.g. aggregate numbers, age, sex and disability disaggregation, location) and socio-economic data (e.g. microdata on income, consumption, skills, health status, economic activity) or used to establish a sampling frame for surveys. UNHCR's Population, Registration and Identity Management Eco-System (PRIMES) contains a data tool (proGres) that uses registration data to generate reports, statistics and trends analysis based on global aggregate data.	There is no public access to UNHCR's registration database. Depending on country-level circumstances, data access may be negotiated through a data sharing agreement, which includes clear data protection safeguards consistent with privacy and data protection principles set out in UNHCR's Data Protection policy. ProGres v4 has the functionality to grant access to partners, including host governments.

Table E.2: Primary data sources – Survey data

Organisation	Data source	Data content	How to access
International Organisation for Migration (IOM)	Displacement Tracking Matrix (DTM), often based on surveys of key informants and/or households.	DTM is a system to track and monitor displacement and population mobility. It is designed to regularly and systematically capture, process and disseminate information to provide a better understanding of the movements and evolving needs of displaced populations, whether on site or en route. The DTM provides mobility tracking (numbers, locations and cross-sectorial needs of observed populations); flow monitoring at key transit points to identify scale and direction of flows and reasons for movement; and surveys to gather specific information through sampling from the population of interest, covering topics such as return intention, displacement solutions, community perception, and other thematic areas.	Data, reports and background information available on DTM's website.
Various humanitarian organizations (including UNHCR, JIPS, etc.)	Various household surveys	Humanitarian organizations carry out a variety of different household surveys.	UNHCR's Microdata Library provides access to microdata from many of the surveys run by the organization and its partners. All listed datasets include comprehensive metadata and supporting documents such as survey questionnaires and analytical reports.

Table E.3: Data Portals for accessing other operational data and secondary analytics

In addition to the online portals already listed in table E.2 for primary data sources (administrative databases and surveys) Table E.3 lists portals that provide additional insights into secondary datasets, aggregate statistics and analytical outputs produced and maintained by humanitarian agencies.

Organisation maintaining the portal	Portal	Data content
UN Office for the Co-ordination of Humanitarian Affairs (OCHA)	Common Operational Datasets (CODs)	The CODs are authoritative reference datasets needed to support operations and decision-making for all actors in a humanitarian response. CODs are 'best available' datasets that ensure consistency and simplify the discovery and exchange of key data. CODs include data on administrative boundaries and the names of places, as well as the best available data on the numbers and (where possible) characteristics of displaced people.
UN Office for the Co-ordination of Humanitarian Affairs (OCHA)	Humanitarian Data Exchange (HDX)	The HDX is a tool for accessing a broad variety of humanitarian datasets. HDX provides access to data from multiple organisations on the numbers and (where possible) characteristics of displaced people, including but not limited to the CODs. It is searchable by location and organisation and includes a visualisation tool. HDX is managed by OCHA's Centre for Humanitarian Data.
United Nations High Commissioner for Refugees (UNHCR)	Refugee Data Finder (RDF)	The refugee data finder website contains information about forcibly displaced populations spanning more than 70 years of statistical activities. It covers displaced populations and their demographics. The database also reflects the different types of solutions for displaced populations such as repatriation or resettlement. It includes data from UNHCR's annual statistical activities, the UN Relief and Works Agency for Palestine Refugees in the Near East (UNRWA) and the Internal Displacement Monitoring Centre.
United Nations High Commissioner for Refugees (UNHCR)	Operational Data Portal	Inter-agency coordination portal showing operational information about selected current emergencies.
World Bank – UNHCR Joint Data Center on Forced Displacement (JDC)	Literature Review Database	The JDC publishes a monthly literature review update, highlighting recent publications, academic scholarship, and thought leadership on issues related to forced displacement which can be found in the Forced Displacement Literature Review Database.
Internal Displacement Monitoring Centre (IDMC)	Global Internal Displacement Database	Aggregates data from multiple sources to publish statistics on stocks and flows of internal displacement. The statistics are based on research, data management, analysis and validation. Where available, it draws on national government data but it also utilises data from many other institutions, in a prioritised list, including local authorities, the UN and other international organisations, civil society organisations, research institutions, national Red Cross and Red Crescent societies, the private sector, media and affected populations. It uses new methods of monitoring displacement, including the use of satellite imagery, natural language processing and machine-learning, to identify new incidents of displacement and triangulate data.
International Organization for Migration (IOM)	Migration Data Portal	The Portal aims to serve as a unique access point to timely, comprehensive migration statistics and reliable information about migration data globally.

USE CASE

F

NON-TRADITIONAL DATA SOURCES

SOUTH SUDAN.
South Sudanese men fleeing Sudan violence try to find a phone signal at the Joda border point near Renk in 2023.
©UNHCR/Andrew McConnell

EGRISS

Non-traditional data sources – general background

191. Non-traditional data sources encompass the huge volumes of data generated by new technologies such as mobile phones and social media. These provide a rich potential source to complement or aid the production of government official statistics, with the benefit that they are often more timely and detailed than traditional statistics. The UN has established a committee of experts on big data and data science for official statistics[39] that advocates for the use of non-traditional data sources to complement and improve national official statistics. The UN Statistics Division has stated that "big data constitute a source of information that cannot be ignored and that the global statistical community must organize itself and take urgent action to exploit the possibilities and harness the challenges effectively"[40].

192. The potential value of non-traditional data sources in relation to displaced people is significant, for example to fill gaps in understanding around the number and location of displaced people in rapidly evolving situations or to provide information that can improve sampling frames for surveys. However, it is important to be aware of the limitations and caveats that apply across most of these sources.

- The data are not produced for statistical purposes so there should be no expectation that data quality is sufficient for the production of official statistics. As with the use of operational data, the statistical authority in charge (e.g. NSO, line ministry, specialised statistical unit within NSS) will need to determine the usefulness of non-traditional data sources within official statistics production processes, understand where the data come from and assess their quality.
- Non-traditional data sources are often owned by private companies and legal use of the data will require negotiation of data sharing agreements, consideration of data protection and may potentially require a payment.
- Many of the sources rely on individuals interacting with technology, for example through a mobile phone, use of social media or the internet. This creates an inherent bias, particularly in relation to age (older age groups and the very young are unlikely to be represented) and access (excludes remote rural locations with limited connectivity) but also potentially in relation to gender, disability, income or other characteristics.
- Identification of refugees and IDPs through technology is not equivalent to identification following a set of survey questions, for example, and usually requires assumptions to be made. There can be important definitional differences which need to be considered, particularly if using non-traditional sources as a source of statistics for comparison or to augment survey results: the two may not be covering the same population, which will make comparisons invalid or problematic.

193. Provided these challenges can be overcome, the sources can hold a lot of promise, particularly in terms of complementing or supporting the production of traditional sources, if not yet as a replacement for them. This section focusses on providing a summary of the non-traditional data sources, what each is useful for, how to arrange access to the data, and provides case studies to illustrate and inspire.

---

39 United Nations. UNBigData

40 Big data and modernization of statistical systems

Mobile phone call detail records

194. *What is it?* When mobile phones are used to make or receive calls (or texts) the telecommunications provider that handles the call generates a record that includes many attributes of the call, such as the time, duration, the phone numbers making and receiving the call, and location details. This generates huge volumes of data, known as Call Detail Records (CDR) and owned by the telecommunications companies.

195. *What is it useful for?* CDRs can provide up to date information about sudden changes in population density in small areas, informing understanding of the scale of displacements and potentially the size of new refugee or IDP settlements. This can help, both directly with estimating stocks and flows of displaced people but also to support sampling and enumeration in a survey or census. With more advanced analysis CDRs can support understanding of refugee integration through analysis of call location patterns and calls within or outside the refugee community.

196. *How to access it?* CDRs are proprietary to the carrier network and cannot be accessed unless a bilateral agreement between the analyst and the carrier network is in place that addresses privacy concerns. This may involve limited release of a restricted data sample, sharing of non-identifiable aggregated or anonymized data, or remote access to anonymized data on a virtual environment controlled by the mobile phone operator. If the population movements of interest cross borders, the number of agreements needed – and the complexity of the challenge – rises and it can be more practical to look for other sources, such as internet-based social media records. The UN Global Working Group on Big Data for Official Statistics has published a *Handbook on the Use of Mobile Phone Data for Official Statistics*[41].

41 Handbook on the Use of Mobile Phone Data for Official Statistics

CASE STUDY: POPULATION DENSITY MAPS IN NEPAL

In April 2015, a devastating earthquake with a magnitude of 7.8 struck Nepal. Following the earthquake, Flowminder was given access to anonymised mobile phone data by Ncell, the largest mobile operator in Nepal. Flowminder accessed the data directly at the operator`s premises using it to produce static population density maps after the earthquake, which were used by UN OCHA and other key relief agencies to estimate the number of people displaced and affected, and coordinate disaster relief operations. The project mapping team estimated that around 1.8 million people above normal levels had left their home districts because of the 2015 earthquake.

Source: Flowminder (2015). Nepal Earthquake 2015 Case Study. Retrieved from https://web.flowminder.org/case-studies/nepal-earthquake-2015.

CASE STUDY: SEGREGATION OF SYRIAN REFUGEES IN TURKEY

Researchers have used CDRs to analyse patterns of spatial segregation of Syrian refugees in Turkey, and how patterns of spatial segregation influence internal mobility decisions of refugees as they move to other regions within the country. This enables the construction of two indices of integration: a dissimilarity index measuring the share of refugees that would have to move from high to low concentration regions to match their average distribution across the country, i.e. to achieve full integration; and a normalized isolation index measuring the probability that refugees interact with the wider population. The analysis is based on anonymized CDRs from Turk Telekom for their Syrian customers and for a large sample of Turkish customers. The database includes phone activity for a sample of nearly one million customers, out of which approximately 185,000 are tagged as refugees. The granularity of the data permits the analysis of calls made/received at a geographically disaggregated level (i.e. for each cell phone tower) for each hour in 2017.

Source: Segregation and internal mobility of Syrian refugees in Turkey: Evidence from mobile phone data. Retrieved from https://www.jointdatacenter.org/literature_review/segregation-of-syrian-refugees-in-turkey-evidence-from-mobile-phone-data/.

CASE STUDY: LEVERAGING BEHAVIOURAL AND HUMANITARIAN DATA SOURCES TO ANALYSE DEVELOPMENT CHALLENGES FACED BY SYRIAN REFUGEES AND HOST COMMUNITIES IN LEBANON

UN ESCWA, in partnership with the Qatar Computing Institute (QCRI) and the Data-Pop Alliance, explored the potential of mobile phone data to generate timely, granular, and cost-effective estimates for the vulnerabilities faced by refugees and host communities in Lebanon. CDRs from two mobile operators, Alfa and Touch, were analyzed through the Ministry of Telecommunications. Data from Touch consisted of the number of outgoing and incoming calls, disaggregated by age and gender; data from Alfa contained statistics about data consumption. The spatial distribution of calls proved to be a good indicator of population distribution, correlating closely with traditional official statistics. The ratio of dial-out-duration to receiving-in-duration in Touch data turned out to be a good predictor of socioeconomic factors, through correlation with the number of Syrian refugee families with debt greater than USD $600 and correlation with the number of self-declared poor in the Lebanese host community.

Source: Leveraging Behavioral and Humanitarian Data Sources to Analyze Development Challenges Faced by Syrian Refugees and Host Communities in Lebanon. Retrieved from https://archive.unescwa.org/publications/big-data-challenges-syrian-refugees-lebanon

Mobile phone GPS data

197. *What is it?* Most smartphones collect Global Positioning System (GPS) data, which is used by many apps to track the phone's location and tailor the service on offer accordingly. GPS data differs from location data available through CDRs: the latter rely on identifying which mobile phone tower / receiver has been used by the phone, so is less precise and the data are owned by the telecommunications provider. By contrast, GPS data uses satellites and is more accurate. The data tend to be collected (and therefore owned) either by the phone system manufacturer (for example, Apple) or by the companies that own the apps that use the GPS data. Users typically allow GPS location data to be collected by responding to notifications on their devices that ask whether they are willing to share their location data. The way this happens varies across apps and across phone brands.

198. *What is it useful for?* GPS data provides detailed accounts of movement patterns, for example to understand daily commutes; it can be used to ensure survey enumeration is occurring in the correct location, and the inclusion of GPS technology in smart phones and tablet computers can integrate navigation and the recording of geographical coordinates into enumeration activities, minimising locational error and human effort.

199. *How to access it?* Third party brokers can provide access to data at a cost. There are few examples of the use of GPS data that are directly relevant to producing statistics on displaced populations, but the possibility is included here for completeness. There are literature reviews available that illustrate how mobile phone data (both CDR and GPS) have been used for statistics, for example from the US Census Bureau[42] and the UK Office for National Statistics[43].

Satellite imagery

200. *What is it?* Satellites are used by governments and businesses to take detailed images of the earth's surface, which can build up a regularly updated photographic record of geographical features.

201. *What is it useful for?* Satellite imagery can be used to generate detailed geographical breakdowns, for example to show the locations of displaced communities, either to inform estimates of stocks or to help construct sampling frames. They can also provide geographic context such as proximity to nearby services, schools and hospitals.

202. *How to access it?* Access to some satellite imagery can be free, while many sources are available at a cost.

CASE STUDY: GENERATING ENUMERATION BLOCKS FROM SATELLITE IMAGERY FOR A SURVEY IN SOMALIA

In Somalia where security considerations prevented the usual sampling techniques to be carried out an innovative technique was used. It was not feasible to conduct a full listing of all households in an enumeration area, as this was too time-intensive and raised security concerns. Instead, enumeration areas were segmented into smaller enumeration blocks using satellite imagery. Enumeration blocks are small enough for enumerators to list and select households immediately before conducting the interview.

Source: Utz Pape and Philip Wollburg. "Estimation of Poverty in Somalia Using Innovative Methodologies." World Bank Policy Research Working Paper 8735, February 2019 https://documents1.worldbank.org/curated/en/509221549985694077/pdf/WPS8735.pdf

42 Use of Mobile Phone Location Data in Official Statistics, Social, Demographic and Health Studies

43 ONS methodology working paper series no. 8- Statistical uses for mobile phone data: literature review

CASE STUDY: USING SATELLITE IMAGERY TO ESTIMATE SIZE OF ENUMERATION AREAS IN SOUTH SUDAN IDP SURVEY

The Crisis Recovery Survey was conducted in 4 IDP camps in South Sudan between May to July 2017. The sample was restricted to Protection of Civilian (PoC) camps, and includes the 4 largest camps with clearly defined boundaries. The sample was designed as a multi-stage stratified random sample. Each camp was selected as a strata, with a target of 600 interviews per camp. Within each camp, 50 enumeration areas (EAs) were selected proportional to size, where the size was defined by the number of structures in the EA. The number of structures was estimated using satellite imagery of the strata (camps). Each EA was divided into 12 blocks, and a micro listing was done in the blocks to randomly select households.

Source: https://microdata.worldbank.org/index.php/catalog/2914

Social media data

203. *What is it?* Social media generates huge volumes of data, through the content of material that individuals post on social media, such as Facebook and Twitter, as well the locations and times of posts and characteristics of the individual posting.

204. *What is it useful for?* All the data has potential analytical value but applying them to inform refugee and IDP statistics is still new and has mostly happened in a research context rather than generating a set of readily repeatable analytical techniques. For example, sentiment analysis of social media posts (through analysing the text to understand positive or negative sentiments) can be used to support analysis of refugee integration in host communities, while geo-tagged tweets and Facebook data have potential to inform understanding of flows of IDPs and refugees. This can be particularly valuable in the context of cross-border flows, where mobile phone data are of limited value as mobile phone SIM cards are linked to national providers, so human mobility calculated from phone records can only be used to estimate movements within countries.

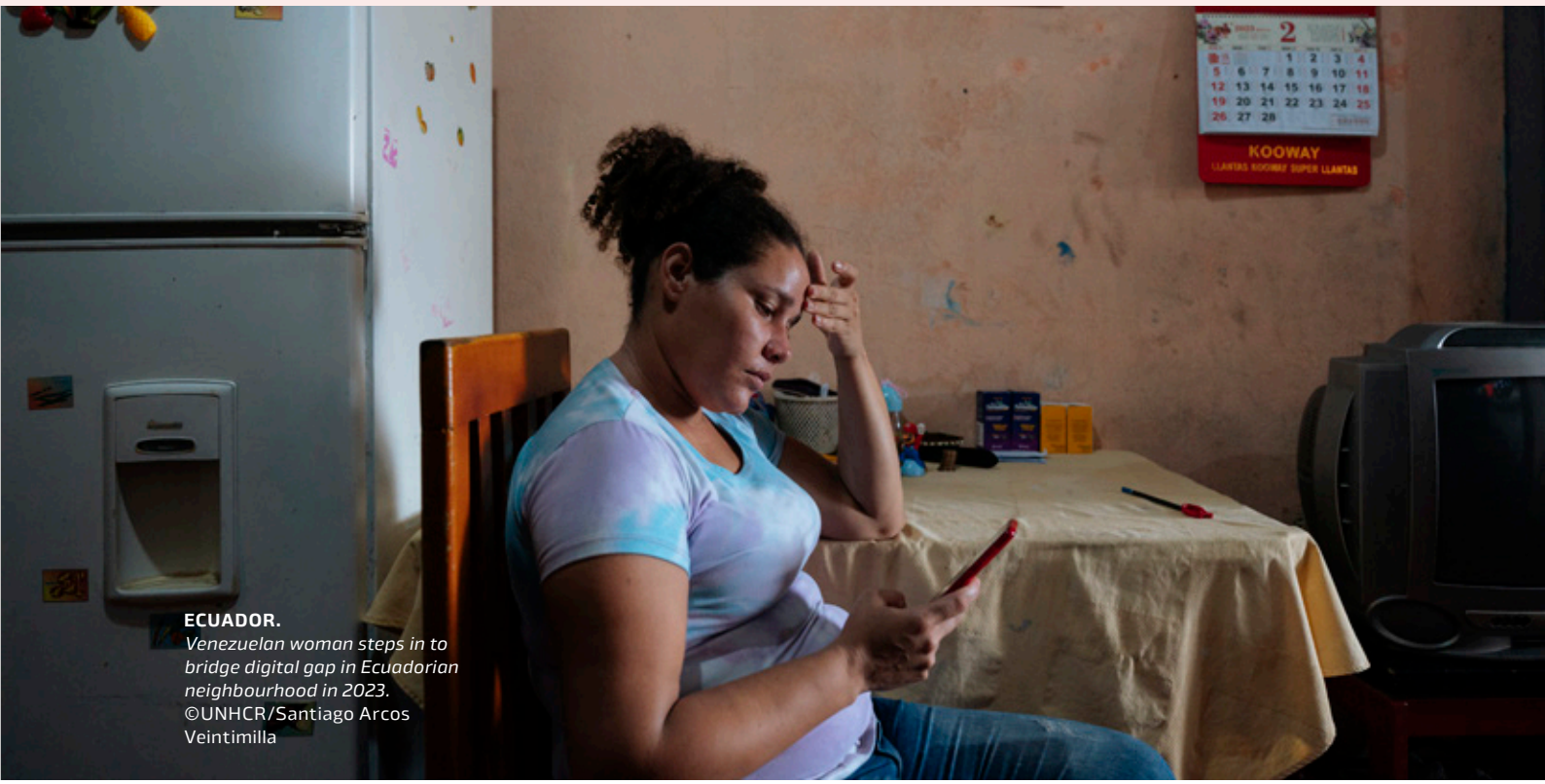
205. Currently there are few examples of social media data providing useful analysis to support official statistics on refugees and IDPs, but there is research that shows the potential, as illustrated in the following case studies. Research focussed on this topic mainly uses Facebook's advertising platform or covers sentiment analysis. The main limitation of using social media sentiment analysis is that the data can be biased by automated or spam tweets: the extent of social media posts published by automated "bots" is unknown.

CASE STUDY: MONITORING OF THE VENEZUELAN EXODUS THROUGH FACEBOOK'S ADVERTISING PLATFORM

Venezuela is going through the worst economical, political and social crisis in its modern history. Basic products like food or medicine are scarce and hyperinflation is combined with economic depression. This situation is creating an unprecedented refugee and migrant crisis in the region. Governments and international agencies have not been able to consistently leverage reliable information using traditional methods. Therefore, to organize and deploy any kind of humanitarian response, it is crucial to evaluate new methodologies to measure the number and location of Venezuelan refugees and migrants across Latin America. Research has proposed using Facebook's advertising platform as an additional data source for monitoring the ongoing crisis. National and sub-national numbers of refugees and migrants have been estimated and validated and socio-economic profiles have been disaggregated to further understand the complexity of the phenomenon. Although limitations exist, the presented methodology may be of value for real-time assessment of refugee and migrant crises world-wide. Similar techniques have been deployed in monitoring displacement caused by the war in Ukraine (see second link).

Source: Monitoring of the Venezuelan exodus through Facebook's advertising platform

Source: Nowcasting daily population displacement in Ukraine through social media advertising data



ECUADOR.
Venezuelan woman steps in to bridge digital gap in Ecuadorian neighbourhood in 2023.
©UNHCR/Santiago Arcos Veintimilla

CASE STUDY: REFUGEE CAMP POPULATION ESTIMATES USING AUTOMATED FEATURE EXTRACTION IN BANGLADESH

High-resolution satellite imagery can be used to map (manually or through automated feature extraction) physical structures in refugee and IDP camps, including changes to the number and type of these structures over time, to support population estimates and geospatial analysis. This technique was applied in a study during the Rohingya refugee crisis, focusing on areas in and around existing refugee communities in two main refugee settlements in Bangladesh. Population estimates for each of the refugee camps were determined by: (a) identifying building features; and then using these features to (b) estimate the camp population based on the total area of the building features and UNHCR 'covered area per person' statistics. Automated feature extraction greatly reduced the average processing time for each camp. However, the accuracy of automated feature extraction methods rely on well-defined classifier definition files (used to classify pixels or group of pixels into different roof types and non-building features based on their spectral, textual or spatial properties) and it is not straightforward to establish well-defined classifier definition files that are geographically and temporally transferable.

Source: Refugee Camp Population Estimates Using Automated Feature Extraction

BANGLADESH. *In 2022, a fire devastates Rohingya refugee settlements in Kutupalong.* ©UNHCR/Amos Halde

CASE STUDY: A FRAMEWORK FOR ESTIMATING MIGRANT STOCKS USING DIGITAL TRACES AND SURVEY DATA: AN APPLICATION IN THE UNITED KINGDOM

An accurate estimation of international migration is hampered by a lack of timely and comprehensive data, and by the use of different definitions and measures of migration in different countries. In an effort to address this situation, traditional data sources for the United Kingdom have been complemented with social media data to understand whether information from digital traces can help measure international migration. The Bayesian framework proposed is used to combine data from the Labour Force Survey (LFS) and the Facebook Advertising Platform to study the number of European migrants in the United Kingdom, with the aim of producing more accurate estimates of the numbers of European migrants. The overarching model is divided into a Theory-Based Model of migration and a Measurement Error Model. The quality of the LFS and Facebook data has been reviewed, paying particular attention to the biases of these sources. The results indicate visible yet uncertain differences between model estimates using the Bayesian framework and individual sources. Sensitivity analysis techniques are used to evaluate the quality of the model. The advantages and limitations of this approach, which can be applied in other contexts, are discussed. No one individual source can necessarily be trusted but combining them through modelling offers valuable insights.

Source: A Framework for Estimating Migrant Stocks Using Digital Traces and Survey Data: An Application in the United Kingdom

OpenStreetMap

206. *What is it?* OpenStreetMap (OSM) is a collaborative project to create a free editable geographic database of the world. It is built through crowdsourced geographic information and in some locations its detail and precision rival "authoritative" datasets from governments and commercial entities. The data from OSM is freely available.

207. *What is it useful for?* As a mapping tool, OSM will potentially be a useful geographic resource to aid sampling and enumeration and potentially as a source of data to feed into other estimates. Ultimately, OSM's usefulness depends on the extent to which it is complete and regularly updated in the locations of interest.

208. *How to access it?* OSM is freely available through its website.

CASE STUDY: EVALUATING THE UTILITY OF OPENSTREETMAP DATA FOR MONITORING SUSTAINABLE DEVELOPMENT GOAL PROGRESS IN REFUGEE SETTLEMENTS IN UGANDA

All available OSM data within 28 refugee settlements and 26 non-refugee settlements in Uganda was collected. The data represents physical features associated with dwellings, schools, clinics, latrines, etc., with metadata on feature creation date, date of most recent edit (version), and descriptive tags. The study created a data model linking 149 OSM features to 11 SDGs. Based on these SDG-OSM pairings, the study: (1) quantified the spatial distribution of SDG-relevant OSM data across and within settlements; (2) measured the chronology of creation and subsequent versions of SDG data, and (3) compared the spatial and temporal coverage of SDG data between refugee and non-refugee settlements. Despite many limitations of this approach, the study concludes that the widespread availability of OSM data make it a promising source of information on SDGs in refugee settlements, as well as in peri-urban informal settlements and internally displaced person (IDP) camps.

Source: Development after Displacement: Evaluating the Utility of OpenStreetMap Data for Monitoring Sustainable Development Goal Progress in Refugee Settlements

UGANDA.
Landscape of the Benet Community in Chemukula village in 2021.
©UNHCR/Esther Ruth Mbabazi

USE CASE

G

COORDINATING AND PLANNING REFUGEE AND IDP STATISTICS IN NATIONAL STATISTICAL SYSTEMS

SOUTH AFRICA.
2021 National Migration and Urbanisation Forum.
©Statistics South Africa. (Stats SA)

EGRISS

Coordinating and planning refugee and IDP statistics – general background

209. This manual, through the preceding Use Cases, provides guidance for producing statistics on refugees and IDPs within the National Statistical System of a country, and also provides some pointers for potential synergies between "traditional" data sources in official statistics and data sources from humanitarian agencies as well as non-traditional data sources. While each individual source is valuable, greater value can be derived from taking a holistic approach to refugee and IDP statistics across the national statistical system, and coordinating across government and more widely to ensure that the needs of users are understood and met in the fullest way possible.

Anchoring coordination in relevant statistical planning processes

210. In essence, coordinating refugee and IDP statistics is no different from coordinating on any other area of official statistics. It should be conceptualized and anchored within the relevant statistical planning processes in each country, such as sectoral statistical strategies[44] and/or the National Strategy for Development of Statistics (NSDS),[45] both of which provide frameworks for assessing data needs, structuring support, enhancing coordination and developing statistical capacity in the national statistical system.

211. In practice to date, refugees and IDPs have tended not to be thought of as a statistical sector in their own right, so can be overlooked in such planning processes. Whether a sectoral statistical strategy or an NSDS should be the main tool for coordinating and enhancing official statistics on refugees and IDPs, will depend on circumstances in a given country. In some cases, a combination of both planning tools (or a sequenced approach) may be appropriate, provided overlap can be avoided.

Coordinating refugee and IDP statistics via an NSDS

212. Planning the coordination and enhancement of displacement statistics via a chapter within the NSDS has various advantages over a sectoral statistical strategy. Firstly, one can capitalize on the buy-in and broad-based consultative nature of the wider NSDS process. Secondly, the comprehensive nature of an NSDS does justice to the often cross-cutting nature of refugee and IDP statistical production, which in a national statistical system may draw heavily on broader statistical products such as population censuses and nationwide surveys. An NSDS can be a strong catalyst of national ownership in statistical production, and secure funding at national and international level.

213. However, a key drawback of coordinating and planning refugee and IDP statistics via an NSDS is that not all countries have a current NSDS, and even if they do the opportunities for introducing a new sector into the national statistical planning process only occur in rare intervals. The mid-term evaluation or a new phase of the NSDS can be the right moment for integrating a new sector related to refugee and IDP statistics. Integration of refugee and IDP statistics into the NSDS should occur at all stages of the design and implementation phases.

---

44 Sectoral strategies for the development of statistics

45 Guidelines for National Strategy for the Development of Statistics

Coordinating refugee and IDP statistics via a sectoral statistical strategy

214. Planning the coordination and enhancement of refugee and IDP statistics via a sectoral statistical strategy, either as a sector in their own right or as part of another sector (e.g. migration statistics), also has its advantages. A sectoral statistical strategy may be the instrument of choice where an NSDS is not available or not currently under review, or where the production of refugee and IDP statistics occurs in a smaller set of entities within the National Statistical System that do not usually partake very actively in the wider NSDS process. The stand-alone nature of a sectoral statistical strategy may also be more appropriate where the relevant leads in the National Statistical System wish to engage comprehensively with government-external stakeholders that may not routinely engage in the wider NSDS process (e.g. humanitarian organizations).

Key steps in coordinating and enhancing refugee and IDP statistics

215. Irrespective of whether the coordination and enhancement of refugee and IDP statistics is anchored and conceptualized in a sectoral statistical strategy or an NSDS, or both, the steps of the planning process are similar. The following discussion offers some pointers that are specific to refugee and IDP statistics along these steps but doesn't intend to provide a comprehensive manual for statistical strategy development[46].

Identifying a forum for key stakeholders to discuss displacement statistics

216. A first stage in coordinating refugee and / or IDP statistics is to establish a forum where displacement statistics can be discussed and prioritised actions agreed. While official statistics should be the primary focus in this forum, and its governance and leadership should be government-led, relevant government-external data producers may also be invited to join the forum.

217. In some countries, it will be possible to expand an established forum to include displacement statistics. This may be done through expansion of the scope of a forum for discussing migration statistics, where one exists. If this is the case, the terms of reference of the group can be expanded to include displacement statistics and it is likely to be necessary and appropriate to invite a few new members to the group, to represent the wider displacement statistics community.

218. If there is no existing group that is suitable to cover displacement statistics then it may be necessary to establish a new statistics committee. Sector Statistics Committees (SSC) are responsible for defining and confirming the statistical needs for sectors. A Refugee and IDP Statistics Committee should be constituted by those leading the sector at the request of the NSO, and a sector Coordinator should be designated. All relevant departments, ministries and agencies could be involved, including migration authorities, social development ministries, urban planning authorities, rural development authorities, and the humanitarian sector. Experience shows that five active members are the ideal number for the committee, but the Sector Statistics Committee should involve actors

46 Guidelines for National Strategy for the Development of Statistics

from the wider data ecosystem and goes beyond the traditional boundaries of the NSS, to humanitarian agencies including UNHCR and UN OCHA.

219. The responsibilities of the discussion group, whether an existing forum or new committee, would be similar to those of other statistical sectors, including:

- proposing a vision and strategic objectives for refugee and IDP statistics;
- identifying entities currently collecting relevant statistics for refugees and IDPs;
- preparing an inventory of existing current data systems;
- identifying major data needs related to the sector;
- foster institutional partnerships in the statistical sector
- develop and prioritise action plans, including technical but also organisational (HR development, IT strategy) and communications-related actions.

Understanding users' needs

220. There are many potential users of statistics on displaced persons, including governments and national authorities, local authorities, international organisations, civil society, NGOs, researchers/academia, the media and the general public. They have different needs and priorities, requiring statistics for different purposes:

- for administrative purposes,
- for policymaking,
- for monitoring the implementation of policies,
- for evaluating policies,
- to facilitate the allocation of resources,
- to enable regional/international comparison,
- to enable comparisons with the wider population,
- to assess progress towards durable solutions or integration with host communities,
- to inform the general public and enhance public debate.

221. It is important to identify the key groups of users to engage in the relevant country context, whose needs will drive the development of these statistics. Some of these groups will already be accessible through existing statistical networks and activities, but others may be new – such as displaced population support groups. Identifying important "needs gaps" requires research, but wider needs assessments undertaken by or on behalf of operational agencies supporting refugees and IDPs can be a rich source of information on the data needs of those supporting the displaced.

222. PARIS21 has developed the Advanced Data Planning Tool[47] (ADAPT) for NSOs to adapt their data production processes to meet the data needs of users. It is a consultative tool, which brings data stakeholders together to identify data gaps, and define requirements within an established monitoring indicator framework such as the UN Sustainable Development Goals (SDGs) indicator framework or the basic statistics proposed by the IRIS and IRRS (see also the Sustainable Development Goal indicators website.)

47 Paris21 Advanced Data Planning Tool

223. The results of ADAPT can contribute to the development of statistical plans and processes to strengthen the coordination of statistical planning at the national level. Based on the information provided, the ADAPT tool can produce two main reports on the status of refugee and IDP statistics and identify any data gaps at national level. The first report is able to summarise the status of refugee and IDP statistics in respect of the relevant national requirements, which could be based on the IRRS or IRIS tabulation indicator frameworks, or a national defined requirement. The second report presents the status of refugee and IDP statistics as a general domain of statistical production, considering all indicators that require a relevant disaggregation variable.

224. National collaboration platforms, for example targeted workshops or thematic conferences, and regional fora can be an opportunity to engage with the identified strategic stakeholders and promote the need for refugee and IDP statistics. Connecting the topics with policy priorities at international and national levels can attract the attendance of high-level political representatives and help to gain government support.

225. The tools developed as part of the Grand Bargain[48] work stream on needs assessments provides a useful framework for considering the roles involved in developing useful and useable statistics and includes question prompts for understanding user needs[49]. This is included in Annex 3.

Identifying existing and potential data sources

226. It is unlikely that any individual team or organization has a complete understanding of data production on refugees and IDPs, either within the NSS or more widely across international partners. As such, a comprehensive mapping of the existing and potential sources that might contribute data on displaced populations, and their strengths and weaknesses, is essential. It will require input from a range of groups and can be one of the first key tasks of the discussion forum. It may also suggest previously unidentified organisations that should be invited to join the discussion forum.

227. As part of this exercise, it is important to establish what relevant questions are already asked in official statistics surveys; what data on displaced populations could be available from administrative sources, operational data and new technologies; and where possible (although not essential at this stage) the legal basis for sharing any of these data. Use Cases D, E and F in this manual can support this exercise, providing information on possible sources of administrative data, operational data and non-traditional data respectively.

228. Table G.1 provides a very high-level guide to the different categories of data source that can be used to produce statistics on displaced populations, the key things to consider around each source and which Use Case to refer to for further information. Ideally this manual would provide comprehensive guidance on which source is optimal for producing different types of statistics but in practice this varies depending on the country context and the existing statistical landscape.

---

48 The Grand Bargain was launched during the World Humanitarian Summit in Istanbul in May 2016 and is an agreement between some of the largest donors and humanitarian organisations who have committed to improve the effectiveness and efficiency of humanitarian action.

49 The tools are available on the website for Workstream 5

Table G.1: Guide to the different categories of data source that can be used to produce statistics on displaced populations

Source	Use Case	What the source can provide	Key considerations
Census	A	Accurate statistics on stocks of refugees and IDPs, analysis of their characteristics, some socioeconomic indicators and comparisons with the wider population. Census data can help understand flows, through collecting data about a person's previous locations.	- Provides the benchmark of the whole resident population so it is important that displaced people are included in census data collection. - The infrequency of censuses and the time lag before data become available can impact on the relevance of the data but can be used to compare with and adjust other more frequent data sources. - Good coverage of displaced people in the census increases its value as a sampling frame for subsequent household surveys. - Competing priorities for space in the census questionnaire can make it challenging to include all the relevant questions necessary for accurate identification of displaced populations
Existing general household survey	B, C	Important data source for measuring the characteristics and living conditions of displaced populations and socioeconomic indicators. Unlikely to be a primary means of measuring stocks, unless good coverage can be achieved, but may be suited as a source to validate or update stock estimates from other sources. Will not capture flows effectively. Examples include labour force surveys, demographic and health surveys, multiple indicator cluster surveys, integrated living conditions surveys, etc.	- Relatively low cost to include displaced populations in pre-existing statistical landscape. - Captures equivalent measures for displaced population, ensuring comparability with other groups in the wider population. - Need to ensure the sample frame provides good coverage of displaced populations. - Can be logistically challenging to adapt existing surveys to reach displaced populations effectively. - Displaced populations are usually a small proportion of overall population, so oversampling of the displaced population may be required to support the desired level of analysis.
Specific refugee or IDP survey	B, C	Detailed coverage of relevant socio-economic indicators to assess needs, integration and progress. potential to collect broader and deeper information	A dedicated survey provides the opportunity to collect broader and deeper information, capturing topics that meet users' needs and are most relevant to displaced populations. Survey can be designed to meet the challenges of identifying and reaching displaced populations, ensuring accurate coverage. Relatively high costs associated with running a dedicated survey, especially if repeating the survey to monitor changes over time.

Government administrative data	D	Potential source of information for measuring both the stock of displaced populations and the flows, showing the direction of displacement trends, and potentially for developing sampling frames. Can contribute to socioeconomic indicators, for example through employment, education and health registers.	- Administrative data can be more timely than other sources. - Data linking (across administrative sources or with surveys or census) increases the potential value of administrative data, enabling them to become part of a rich source of regular official statistics. - Administrative registers may not include displaced populations (particularly refugees) or will often fail to identify them effectively (including IDPs, unless there are IDP-specific registers) without adjustments to the registers, so good dialogue with data owners is essential. - Usually have to use existing systems that have not been designed with statistics in mind, so quality can be a concern. Setting up new systems is expensive. - Unlikely to provide indicators of needs or progress towards durable solutions.
Operational data	E	Stocks and flows and potentially socioeconomic indicators such as employment, education, health.	- In some contexts, operational data can be the most reliable and timely statistics available on refugees or IDPs, while in others they can be a valuable source for data linking or for comparison and triangulation with official statistics. - Operational data are produced during regular activities of humanitarian agencies and are often intended primarily for internal or inter-agency use. Not designed for use as official statistics, so data quality needs to be carefully evaluated and may mean that the sources are not suitable to be used.
Non-traditional data sources	F	Indicators that support understanding of stocks or flows or that aid effective sampling and enumeration for surveys.	- Can fill gaps in understanding and provide creative solutions to statistical challenges, for example to help understand rapid changes after large disruptive events. - Unlikely to be a source of regular, accurate official statistics in their own right. - Data quality and completeness can be unknown. - Data are often owned by private companies and their use depends on negotiating access, potentially with a cost attached.

Developing and prioritizing action plans to improve data availability

229. The production of refugee and IDP statistics often requires the combination of data from several data producers within the national statistical system, including line ministries, government departments and agencies, and civil society. National statistical systems differ in their level of centralization and the effort needed to foster consultation and collaboration between statistical producers varies accordingly but it is important to that the statistical programme of work is developed jointly by the statistical producers.

230. The Use Cases in this manual provide guidance on improving the availability of statistics on displaced populations through census, surveys, administrative data, operational data from international organisations and non-traditional data sources. Other points to note include:

- Identify opportunities to co-ordinate and harmonise the approach to refugees and IDPs across existing NSO-led household surveys, and census.
- Establish the legal framework for accessing other data sources and where appropriate engage owners of data sources that are potentially useful (administrative / operational / academic / other data) – for data access, and potentially to propose improvements to make the sources more useful in future.
- Consider partnerships with non-traditional data sources such as data science firms, telecommunication companies, and other service providers to share data and/or help develop applications to extract data.
- Develop a business case, if required, for new data sources.

Disseminating statistics and seeking feedback

231. The key objective of coordinating refugee and IDP statistics is to improve the availability of statistics to meet user needs – and this can only be achieved fully through the dissemination of statistics. As set out in the Use Cases in this manual, it is important to ensure that displaced populations are represented in standard disaggregations, produce thematic reports and visualisations and engage users to seek feedback.

232. It is also worth noting that dissemination involves both publishing statistics and the promotion of the statistical products to help them reach the widest possible audience, for example though social media and attending relevant conferences and public debates. There are many sources of information[50] to support effective dissemination of statistics.

233. It is important to consider how statistics can be shared with the refugee communities themselves. This feedback loop is significant: as well as meeting the needs of a key user community it demonstrates what the survey is for and encourages future engagement and participation.

50 For example, Eurostat Social Media Guidelines and UNECE Recommendations for Promoting, Measuring and Communicating the Value of Official Statistics

CASE STUDY: COORDINATION OF IDP STATISTICS IN SOMALIA

A combination of conflict and natural disasters has displaced over 2.4 million Somalis, including 549,000 in 2021 alone, exacerbating the impact of poverty. The government published a four-year national durable solutions strategy in 2021, intended to address the causes and impacts of the country's internal displacement crisis. To support this agenda, Somalia is progressing towards increasing its capacities on national statistics and analysis on displacement.

The Somalia National Bureau of Statistics (SNBS) is leading development of a National Strategy for the Development of Statistics. In 2021 the SNBS requested support from EGRISS to improve the availability and robustness of data on IDPs. The aim is to increase the visibility of internal displacement and contribute to evidence-based public policy.

JIPS has been providing technical support on behalf of EGRISS. A first step was to conduct a mapping exercise of existing relevant data sources on IDPs in Somalia, both from government and humanitarian sources.

In parallel, in 2022 a workshop was held that brought together different actors from SNBS, other government entities and international actors working with IDP data in Somalia. The aims of the meeting were to develop a shared understanding of the requirements IRIS sets out for the production of official IDP statistics; continue mapping the existing data ecosystem on internal displacement in Somalia; and to discuss how the international community can best support government-owned statistics on internal displacement.

A key outcome of the meeting and the data mapping was a common understanding of the limitations of operational data sources for official statistics and the need to integrate IDP identification in national household surveys run by the SNBS.

Based on this common understanding an IDP Statistics Sector Plan has been drafted. National policy documents such as the National Development Plan and the National Durable Solutions Strategy have been used as guides, as well as international frameworks such as the Sustainable Development Goals. The plan covers coordination, legal and statistical frameworks, capacity building and a costed implementation plan. Key objectives of the plan will be the establishment of an IDP sampling frame and integration of IDP identification in major household surveys. This will facilitate regular production of both estimates of the number of IDPs and regular socio-economic statistics on IDPs.

JIPS on behalf of EGRISS will continue leading on building capacity, including coordinating a workshop with Government actors to raise awareness around the IRIS and identify opportunities for potential government-led data sources to support official statistics; and to continue technical meetings with the SNBS to provide technical support on the upcoming data collection exercises.

ANNEXES

Annex 1: Administrative data sources, detailed case studies

CASE STUDY: PRODUCING REFUGEE STATISTICS FROM A POPULATION REGISTER AND LINKS TO OTHER ADMINISTRATIVE REGISTERS IN NORWAY

Refugee statistics in Norway are produced from the Central Population Register (CPR). The CPR of Norway was established in 1964 based on the 1960 Population Census. A unique 11-digit personal identification number (PIN) was introduced at the same time. The CPR includes all persons who has ever been a (legal) resident of Norway since 1960, regardless of their citizenship. It provides information on their name, address, residence status, place of birth, citizenship, country of immigration or emigration, marital status and PIN numbers for spouse, mother and father. Persons who die or emigrate are not deleted from the register, but a code for their status is changed. All vital events and migrations and address changes are registered in the CPR.

Data on asylum seekers is transferred to the CPR from the Norwegian Directorate of Immigration (UDI). All asylum seekers must register with the Norwegian immigration authorities, and their case data and personal data are stored in a database that contains data on those with an existing refugee status as well as those who enter Norway on their own and apply for asylum.

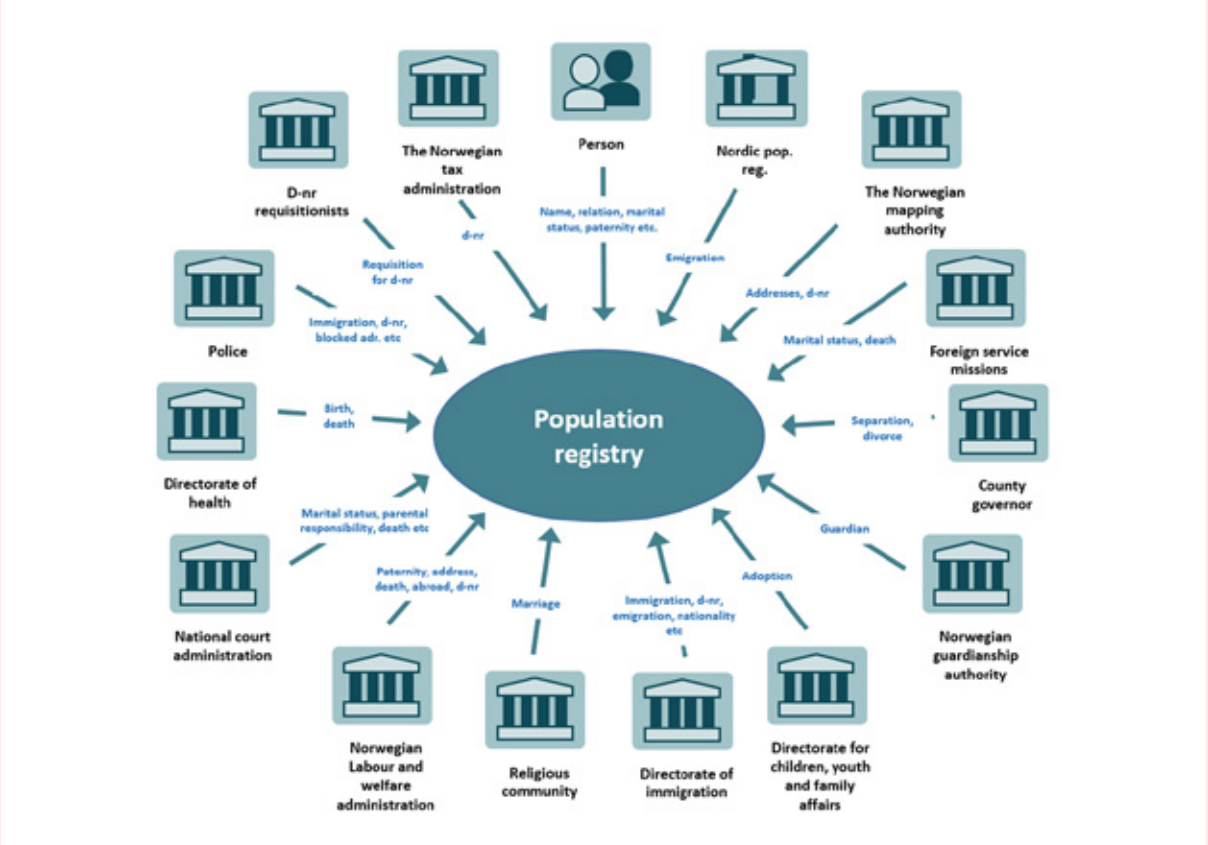
Data are transferred from the CPR to Statistics Norway, Norway's National Statistics Institute, which together with processed data from UDI enables Statistics Norway to produce statistics on stocks and flows of people with a refugee background.

Persons who have come to Norway as refugees or asylum seekers are included in the CPR once they are recognized as legal residents. Refugees are given a PIN on arrival, whereas asylum seekers are assigned another ID number, called the D-number, when they apply for asylum. The D-number is used for administrative purposes for people who are not residents of Norway but who have economic or other links to Norway. It is only when a person is granted asylum that he or she is assigned a PIN number. Persons with D-numbers are not included in the official population statistics for Norway.

The inclusion of the PINs of parents and spouses in the CPR makes it possible to establish links between siblings, cousins, children, grandparents and other relatives. The PIN is used in a large number of other administrative registers, which makes it possible to link information in different registers for statistics and research. The CPR is also used to draw samples for sample surveys. The contents, coverage and quality of the administrative registers have become so good that it is no longer necessary to conduct traditional population and housing censuses.

Figure 1 shows the extent of data flows to and from the CPR. It should be noted that the population register does not receive any microdata from Statistics Norway, with a few exceptions, in keeping with the fundamental principles of statistics that individual data collected by statistical agencies for statistical compilation are to be strictly confidential and used exclusively for statistical purposes.

Figure 1: Data flows to the Norway Population Register



Statistics Norway regularly collects data on new vital events and other changes from the CPR and use these data to update the statistical population register. Statistics Norway has access to many other administrative and statistical data sources and can link them to the copy of the central population register to check for possible errors or add information, such as country of birth if not registered by the CPR.

Statistics Norway publishes regular analysis, including on persons with a refugee background[51] and on refugees in the labour market. The report *Refugees inside and outside the labour market*[52] analyses the labour market participation of persons with a refugee background and is an integral part of

51 Statistics Norway Persons with refugee background

52 Statistics Norway Flyktninger i og utenfor arbeidsmarkedet 2020

the register-based labour market statistics in Norway. The underlying data sets on employment/unemployment used for the statistics are the same for refugees, migrants and the general population and are used for publishing annual statistics. Statistics Norway has published statistics on labour market participation of persons with a refugee background since the 1990s.

The analysis includes comparison with the general population as well as the immigrant population and includes estimates for the refugee population by country of origin, age group, gender, and time since arrival in Norway, using cross tabulations and statistical inference. The report also examines which industries refugees work in, whether they work part time or full time, and the number of hours worked.

The analysis shows the impact on the integration into local labour markets of factors such as educational attainment and stock of refugees of the same nationality. The report also includes a chapter on the characteristics of persons with a refugee background who are not part of the labour force.

The following boxes show example extracts from publications on persons with refugee background and refugees in the labour market.

Statistics Norway, published statistics on persons with refugee background, extract

Persons with refugee background

Updated: 7 July 2022

Next update: Not yet determined

Proportion of persons of the population with refugee background

4.5 %

Download table as ...

Persons with refugee background. 1 January

	2022	Change in per cent	2022	
		2021 - 2022	Proportion of persons with refugee background of all immigrants. Per cent	Proportion of persons with refugee background of the whole population. Per cent
Total	244 660	1.8	29.9	4.5
Principal applicants	177 346	1.6	21.6	3.3
Asylum seekers	114 825	-0.2	14.0	2.1
Resettlement refugees	45 063	7.8	5.5	0.8
Other refugees	9 444	-0.9	1.2	0.2
Unspecified	8 014	-0.9	1.0	0.1
By family connection to refugee	67 314	2.4	8.2	1.2
Family enlargement	18 390	2.5	2.2	0.3
Family reunification	48 877	2.4	6.0	0.9
Family unspecified	47	2.2	0.0	0.0

Source: https://www.ssb.no/en/befolkning/innvandrere/statistikk/personer-med-flyktningbakgrunn

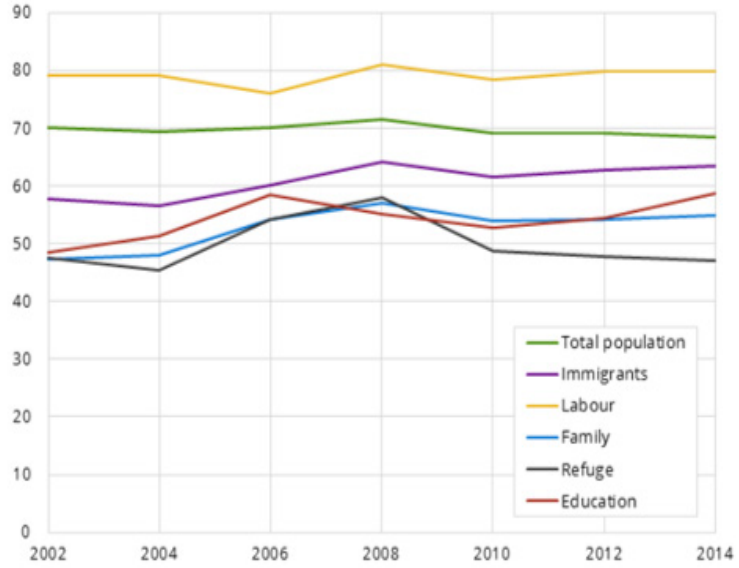
Statistics Norway, published statistics on refugees in the labour market, extract

Employment rate increases with length of residence

How have the refugees in Norway faired? Employment levels among immigrants are lower than among non-immigrants, with 63 and 69 per cent respectively (Statistics Norway 2015b). As expected, labour immigrants have the highest employment rate, while refugees have the lowest (see figure 7). Employment among refugees has declined somewhat in recent years, partly due to the increase in the share of new arrivals, and partly because there have been an increasing number from countries with a low employment rate

Figure 7

Percentage employed, by reason for immigration



Source: Statistics Norway.

Source: https://www.ssb.no/en/befolkning/artikler-og-publikasjoner/refugees-in-norway

CASE STUDY: DATA INTEGRATION IN THE ABSENCE OF UNIQUE IDENTIFIERS - USING MACHINE LEARNING TO IMPUTE REFUGEE STATUS IN SURVEY DATA IN GERMANY

Following the unprecedented increase in asylum applications in 2015 and 2016, the integration of refugees in housing markets, labour markets and in civic society is still a major priority for policy makers in Germany. Accordingly, policy makers demand for data on the situation of refugees, in order to monitor the progress of integration and to derive implications for integration policies. The idea of integrating data sources to satisfy this demand is intuitively appealing. Data integration is cost effective and imposes no additional response burden since information is produced by making better use of already existing data.

In general, countries differ with respect to their data environment and hence face different possibilities for integrating data sources. Differences largely depend on whether or not unique person identifiers exist in administrative data bases and whether or not the legal framework allows their use in official statistics. While matching datasets via unique identifiers represents the gold standard, the vast majority of countries will not be able to link information on refugees via unique identifiers within the foreseeable future. In this data environment statistical matching offers an alternative tool for integrating datasets. Statistical matching is usually described as the joint analysis of variables that are not jointly observed but available in different datasets. In contrast to record linkage that links information on identical units, statistical matching can hence be framed as an imputation problem in which a target variable in a recipient data set is predicted based on information on similar units – "statistical twins" – in a donor data set.

The data environment in which statistics on refugees and people in refugee-like situations are produced in Germany is typical for many countries: Germany uses administrative data on residence permits from the Central Register of Foreigners (CRF) to identify refugees and foreigners in refugee like situations. While the CRF confidently identifies and comprehensively covers refugees, it does not collect information on their socioeconomic characteristics or housing and living conditions. In addition, German official statistics uses the Labour Force Survey (LFS) for shedding light on the socioeconomic characteristics and the structural integration of first- and second-generation immigrants in Germany. While the LFS provides a detailed socioeconomic picture on these immigrants, it does neither confidently identify nor comprehensively cover refugees and foreigners in refugee-like situations.

Considering the production of statistics on refugees and people in refugee-like situations, the problem can be formally stated as follows: the information from the administrative data set (donor) are used in order to impute refugee status in the survey data set (recipient). For imputation a set of common variables (C) is used. The final aim is analysing the unobserved joint distributions of refugee status (R) and socioeconomic characteristics (S).

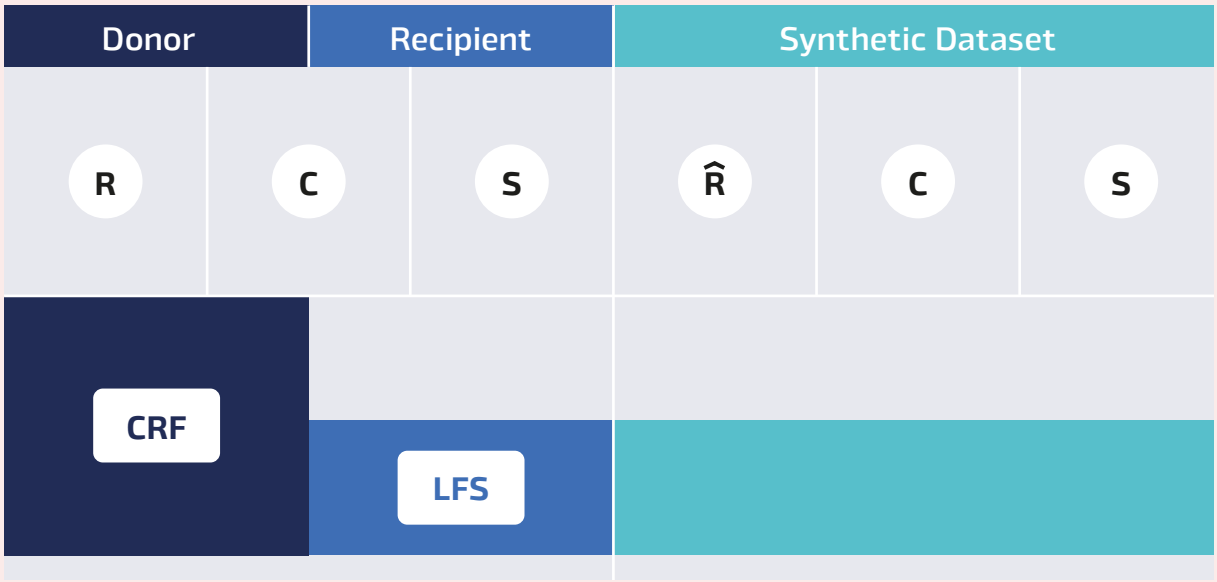
(1) Donor with C_i, R_i

(2) Recipient with C_i, S_i

(3) Prediction $\widehat{R}_i = f(C_i)$

(4) Synthetic data set with $C_i, S_i, \widehat{R}_i$

Figure 2: The structure for imputing refugee status



This Case Study demonstrates the possibility of integrating the CRF and LFS via statistical matching to combine reliable information on residence status with a broad set of socioeconomic characteristics, e.g. educational background and labour market participation. The common information on citizenship, date of entry, age and gender are used to impute refugee status for respondents of the LFS. For this task, a boosted and post-pruned Classification and Regression Tree (CART - C5.0) in the CRF is developed and used to identify refugees among LFS respondents. Statistical inference after statistical matching is implicitly based on the restrictive but untestable assumptions of conditional independence. Therefore, it is important to carefully evaluate the performance of the model.

Annex 2: Capture-recapture method for estimating target population

The estimation of the unknown size of a target population is very important for official statistics. This is the situation when multiple sources (at least partially overlapping) are available but the combined data entail under-coverage of the target population even in an ideal error-free state. In this case, the first statistical objective of the analysis is to estimate the unknown size of the target population collected in the different sources. The most common approach to address this task is the capture-recapture (CRC) method. In case of two lists, the basic CRC method relies on the following assumptions (Wolters, 1986):

- the population is closed, so the population measured in both sources is the same;
- records from both sources can be linked without errors;
- the inclusion probability of being registered in the first source is independent of the inclusion probability in the second one;
- units have the same capture probabilities within each source (homogeneity probability assumption);
- over-count in both sources is negligible.

In the case of refugees or IDPs statistics, the first assumption can be considered as true if the time between the capture and recapture is limited. Otherwise, deaths, births, and moves of the population of interest will break the hypothesis of a closed population. The validity of the second assumption depends on the quality of the available identifier for linking the sources. The third assumption is perhaps the most challenging, as those refugees or IDPs who are difficult to reach through surveys are also likely to be missing from administrative registers, which implies a problem of persistent under-coverage for specific groups and resulting CRC estimates will be weak, particularly if the second register overlaps greatly with the first register and adds relatively few new records to it. The fourth assumption corresponds to the fact that for example children, women and men have the same probability of inclusion. In case this is not verified, the CRC method can be applied to each identified specific subgroup before aggregating the results. Lastly, the fifth hypothesis could be assumed to be verified in most cases.

The violation of these assumptions can lead to serious bias in the CRC-estimation of the population size. Extensions of the CRC method have been proposed in order to address problems connected to violation of the basic assumptions and can be found in statistical literature. For example, the CRC model was explicitly designed and developed to estimate under-coverage, but the measurement of over-coverage in the context of population size estimates has also emerged as an important topic (e.g. Statistics Canada, 2015 and ONS, 2012). This reflects the risk of duplicated records in administrative data, and the possibility for administrative data to contain units outside the target population.

Example: Basic capture recapture method

Assume there are two different sources for a given area (survey and administrative sources for example) and the following data are collected:

- Survey: 500 IDPs identified
- Administrative list: 1000 IDPs identified, among which 140 can be linked to IDPs identified during the survey (recapture).

Based on these data, the following contingency table can be built:

Collected data	Not captured in administrative data	Captured in administrative data
Not captured during survey	N_{00} = ?	N_{01} = 1000 - 140= 860
Captured during survey	N_{10} = 500 - 140 = 360	N_{11} = 140

The **estimated population** in this case would be:

$N^* = (N_{10} + N_{11}) / [N_{11} / (N_{01} + N_{11})] = 500 / (140/1000) = 3{,}571$

Where

- $N_{10} + N_{11}$ = the number of IDPs identified during the survey
- $N_{11} / (N_{01} + N_{11})$ = the number of IDPs identified during the survey as a share of the total number of IDPs identified in the administrative list.

The **variance of the estimator** in this case would be:

$V(N^*) = N^* N_{01} N_{10} / N^2_{11} = 3571 \times 860 \times 360 / 140^2 = 56{,}414$

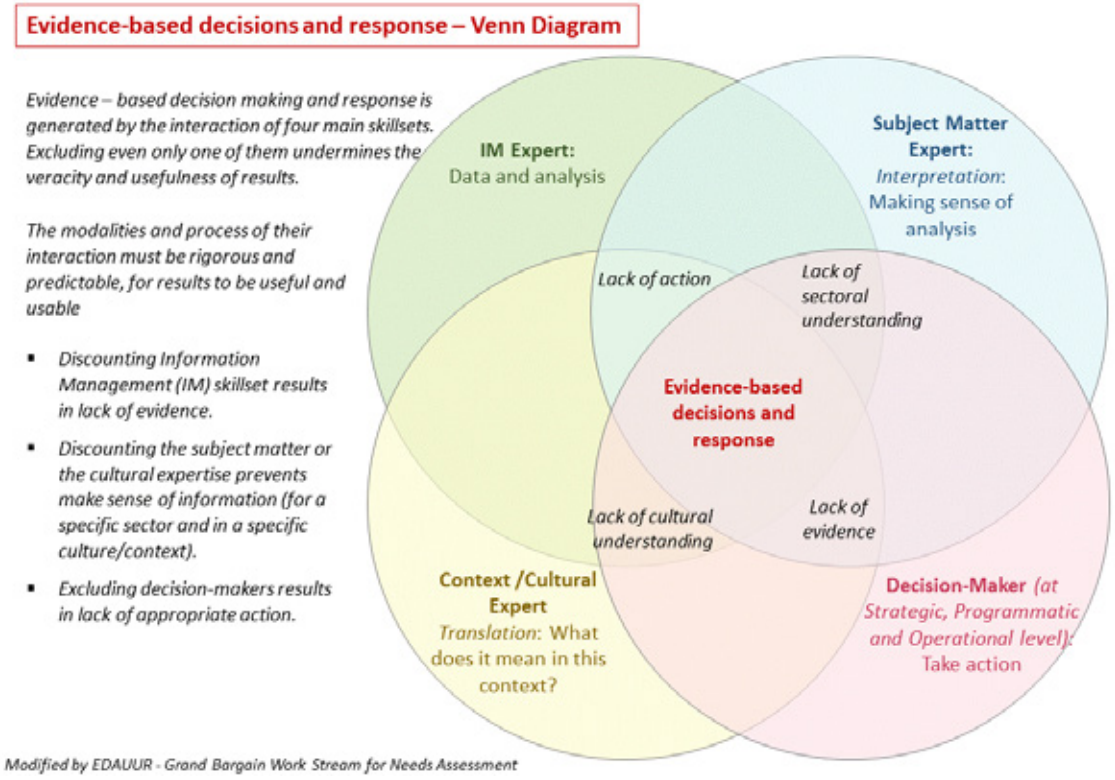
Which gives the following **95% confidence range**: [3096; 4046]

Annex 3: Tools for developing useful and usable data

The Grand Bargain was launched during the World Humanitarian Summit in Istanbul in May 2016 and is an agreement between some of the largest donors and humanitarian organisations who have committed to improve the effectiveness and efficiency of humanitarian action. The following tools are an output of the *Working Group on Useful and Usable Data and Analysis (EDAUUR)* under the Grand Bargain work stream on Needs Assessment[53].

a. Roles and questions

Four roles are needed for evidence-based decision-making: decision-makers, sectoral experts, context/cultural experts, statistical experts. Skillsets of each one and consequences of excluding one of these roles from the process is illustrated in the following diagram.



Close cooperation among the four roles is required in all stages of the process. Cooperation is particularly important during four key steps of the process:

- Define specific information needs,
- Methodology design,
- Tools and data analysis plan design,
- Analysis considered as one 'composite' step.

53 The tools are available on the website for Workstream 5

b. Define specific information needs: roles & responsibilities

Decision-makers clearly identify the decisions they have to make and communicate their data and information needs for these decisions to the statisticians, sectoral and cultural/context experts. Decision makers may also include a wider group of users including the media and the public.

Sectoral experts identify what information is available and accessible, and where the gaps are. They identify the essential building blocks of sectoral information and help statisticians identify the most appropriate sources of sectoral information, unit of analysis and modality of data collection, providing the sectoral perspective to the conversation.

Cultural/context experts help turn abstract information needs into specific local 'language'. They identify the most appropriate way of translating information needs into questions that will be correctly understood in that specific context/culture.

Statisticians contribute their technical know-how on questionnaire development, data collection modalities and data analysis. They translate the information needs of decision makers and the sectoral and cultural perspectives into questions that can be analysed to provide the information needed by decision-makers.

In the design phase, identifying information needs before developing questions is essential to get the needed results for response. Questions will thus be more targeted and obtain the right data.

Guiding questions (and who should answer):

- What is it that you have to do? (Decision makers)
- What are the decisions you need to make? (Decision makers)
- What information do you miss in order to make that decision? (Decision makers)
- How often should that information be updated, at a minimum, to be still usable? (Decision makers, sectoral experts and context experts)
- Is that information already available/ accessible? (Sectoral experts, context experts and statisticians)
- How will this information help in the decision-making *(What are logical flow & benchmarks)?* (Decision makers)
- What are the components of the information (e.g., data that can be analysed to obtain the needed information)? (Sectoral experts, context experts and statisticians)
- Are any of these data already available/accessible? (Sectoral experts, context experts and statisticians)
- What are the missing data we need to collect? (Sectoral experts, context experts and statisticians).

c. Methodology design: roles & responsibilities

Statisticians are responsible for choosing the appropriate methodology for the specific context and purposes. The choice will be based on information needed, the needed level of detail, the time frame required, data access and available resources. Information from sectoral, cultural experts and decision makers are necessary for statisticians to choose the methodology. Statisticians explain to other actors the impact on the results of different data collection methods and the proposed unit of analysis.

Sectoral experts are fully involved in the process to define the methodology, as their expertise is needed to identify, among others, appropriate unit of measurement, method of data collection, typology of respondents, resolution of data.

Cultural/context experts are crucial part of the conversation, as they provide insight on information that lead to the choice of methodology: for example, on most appropriate respondents, affected groups, geographic boundaries (e.g., groups living across admin boundaries, control of territory), accessing gatekeepers, respondents' likely reaction to a type of interview.

Decision makers' involvement in deciding on methodology greatly increases their understanding of meaning and reliability of results. It also increases their level of trust in the results. If decision-makers cannot participate, they take the necessary time to understand the limitations and strengths of results and reasons for agreed methodology. The limitations and strengths of results as well as the reasons for choosing such methodology should be presented to decision-makers in a format they can easily absorb and communicate to others.

Guiding questions (and who should answer):

Statisticians should use these questions to facilitate the identification of a sound and appropriate methodology. Identification is done jointly with sectoral experts, cultural experts and decision-makers. The statistician's skillset is essential in developing a methodology that can provide high quality results, to inform decision-makers.

WHERE:

- What are the geographic boundaries?

WHEN:

- What is the timeframe for the results?
- Should this be a 'one off' or an 'on-going' exercise?

HOW MUCH:

- How much will it cost?
- Who /how many staff will need to work on it?

HOW - Data collection methods:

- What is the most appropriate method of data collection?
- What is the most appropriate unit of analysis (e.g., individual, household, community or facility), or a structured phased approach is to be used over time, starting with community-level information, then household /individual- level?
- What method of data collection is most appropriate?
- What level of precision is required?
- Is it likely that the data will be re-used for other statistical purposes?
- What is the most appropriate interviewing method – are specialist interviewers required?

WHO:

- Who do we ask? (e.g. which members of the household, which key informants)
- Can we access the relevant gatekeepers?
- Define the scope of the survey, the unit of analysis (affected groups, communities, areas)
- Define the resolution or level of disaggregation of the data (villages, sub-districts, districts, etc.)

AND VERIFYING:

- Would decision-makers be able to make identified decisions based on data from these sources, methodologies/modalities and unit of analysis?
- How will the selected method, timeframe, geographic boundaries, selected respondents impact the results and ability to accurately inform decision-making?
- Would decision-makers trust the results obtained with this methodology?
- What are the possible negative consequences that may result from data collection/analysis using methodology and how can these be mitigated? (this must be including all actors, sectoral experts, context/cultural experts, IM, decision-makers, in addition to protection experts)

d. Data analysis plan and assessment tool design: roles & responsibilities

The purpose of developing and sharing a data analysis plan is to visualize the final descriptive analysis and verify whether it provides the information originally identified as needed by decision makers.

Statisticians develop draft questions with support of sectoral and cultural experts, in coordination with decision-makers, to ensure the link between objectives and draft questions is clear and correct. Statisticians then visualize potential results using fake data. Results are discussed with decision makers, cultural and sectoral experts to verify that the questions can provide needed information. Statisticians make necessary changes to the questions, finalize and share the data plan. Statistician also designs the tool, according to best practice.

Guiding questions and tasks (and who should answer/do it):

- In this context, using this methodology/modality at this level of analysis, what question/questions should we ask in order to obtain the data we need? (Sectoral experts, context experts and statisticians)
- In this context, using this methodology/modality at this level of analysis, what options for answers should we give in order to obtain the data we need? (Sectoral experts, context experts and statisticians)
- In this context, using this methodology/modality at this level of analysis, can this question and its analysis do harm? (Protection experts, decision-makers, sectoral experts, context experts and statistician)
- Questions and answers are included in a data plan, which may also include dissemination modality, source of data and more
- Fake values are attributed to each question and mock-up visuals (e.g., charts, maps and tables) are created for each question, and combination of questions, as deemed appropriate. (Statistician)
- Mock-up visuals are discussed with decision- makers to verify whether the analysis of the questions would provide the information identified as needed by decision makers. (Decision makers, sectoral experts, context/cultural experts and statisticians)
- Changes to questions, options for answers and visualizations are made according to outcome of previous discussion, and final version is shared with other actors. (Statistician)
- Assessment tool is designed, according to best practices (Statistician).

e. Analysis: roles and responsibilities

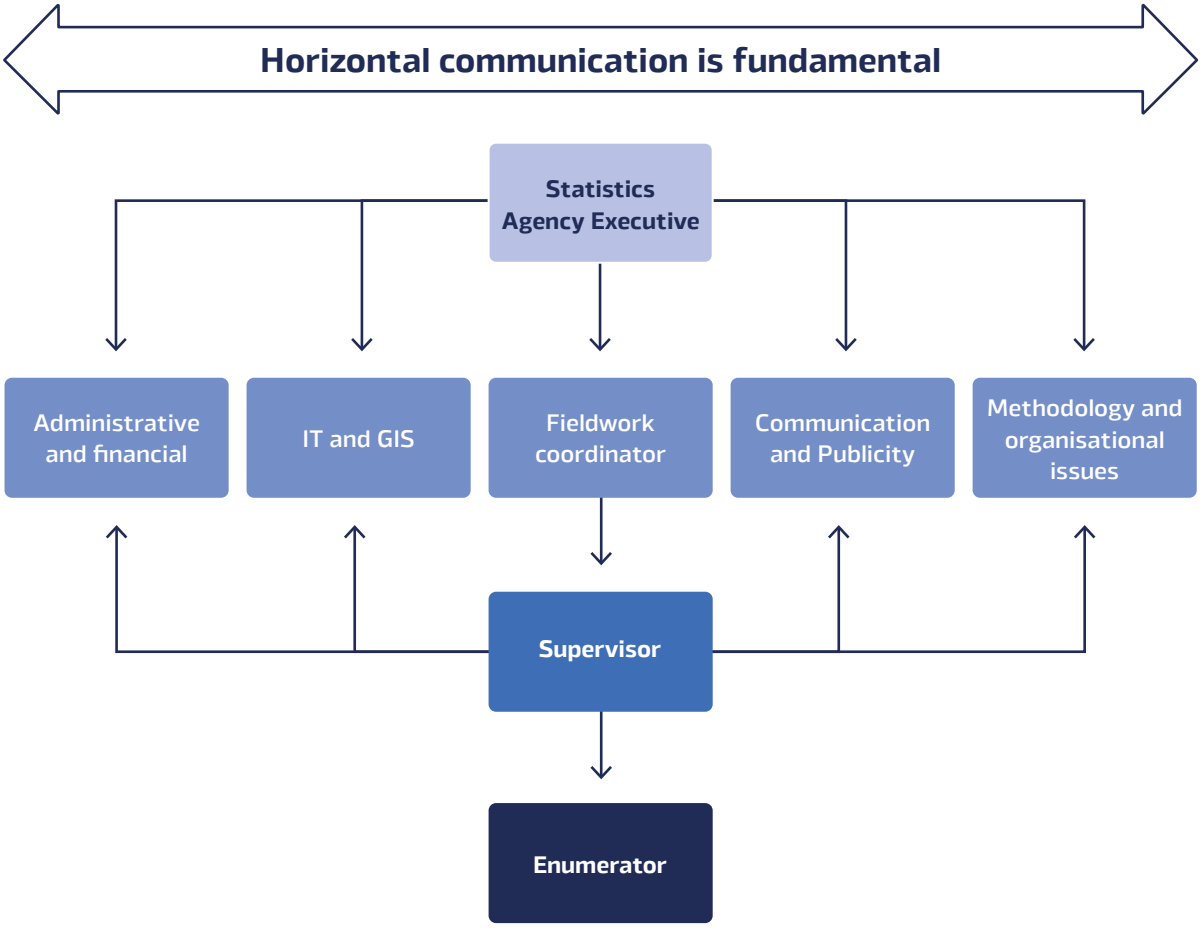
Analysis is carried at different levels, this should be clear to all actors. In each level of analysis, different actors/ skillsets answer different questions. Each level relies on the preceding ones.

- **Descriptive analysis:** Describing data means to summarize and reduce large amount of data to a representation where it is easier to compare between them and identify the main points, important stories and relevant messages, e.g. a percentage, average, mean, mode, etc.
- **Explanatory analysis** looks for associations, correlations and more generally for connections between observations and measurements. It allows for formulation of better hypothesis or theories, based on careful investigation of relationships, underlying processes or causal mechanisms.
- **Interpretive analysis** aims at moving beyond findings to identify key messages and drawing well-supported conclusions, through careful argumentation, evaluation of the strength of evidence available and attention to plausibility in context.
- **Anticipatory analysis** identifies the likelihood of future events and outcomes at a specific time, based on current and historical data. It combines predictions (a one-off estimate of a specific event in the future – What will happen?) and forecast (a set of possible futures that include probability estimates of occurring – What else might happen?)
- **Prescriptive analysis** entails both response analysis and planning. This process is generally conducted in a workshop setting and uses results from both secondary and primary data collection.

Guiding questions, roles & responsibilities:

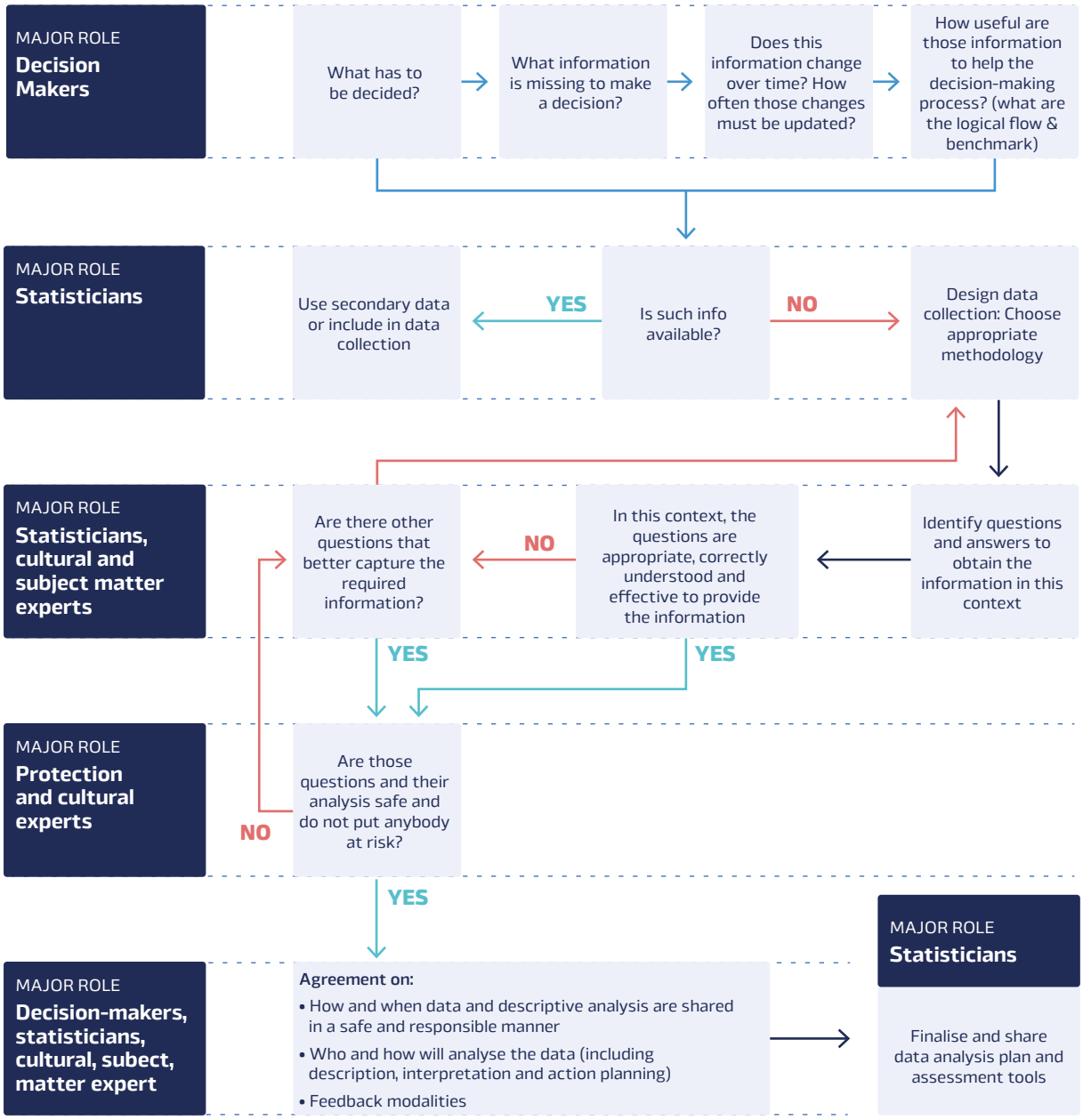
- **Descriptive analysis:** Who? What? Where? When? How many? (mainly statistical, with support by context- and subject-matter experts)
- **Explanatory analysis:** Why? How come? (mainly context- and subject-matter experts, with support by statistical experts)
- **Interpretive analysis** What does it mean? What else could it mean? (mainly Context- and Subject-matter experts, with support by statistical experts)
- **Anticipatory analysis**: What will happen next? What else might happen? What if? What then? (mainly context- and subject-matter experts, possibly with decision makers)
- **Prescriptive analysis:** What can be done? What should be done? (Decision makers, with support by context- and subject-matter experts)

Based on the model of organization for the statistical agency, the communication flows should be horizontal across all units and diagonal with regards to the supervisor in the field. It should also be a two -way communication, as shown in the following diagram.



For communication diagram to be more effective, depending on the context, a Memorandum of Understanding for data sharing between statistical agencies is required, in accordance with prevailing statistical and data protection laws. The NSO or member of the NSS responsible for publishing the results should have access to the data, under statistical legislation.

An activity tracking system is desirable. A Gantt chart is recommended, detailing the activities and moment of implementation. There are several software tools that allow for project management, or it can be done in a simple Excel spreadsheet. The activity tracking system should ideally include a financial component tied to disbursements and payments. A more advanced tracking system would include dependencies in the activities.


MAJOR ROLE
Decision Makers
What has to be decided?
What information is missing to make a decision?
Does this information change over time? How often those changes must be updated?
How useful are those information to help the decision-making process? (what are the logical flow & benchmark)
MAJOR ROLE
Statisticians
Use secondary data or include in data collection
YES
Is such info available?
NO
Design data collection: Choose appropriate methodology
MAJOR ROLE
Statisticians, cultural and subject matter experts
Are there other questions that better capture the required information?
NO
In this context, the questions are appropriate, correctly understood and effective to provide the information
Identify questions and answers to obtain the information in this context
YES
YES
MAJOR ROLE
Protection and cultural experts
Are those questions and their analysis safe and do not put anybody at risk?
NO
YES
MAJOR ROLE
Decision-makers, statisticians, cultural, subect, matter expert
Agreement on:
• How and when data and descriptive analysis are shared in a safe and responsible manner
• Who and how will analyse the data (including description, interpretation and action planning)
• Feedback modalities
MAJOR ROLE
Statisticians
Finalise and share data analysis plan and assessment tools

REFERENCES

Bertoli, S., Ozden, C. Packard, M. (2021). Segregation of Syrian Refugees in Turkey: Evidence from Mobile Phone Data.

Department for Economic and Social Information and Policy Analysis. (1993). National Household Survey Capability Programme: Sampling Rare and Elusive Populations. New York: United Nations.

Department of Economic and Social Affairs Statistics Division Studies. (2005). Studies in Methods Series F No. 96 Household Sample Surveys in Developing and Transition Countries. New York: United Nations.

Department of Economic and Social Affairs Statistics Division. (2008). Studies in Methods Series F No. 98 Designing Household Survey Samples: Practical Guidelines. New York: United Nations.

Department of Economic and Social Affairs Statistics Division. (2019). Studies in Methods Series M No. 100. United Nations National Quality Assurance Frameworks Manual for Official Statistics Including recommendations, the framework and implementation guidance. New York: United Nations.

Eckman, S. and Himelein, K. (2022). Innovative Sample Designs for Studies of Refugees and Internally Displaced Persons. Springer.

EGRISS. Membership.

EGRISS. (2018). International Recommendations on Refugee Statistics (IRRS).

EGRISS. (2020). International Recommendations on Internally Displaced Persons Statistics (IRIS).

EGRISS. (2023). International Recommendations on Statelessness Statistics (IROSS).

EGRISS. (2023). Towards a standardized approach to identify IDPs, refugees and related populations in household surveys.

ESCWA, Data Pop Alliance, Lebanon Central Administration of Statistics, Qatar Computing Research Institute (QCRI). (2020). Big Data for Good. Can Big Data Illustrate the Challenges Facing Syrian Refugees in Lebanon?

Flowminder. (2015). Monitoring population movements post disaster: 2015 Nepal Earthquake.

Green, B., Blanford, J. (2020). Refugee Camp Population Estimates Using Automated Feature Extraction. Proceedings of the 53rd Hawaii International Conference on System Sciences.

IASC. (2019). Workstream 5: Tools and guidance for advancing coordinated needs assessment and analysis through the Grand Bargain.

Instituto Nacional de Estadística e Informática. (2019). Condiciones de Vida de la Población Venezolana que reside en Perú. Resultados de la 'Encuesta dirigida a la población venezolana que vive en el pais' ENPOVE 2018. Lima: INEI.

Inter-Agency and Expert Group on Sustainable Development Goal Indicators. (2019). Data Disaggregation and SDG Indicators: Policy Priorities and Current and Future Disaggregation Plans. UNSC 50th session.

IDMC. Global Internal Displacement Database.

ILO. Labour force survey (LFS) resources.

ICRC. (2020). Handbook on Data Protection in Humanitarian Action Second Edition. Switzerland: ICRC.

IOM. (2018). IOM Data Protection Manual.

IOM. Displacement Tracking Matrix.

IOM. Migration Data portal.

JDC. JDC Literature Review.

JDC. Central African Republic (CAR): IDP survey as part of the 2021 Harmonized Household Living Conditions Survey.

JDC, World Bank Group. (2021). Burkina Faso – COVID-19 Impact Monitoring of Internally Displaced Persons | Bulletin No.1.

JIPS. (2020). Sampling Guide for Displacement Situations and Practical Examples.

JIPS. (2022). Community Engagement in Data Processes in Displacement Contexts: Technical Brief with Key Considerations & Case Studies Based on JIPS' Country Support.

Leasure, D. R., Kashyap, R., Rampazzo, F., Dooley, C., Elbers, B., Bondarenko, M., ... Mills, M. C. (2022). Nowcasting daily population displacement in Ukraine through social media advertising data.

Ministry of Justice and Public Safety Government of El Salvador. (2018). Profiling Study on Internal Mobility Due to Violence in El Salvador Final Report Match 2018. DIGESTYC, FLACSO El Salvador, UNHCR, JIPS, EU.

National Statistics Office of Georgia. (2019). Georgia Multiple Indicator Cluster Survey 2018, Survey Findings Report. Tbilisi, Georgia: National Statistics Office of Georgia.

OCHA. (2004). Guiding Principles on Internal Displacement. New York: United Nations.

OCHA. Common Operational Datasets.

OCHA. The Centre for Humanitarian Data.

OCHA. The Humanitarian Data Exchange.

Office for National Statistics. ONS methodology working paper series no. 8- Statistical uses for mobile phone data: literature review.

OpenStreetMap.

Østby, L. (2015). The population with an immigrant background in 13 municipalities in Norway Reports 2015/04. Oslo: Statistics Norway.

Palotti J, Adler N, Morales-Guzman A, Villaveces J, Sekara V, et al. (2020) Monitoring of the Venezuelan exodus through Facebook's advertising platform.

PARIS21. Advanced Data Planning Tool.

PARIS21. (2017). Guidelines for the elaboration of a National Strategy for the Development of Statistics.

PARIS21. NSDS Guidelines Sectoral strategies for the development of statistics.

Pape, U. and Wollburg, P. (2019). Estimation of Poverty in Somalia Using Innovative Methodologies Policy Research Working Paper 8735. The World Bank.

Peacebuilding Fund (PBF) Sudan, DSWG Sudan, UNHCR, JIPS. (2021). Peacebuilding With Durable Solutions for Darfur's Displaced At The Core: Insights From Eight Localities. For the UN Peacebuilding Fund (PBF) Sudan and the Durable Solutions Working Group (DSWG), 2020-2021.

Radovich E, Benova L, Penn-Kekana L, et al. (2019). 'Who assisted with the delivery of (NAME)?' Issues in estimating skilled birth attendant coverage through population-based surveys and implications for improving global tracking. BMJ Glob Health 2019.

Rampazzo, F., Bijak, J. Vitali, A. Weber, I., Zagheni E. (2021) A Framework for Estimating Migrant Stocks Using Digital Traces and Survey Data: An Application in the United Kingdom. Demography.

Statistics Norway. (2016). Refugees in Norway.

Statistics Norway. (2022). Persons with refugee background.

Statistics Norway. (2022). Refugees inside and outside the labour market.

The World Bank. (2016). High Frequency Survey 2016, Wave 3. South Sudan, 2016 – 2017.

The World Bank. (2018). Afghanistan World Bank Phone Survey of Afghan Returnees Methodology and representativeness. Washington: The World Bank.

The World Bank. (2002). Guidelines for Constructing Consumption Aggregates for Welfare Analysis LSMS Working Paper Number 135. Washington: The World Bank.

Thompson, Steven K. (1990). "Adaptive Cluster Sampling." Journal of the American Statistical Association 85, no. 412: 1050–59.

UN Global Working Group on Big Data for Official Statistics. (2019). Handbook on the Use of Mobile Phone Data for Official Statistics.

UNECE. (2015). Conference of European Statisticians Recommendations for the 2020 Census of Population and Housing. New York and Geneva: United Nations.

UNECE. (2015). Measuring Change in the Socio-Economic Conditions of Migrants (ECE/CES/42).

UNECE. (2018). Recommendations for Promoting, Measuring and Communicating the Value of Official Statistics. New York and Geneva: United Nations.

UNESCO Institute for Statistics. (2019). Minimum Proficiency Levels.

UNESCO Institute for Statistics, UNHCR. (2021). Refugee Education Statistics: Status, Challenges and Limitations. Montreal and Copenhagen, UIS and UNHCR.

UNHCR. (2015). Policy on the Protection of Personal Data of Persons of Concern to UNHCR.

UNHCR. (2018). Global Compact on Refugees. Booklet. New York: United Nations High Commissioner for Refugees.

UNHCR. (2020). Data disaggregation of SDG indicators by forced displacement. Copenhagen, UNHCR.

UNHCR. (2020). UNHCR proGres Database: Handling differences between cases and households.

UNHCR. (2021). Developing a Standardized Employment Module: Capturing key employment indicators and ways to use the module.

UNHCR. (2022). Global Trends Forced Displacement in 2021. Copenhagen: United Nations High Commissioner for Refugees.

UNHCR. Microdata Library.

UNHCR. Operational Data portal.

UNHCR. Refugee Data Finder.

UNICEF. (2016). MICS Manual for Water Quality Testing.

UNICEF, WHO. (2018). Core questions on drinking water, sanitation and hygiene for household surveys: 2018 update. New York: United Nations Children's Fund (UNICEF) and World Health Organization, 2018.

United Nations. (1951). The 1951 Convention relating to the Status of Refugees and its 1967 Protocol.

United Nations. (1998). Recommendations on Statistics of International Migration Revision 1. New York: United Nations.

United Nations. (2017). Principles and Recommendations for Population and Housing Censuses Revision 3. New York: United Nations.

United Nations. (2015). Fundamental Principles of Official Statistics Implementation Guidelines. United Nations.

United Nations. (2018). Personal Data Protection and Privacy Principles.

United Nations. (2022). The United Nations Secretary-General's Action Agenda on Internal Displacement Follow-Up to the Report of the UN Secretary-General's High-Level Panel on Internal Displacement.

United Nations. SDG indicator metadata.

United Nations. UNBigData.

United Nations Secretary-General's High-Level Panel on Internal Displacement. (2021). Shining a Light on Internal Displacement A Vision for the Future. United Nations.

UNRWA.

UNSD. Statistical Commission Decision.

US Census Bureau Center for Behavioral Science Methods Research and Methodology Directorate. (2023). Research Report Series. Use of Mobile Phone Location Data in Official Statistics, Social, Demographic and Health Studies. Washington: US Census Bureau.

USAID. (2018). Best practices for Quality Anthropometric Data Collection at the DHS Program. Rockville: The Demographic and Health Surveys Program.

Van den Hoek, J. Friedrich, H. Ballasiotes, A. Peters L., Wrathall, D. (2021). Development after Displacement: Evaluating the Utility of OpenStreetMap Data for Monitoring Sustainable Development Goal Progress in Refugee Settlements. International Journal of Geo-Information, Volume 10, Issue 3.

World Bank. (2020). Data Collection in Fragile States.

United Nations. SDG indicator metadata.

United Nations. UNBigData. Retrieved from https://unstats.un.org/bigdata/about/index.cshtml.

United Nations Secretary-General's High-Level Panel on Internal Displacement. (2021). Shining a Light on Internal Displacement A Vision for the Future. United Nations.

UNRWA. Retrieved from https://www.unrwa.org/.

UNSD. Statistical Commission Decision. Retrieved from https://unstats.un.org/unsd/statcom/decisions-ref/?code=45/110.

US Census Bureau Center for Behavioral Science Methods Research and Methodology Directorate. (2023). Research Report Series. Use of Mobile Phone Location Data in Official Statistics, Social, Demographic and Health Studies. Washington: US Census Bureau.

USAID. (2018). Best practices for Quality Anthropometric Data Collection at the DHS Program. Rockville: The Demographic and Health Surveys Program.

Van den Hoek, J. Friedrich, H. Ballasiotes, A. Peters L., Wrathall, D. (2021). Development after Displacement: Evaluating the Utility of OpenStreetMap Data for Monitoring Sustainable Development Goal Progress in Refugee Settlements. International Journal of Geo-Information, Volume 10, Issue 3. Retrieved from https://www.jointdatacenter.org/jdc-literature-review/.

World Bank. (2020). Data Collection in Fragile States. Retrieved from https://www.worldbank.org/en/topic/poverty/publication/data-collection-in-fragile-states.

CHAD.
New Sudanese refugee arrivals from Sudan in 2023.
©UNHCR/Colin Delfosse

THE EXPERT GROUP ON REFUGEE, IDP AND STATELESSNESS STATISTICS

The Expert Group on Refugee, IDP and Statelessness Statistics is a UN Statistical Commission mandated, multi-stakeholder group that works with National Statistical Offices, international organizations and civil society to develop and support implementation of international standards and guidance to improve official statistics on forcibly displaced and stateless persons.

egrisstats@unhcr.org

https://egrisstats.org

twitter.com/EGRISStats

THE COMPILERS' MANUAL

The Compilers' Manual offers clear operational instructions on producing official statistics on refugees, asylum seekers, IDPs and related populations. It complements the content of the International Recommendations on Refugee Statistics and the International Recommendations on Internally Displaced Persons Statistics by providing hands on guidance for practitioners working in National Statistical Offices and other relevant institutions.

EGRISS
EXPERT GROUP ON REFUGEE, IDP AND STATELESSNESS STATISTICS