

Statistical Commission
Fifty-first session
3 – 6 March 2020
Item 3(n) of the provisional agenda
Item for discussion and decision: Refugee statistics

Background document
Available in English only

Compilers' Manual on Displacement Statistics

Prepared by Expert Group on Refugee and IDP Statistics (EGRIS)

Expert Group on Refugee and Internally Displaced Persons Statistics

Compilers' Manual on Displacement Statistics

February 2020

TABLE OF CONTENTS

BACKGROUND	10
ORGANIZATION OF THESE RECOMMENDATIONS	11
ACKNOWLEDGEMENTS	13
Phase 0. International Coordination	16
0.1 Mainstreaming Displacement Statistics	16
a. Introduction.....	16
0.2 Coordination of International Partners.....	17
Phase 1. Specify needs.....	19
1.1 Identify the Needs of Users and Potential Users	20
1.2 Consult and Confirm Needs.....	22
a. Establishing a Sector Statistics Committee for Displacement Statistics	23
b. Using operational data from humanitarian agencies for official statistics.....	24
1.3 Establish Output Objectives	26
1.4 Identify Concepts	26
1.5 Check Data Availability	27
Phase 2. Design	32
2.1. Design Outputs	32
a. Refugee statistics outputs.....	32
a. IDP statistics outputs.....	36
b. IDP and IDP-related population stocks tabulations.....	38
c. Key IDP Stock Indicators	39
d. Examples of Outputs Illustrated in PART II.....	40
2.2. Design variable descriptions	41
a. Variables for refugee statistics.....	41
b. Variables for IDP statistics	43
Data Source A: Censuses of Population and Housing	47
2.3. Design collection.....	47
a. EGRIS recommendations.....	47
b. UNECE Recommendations for the 20202 Census Round.....	49
c. Country Examples of Census Questions for Identifying Refugees and IDPs.....	51
d. Hybrid census techniques	54
2.4 Design Frame.....	55
a. Dynamic population mapping.....	55
2.5 Design Processing and Analysis.....	55
a. Quality Assessment – the Post Enumeration Survey	55
b. PES Design	56
c. PES Analysis.....	57
Data Source B: Sample Surveys	59
2.3 Design Collection.....	59
a. Refugee questionnaires and methodology	59
b. IDP Survey questionnaires.....	60
c. Country Examples of Questionnaires found in PART II.....	61
2.4 Design Frame.....	62
a. Sampling Frames	62
b. Challenges in defining the target population and constructing a sampling frame and sample size.....	66
Data Source C: Administrative Data Including Population Registers	69

2.3. Design Collection	69
a. Population Registers	71
b. Refugee and IDP Registers	73
Data Source D: Integrated Data, Combining Data from Different Sources	77
2.3 Design Collection	77
a. Legal environment and data protection legislation	79
b. Cooperation with external data providers	80
c. IT Resource requirements	80
d. Skills needed	80
e. Statistical requirements	81
f. Collaboration with external data providers and other stakeholders	81
g. Capacity requirements	82
g. Security environment	83
2.3.3 Possible sources of data for integration	83
2.3.4 Overview on linking several data sources	85
2.5.1 Data editing and imputation.....	87
2.5.2 Creation of joint statistical micro data	88
2.5.3 Alignment of populations and measurements.....	89
a. Alignment of populations: Unit harmonisation.....	89
b. Alignment of measurements: Variable harmonisation	90
2.5.4 Multisource estimation at aggregated level	90
a. Population size estimation: multiple lists with imperfect coverage of target population.	90
b. Reconciliation	91
2.6. Design production systems and workflow	99
2.6.1. Introduction.....	99
2.6.2. The statistical production environment.....	99
2.6.3. The statistical database environment	101
2.6.4. Statistical dissemination process.....	102
2.6.6 Indicator system for measuring the design production systems and workflow ...	102
Phase 3. Build a system for collection of data and production of statistics	105
3.1 Build collection instrument	105
3.1.1 Development of questionnaires.....	105
3.1.3 Development of the collection methodology	106
3.1.3 Development of training plan, evaluation and selection of human resources	107
3.1.3 Development of collaboration agreements	108
3.4 Test production system	109
Phase 4. Data collection phase	111
4.1 Create frame and select sample for surveys	111
4.2 Set up collection	111
4.3 Run Data Collection	116
Phase 5. Process data	122
5.1. Integrate data	122
5.2 Classify and Code	122
5.3 Review and validate	125
5.5 Derive new variables and units	126
5.6. Calculate weights for data derived from probability samples	127
Phase 6. Analyse Data	128

6.1. Prepare draft outputs	128
a. Basic tabulations - estimates of counts and proportions of stocks and flows	129
b. Basic indicators (statistics, proportions and estimates)	133
c. SDG indicators recommended for refugees and IDPs	133
d. More complex analysis and commentary	135
6.2. Validate outputs	142
6.3 Interpret and explain outputs	145
6.4. Apply disclosure control.....	147
6.5. Finalise outputs	148
Phase 7. Dissemination	149
7.1. Update output systems.....	150
7.1.1 Existing output systems (portals and platforms).....	151
7.1.2 New Output system.....	153
7.1.3 Data Portals – guidelines	153
7.2. Produce dissemination products.....	154
7.2.2. Users / recipients of the dissemination products (“for whom”).....	158
7.2.3. Process ensuring proper creation of the dissemination products and best practices, guidelines (“how”).....	159
7.3. Manage release of dissemination products	161
7.4. Promote dissemination products	164
7.4.1. Conferences, public debates, fairs	164
7.4.2. Social media.....	165
7.4.3. Other forms of dissemination - visualisation tools, mobile apps, GIS portals	166
7.5. Manage user support	166
PART II – Examples and Case Studies.....	168
Country case studies	169
a. Bosnia and Herzegovina	169
b. Cameroon.....	174
c. Colombia.....	176
d. Georgia.....	178
e. Hungary.....	181
f. Nepal	185
g. Norway.....	186
h. South Sudan High Frequency Survey	189
i. Turkey	194
j. Ukraine.....	200
k. Yemen IDP Example	203
Regional Capacity Building Examples.....	208
a. MED-HIMS	208
b. SESRIC	221
c. Harmonising Refugee Statistics in the European Union.....	222
d. UNHCR Agreements and Memoranda of Understanding for the transfer of data to Governments	224
PART III – Additional Technical Material	231
PART III – PHASE 1A: Development of a stakeholder plan for displacement statistics.....	232
a. Roles and Questions.....	232
b. Define Specific Information needs: Roles & Responsibilities	232

c.	Methodology design: Roles & Responsibilities.....	233
d.	Data Analysis Plan and Assessment Tool Design: Roles & Responsibilities	235
e.	Analysis: Roles and Responsibilities	236
	PART III – PHASE 1 IDENTIFY NEEDS USING THE ADAPT TOOL	239
a.	Option 1. Cloning established indicator frameworks (logframes).....	239
b.	Option 2. Flagging indicators by collection themes (tags)	239
c.	Option 3. Identification of indicators by disaggregation variable	240
d.	Refugee and IDP statistics output in ADAPT.....	241
	PART III – PHASE 1 NSS ASSESSMENT TOOLS	242
	PART III –PHASE 2: SAMPLING METHODOLOGY	244
	PART III – PHASE 2C UNHCR Registration	250
a.	Purpose of registration data	250
b.	Registration data management.....	251
c.	Statistical Reporting.....	252
d.	Registration and Assistance Management	252
e.	Assistance management tools	253
f.	Targeting for assistance	254
g.	Registration data for Livelihood Planning and National Development.....	254
	PART III – 2D Data Integration Techniques	258
	2.3 Design Collection.....	258
2.3.1	Output objectives of data integration.....	258
2.3.2	Data availability, accessibility and other requirements	260
a.	Legal environment and data protection legislation	260
b.	Cooperation with external data providers	261
c.	IT Resource requirements	261
d.	Skills needed	261
e.	Statistical requirements	262
f.	Collaboration with external data providers and other stakeholders.....	262
g.	Capacity requirements	263
e.	Security environment	264
2.3.3	Confidentiality	264
2.3.4	Design, processing and analysis	267
a.	Data editing and imputation.....	267
b.	Creation of joint statistical micro data.....	268
2.3.4	Alignment of populations and measurements.....	275
a.	Alignment of populations: Unit harmonisation	275
b.	Alignment of measurements: Variable harmonisation	275
2.3.5	Multisource estimation at aggregated level	276
a.	Population size estimation: multiple lists with imperfect coverage of target population	276
2.3.6	Reconciliation	279
	2.6 Design production systems and workflow	286
2.6.1.	Introduction.....	286
2.6.2.	The statistical production environment.....	286
2.6.3.	The statistical database environment	288
2.6.4.	Statistical dissemination process.....	289
2.6.6	Indicator system for measuring the design production systems and workflow	290
	PART III – PHASE 4 – COLLECT	292
a.	Training Needs Plan.....	292

4.2.4 Training Manuals	294
4.3.2 Developing Supervision Forms.....	295
4.3.5 Duties of Survey Team Members	297

ACRONYMS

ABS	Australian Bureau of Statistics
AHM	Ad Hoc Module
ALMP	Active Labour Market Programme
ASEAN	Association of Southeast Asian Nations
CCSA	Committee for the Coordination of Statistical Activities
CEAS	Common European Asylum System
CRRF	Comprehensive Refugee Response Framework
DHS	Demographic Household Survey
DQAF	Data Quality Assessment Framework
EASO	European Asylum Support Office
ECA	United Nations Economic Commission for Africa
ECLAC	Economic Commission for Latin America
ECOSOC	United Nations Economic and Social Council
EFTA	European Free Trade Association
EGRIS	Expert Group on Refugee and Internally Displaced Persons Statistics
ELIPA	Enquête Longitudinale sur l'Intégration des Primo-Arrivants
EPS	Early warning and Preparedness System
ESCWA	United Nations Economic and Social Commission for West Asia
ESS	European Statistical System
EU	European Union
EU LFS	European Union Labour Force Survey
EU-SILC	European Union Statistics on Income and Living Conditions
GIS	Geographic Information System
GPS	Global Positioning System
GPS	Group for the Provision of Statistics
IBR	Individual Basic Registration
ICR	Individual Comprehensive Registration
IDMC	Internal Displacement Monitoring Centre
IDP	Internally Displaced Person
IEHR	Individual Enhanced Registration
IER	Individual Emergency Registration
IOM	International Organisation for Migration
ILO	International Labour Organisation
IMF	International Monetary Fund
IRRS	International Recommendations on Refugee Statistics
ISCED	International Standard Classification of Education
ISCO	International Standard Classification of Occupations
JIPS	Joint IDP Profiling Service
LAMP	Latin American Migration Project
LFS	Labour Force Survey
LSIC	Longitudinal Survey of Immigrants to Canada
LSMS	Living Standards Measurement Surveys
MED-HIMS	Mediterranean Household International Migration Survey Programme
MEDSTAT	Mediterranean Statistical Cooperation Program
MICS	Multiple Indicators Cluster Surveys
NEET	Not in Employment, Education or Training
NGO	Non-Governmental Organisation
NQAF	National Quality Assessment Framework

NSDS	National Strategy for the Development of Statistics
NSO	National Statistical Office
NSI	National Statistical Institute
OAU	Organisation of African Unity
OECD	Organisation for Economic Co-operation and Development
PIAAC	Programme for International Assessment of Adult Competencies
PIN	Personal Identification Number
PPP	Purchasing Power Parity
QAF	Quality Assurance Framework
RDS	Respondent Driven Sampling
RSD	Refugee Status Determination
SDG	Sustainable Development Goal
TAG	Technical Advisory Group
TurkStat	Turkish Statistical Institute
UDI	Norwegian Directorate of Immigration
UK	United Kingdom
UN IGME	United Nations' Inter-Agency Group for Child Mortality Estimation
UNECA	United Nations Economic Commission for Africa
UNECE	United Nations Economic Commission for Europe
UNESCAP	United Nations Economic and Social Commission for Asia and the Pacific
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNESCWA	United Nations Economic and Social Commission for Western Asia
UNFPA	United Nations Population Fund
UNHCR	United Nations High Commissioner for Refugees
UNICEF	United Nations Children's Fund
UNOCHA	United Nations Office for the Coordination of Humanitarian Affairs
UNRWA	United Nations Relief and Works Agency for Palestine Refugees in the Near East
UNSD	United Nations Statistics Division
WFP	World Food Programme
WHO	World Health Organization

Introduction to the Compilers' Manual

BACKGROUND

1. This Compilers' Manual is designed to support those using the International Recommendations on Refugee Statistics (IRRS:2018) and the International Recommendations on IDP Statistics (IRIS:2020) to compile statistics on displaced persons. It is based on contributions and examples provided by the Expert Group on Refugee and Internationally Displaced Persons Statistics (EGRIS).
2. The EGRIS group was established In July 2016, when the Terms of Reference for the Expert Group on Refugee and Internally Displaced Persons Statistics (EGRIS) were approved by the Bureau of the UN Statistical Commission.
3. The overall objective of EGRIS is to improve international refugee statistics through the development of international recommendations on how to collect, compile and disseminate statistics on refugees, asylum seekers and related populations. Specifically, EGRIS was mandated to develop: (a) International Recommendations on Refugee Statistics, to serve as a reference guide for national and international work concerning statistics on refugees, asylum seekers and related populations [approved March 2018 at the 49th session of the Statistical Commission]; and (b) a Technical Report outlining a way forward for the development of comparable international standards for statistics on IDPs [approved in March 2018 at the 49th session of the Statistical Commission]. EGRIS comprises over 40 national statistical and immigration authorities, the Steering Committee and around 25 regional/international organizations¹.
4. In March 2018 the mandate of EGRIS was extended to develop (a) The International

¹ 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.

¹ 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 Affairs (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 Affairs (UNSD); the World Bank; the World Food Programme (WFP) and Academia.

Recommendations on IDP Statistics; and (b) A Compilers' Manual, with operational instructions on how to collect and disseminate statistics on refugees, asylum seekers and related populations and IDPs[to be submitted in March 2020 at the 51st session of the Statistical Commission]. This document is the Compilers' Manual and is intended to be a 'living document' which will be amended and extended as the body of expertise and knowledge develops worldwide. An online version of the Compilers' Manual will be available shortly.

ORGANIZATION OF THESE RECOMMENDATIONS

5. This Compilers' Manual is divided into three sections Part I - The Manual; Part II – Case Studies and examples drawn from national and international practice; and Part III – Technical Information. The Compilers' Manual applies to both Refugee and Refugee related statistics and to statistics of Internally Displaced Persons. The two populations have different characteristics and different classificatory requirements, but there are also many similarities. The Manual clearly signposts the populations to which each section applies.

6. Part I of The Compilers' Manual is structured according to the Generic Statistical Business Process Model (GSBPM), with each chapter describing a Phase shown in the figure below in blue. These processes are described in UNECE's wiki and at the time of writing version 5.1 was current

<https://statswiki.unece.org/display/GSBPM/Generic+Statistical+Business+Process+Model>

and shown below in Figure 0.1 The Generic Statistical Business Process Model.

Table 0.1: General Statistical Business Process Model

The Generic Statistical Business Process Model (GSBPM) Version 5.0

Quality Management / Metadata Management							
Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate
1.1 Identify needs	2.1 Design outputs	3.1 Build collection instrument	4.1 Create frame & select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process components	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design collection	3.3 Build or enhance dissemination components	4.3 Run collection	5.3 Review & validate	6.3 Interpret & explain outputs	7.3 Manage release of dissemination products	8.3 Agree an action plan
1.4 Identify concepts	2.4 Design frame & sample	3.4 Configure workflows	4.4 Finalise collection	5.4 Edit & impute	6.4 Apply disclosure control	7.4 Promote dissemination products	
1.5 Check data availability	2.5 Design processing & analysis	3.5 Test production system		5.5 Derive new variables & units	6.5 Finalise outputs	7.5 Manage user support	
1.6 Prepare business case	2.6 Design production systems & workflow	3.6 Test statistical business process		5.6 Calculate weights			
		3.7 Finalise production system		5.7 Calculate aggregates			
				5.8 Finalise data files			

7. In this manual we add an additional topic, Phase 0, to cover coordination of statistical production of displacement statistics at the international level, and in particular on improving the coherence of global statistics. This Phase is not explicitly included in the GSBPM. It is however included in GAMS0 (Generic Activity Model for Statistical Organisations) under the Strategy and Leadership Activity which includes the component Manage strategic collaboration and cooperation.

8. Within the Compilers' Manual Generic Statistical Business Process Model structure, components are often data source specific. The 'Design Phase 2' is particularly data source specific and arranged under the following subheadings; 2A Census, 2B Sample Surveys, 2C Administrative Systems and 2D Data Integration.

9. Some of the sub-processes included in the GSBPM are generic to all official statistics and do not require specific content for statistics about displaced persons and are therefore not included in the manual. Statistics of displaced persons will also need to conform to whatever business processes the national statistics offices have in place for all its statistical outputs and guidance on this is therefore not included in this Manual.

ACKNOWLEDGEMENTS

The **Compilers' Manual** is developed by the **Expert Group on Refugee and Internally Displaced Persons Statistics (EGRIS)**. The recommendations are developed in close cooperation between experts from national governments and regional and international organizations, ensuring the representation of a wide array of perspectives.

From **country level** the following have provided valuable contributions to the finalization of the IRRS: Mowahed Hasibullah and Suraya Ahmadian (National Statistics and Information Authority, Afghanistan); Stephan Marik-Lebeck (Statistics Austria); Zabiha Asgar (State Statistical Commission, Republic of Azerbaijan); Abual Hossain and Bikash Kishore Das (Bangladesh Bureau of Statistics); Nicolas Perrin (Immigration Office, Belgium); Zlatan Hadžić (Agency for Statistics, Bosnia and Herzegovina); Sok Kosal (National Institute of Statistics, Cambodia); Christophe Tatsinkou (Institut National de la Statistique Republique du Cameroun); Ima Okonny (Citizenship and Immigration, Canada); Ahmat Abderahim Abbo (National Institute for Statistics, Economic and Demographic Studies, Chad); Oscar Ivan Rico Valencia (Victims' Unit of the Government of Colombia); Kô Fié Didier Laurent Kra (Institut National de la Statistique, Cote d'Ivoire); Idriss Ali Soultan (Direction de la Statistique et des Etudes Démographiques, Djibouti); Estefania Encalada and Gabriela Lugmana (National Institute of Statistics and Census, Ecuador); Abdelfattah Mohamed Ali Khallaf (Central Agency for Public Mobilization and Statistics, Egypt); Mesfin Tefera Habe (Central Statistical Agency of Ethiopia); Paata Shavishvili and Shorena Tsiklauri (National Statistics Office of Georgia); Gunter Bruckner and Jan Eberle (Federal Statistical Office of Germany); Athanasios Thanopoulos and Konstantinos Voulgaris (Hellenic Statistical Authority, Greece); Ferenc Urban and Gabrielle Vukovich (Hungarian Central Statistical Office); Dhiaa Khadum (Country Statistical Office, Iraq); Gohdar Mohamed and Serwan Mohamed (Kurdistan Region Statistics Office, Iraq); Ayman Mohammed Al-Qasem and Eman Bany Mfarej (Jordan Department of Statistics); McDonalds Obudho and Stanley Wambua (Kenya National Bureau of Statistics); Avni Kastrati (Kosovo Agency of Statistics); Lara Badre (Central Administration of Statistics, Lebanon); Abdal Majid Mohamed Ali (Libya); Mohd Sofi Ali and Mohamed Shukor Mat Lassim (Department of Statistics, Malaysia); Seydou Doumbia (Mali); Bouchra Bouziani (High Commission for Planning, Morocco); Hamimatou Amadou Garba (Institut National de la Statistique, Niger); Semiu Adeyemi Adeniran (National Bureau of Statistics, Nigeria); Dag Roll-Hansen, Hans Viggo Sæbø, Helge Brunborg, Janne Utkilen, Live Margrethe Rognerud, Minja Tea Dzamarija, Per Schøning, and Vebjørn Aalandslid (Statistics Norway); Magnar Naustdalslid and Peter Akre (Norwegian Directorate of Immigration); Habib Ullah (Pakistan Bureau of Statistics); Hana Al Bukhari and Mohammed Draid (Palestine Central Bureau of Statistics); Editha R. Orcilla, Henedin Palabras, Lourdines Dela Cruz and Rosalinda Bautista (Philippine Statistics Authority); Dahir Mohamed Noor and Mohamed Moalim (Ministry of Humanitarian Affairs and Disaster Management, Somalia); Abdirahman Mohamed Sheikh (Ministry of Planning, Investment and Economic Development, Somalia); Ramadimetja Matji and Tshwaro Gae (Statistics South Africa); Pattama Amornsirisomboon (Thai National Statistical Office, Thailand); Alper Acar, Bora Boranlioğlu, Enver Taştı, Neriman Can Ergan, Şebnem Beşe Canpolat and Serife Dilek Yilmaz (Turkstat, Turkey); Vadym Pishcheiko (State Statistics Service, Ukraine); Victoria Prieto (Universidad de la Republica, Uruguay); Eric Jensen (United States Census Bureau); Funda Ustek-Spilda (Arithmus project) and Mary Strode (Independent senior consultant).

From **regional and international organizations**, the following have taken part in the

work: Dorothee Ouissika and Maurice Mubila (African Development Bank); Alexander David (Comisión Centroamericana de Estadística del Sistema de la Integración Centroamericana); Laurent Aujean, Luca Pappalardo and Zsolt Tasnadi (European Commission); David Thorogood, Kasia Kraszewska, Fabrice Gras and Piotr Juchno (Eurostat); Adrián Calvo-Valderrama, Avigail Shai, Elizabeth Rushing, Justin Ginnetti and Luisa Meneghetti (Internal Displacement Monitoring Centre); Eduardo Zambrano, Denis Kierans, Duncan Sullivan and Vladko Avromovski (International Organization for Migration); Tilman Brück (International Security and Development Center); Corina Demottaz, Devora Levakova, Hanna Mark, Laura Kivelä, Mihaela Darii Sposato and Natalia Krynsky Baal (Joint IDP Profiling Service); Samir Farid (MED-HIMS); Giambattista Cantisani (MEDSTAT IV Project); Cecile Thoreau and Yves Brem (Organization for Economic Co-operation and Development); Martina Caterina (Office of the Special Rapporteur on the Human Rights of IDPs); El Iza Mohamedou (PARIS21); Atilla Karaman (Statistical, Economic and Social Research and Training Centre for Islamic Countries); Danzhen You and Jan Beise (United Nations Children Fund); Leandre Foster Ngogang Wandji (United Nations Economic Commission for Africa); Marwan Khawaja, (United Nations Economic and Social Commission for Western Asia); Mengjia Liang, Romesh Silva and Sabrina Juran (United Nations Population Fund); Ahmed Abukhashaba, Aina Saetre, Carolina Ferrari, Daniel Macguire, Edouard Legoupil, Hussain Watfa, Jallah Faciann, Kimberly Roberson, Petra Nahmias, Sebastian Steinmüller, Tarek Chabake and Vincent Ricouart (United Nations High Commissioner for Refugees); Greta Zeender and Javier Teran (United Nations Office for the Coordination of Humanitarian Affairs); Vibeke Oestreich Nielsen (United Nations Statistical Division); Rami Ibrahim (United Nations Relief and Works Agency for Palestine Refugees in the Near East); Caroline Sergeant, Emi Suzuki and Harriet Muger (World Bank).

The above mentioned offices and organizations have made substantial contributions through their staff; both members of the EGRIS, as well as non-members closely associated with the EGRIS work. Financial and non-financial contributions have been made available by UNHCR, Eurostat, EFTA, JIPS, NRC, NORCAP, OECD, Statistics Norway, SESRIC and the World Bank.

**PART I: Compilers' Manual Organised by the Generic
Statistical Business Process Mode**

Phase 0. INTERNATIONAL COORDINATION

0.1 Mainstreaming Displacement Statistics

a. Introduction

10. The numbers of refugees, asylum seekers and internally displaced persons (IDPs) have increased rapidly in recent years. Moreover, 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. However, developing countries are disproportionately affected by forced displacement, and there is a growing consensus that greater international cooperation is required to assist host countries and affected host communities.

11. This manual is concerned with forced displacement, people who are forcibly displaced from their homes either by disaster or by conflict. These displaced people include those who cross an international border (refugees and related populations covered by the IRRS), and those who remain in the country from which they were displaced (Internally Displaced Persons covered by IRIS). Recent global estimates suggest that the majority of displaced people are displaced within the borders of their own countries (IDMC 2019).

12. The scale of the problem is growing, although it is difficult to obtain a global estimate for all types of displaced people as the mandates of the various international agencies providing estimates do not cover the universe of displaced people. The Institute of Migration tells us that *“by the end of 2017, 68.5 million individuals were forcibly displaced worldwide due to persecution, conflict, generalized violence and human rights violations. This is almost double the number of forcibly displaced people recorded in 1997². The IoM goes on to say that “By the end of 2017, the world was hosting 25.4 million refugees. During 2017 alone, some 2.7 million people became newly displaced to another country, double the amount in 2016”.*

13. Using another more recent source for internal displacements, rather than for refugees seeking protection in another country, the Internal Displacement Monitoring Centre estimated that in 2018 *“there were 28 million new internal displacements associated with conflict and disasters across 148 countries and territories were recorded in 2018, with nine countries each accounting for more than a million”.* It should be noted that there were more displacements associated with natural disasters (17.2 million), than with conflict (10.8 million).

14. Forced displacement has gained prominence on the international agenda. The New York Declaration for Refugees and Migrants, adopted by the United Nations General Assembly on 19 September 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. Specifically, under the Sustainable Development Goal 10 to reduce inequality within and between countries, a key target is *“to facilitate the orderly, safe, regular and responsible migration and mobility of people, including through the implementation of planned and well-managed migration policies.”* The New York Declaration for Refugees and Migrants explicitly recognizes the needs of refugees, IDPs and migrants in the pursuit of this

² <https://publications.iom.int/system/files/pdf/gmdacbulletins.pdf>

target.

15. 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 refugee related populations, including complete, accurate, timely and internationally comparable estimates of the numbers of people displaced. Estimates of these populations are increasingly relevant for official statistics, which need to take account of forcibly displaced populations in a consistent manner. Incomplete or inadequate statistics on displaced populations undermine the estimation of population stocks, which in turn affects other statistics, particularly those measured in per capita terms.

16. Much of the available data on refugees and IDPs is not generated as official statistics produced by the statistical systems of national governments but arise from operational data that is collected by a variety of actors in response to a national emergency. This is data collected for the purpose of supporting displaced people, rather than for statistical purposes. Displacement contexts can pose specific challenges to NSOs, as collecting data during conflict or disaster can be challenging and impose obstacles which go beyond the usual planning for a new statistical series.

17. Given this, particular attention should be paid to coordination between the various partners operating in a country and national authorities. In emergency situations some humanitarian data is processed by international partners to become international estimates and, these estimates need to conform with the standards set for official statistics in order to be recognised as international statistics. Therefore, at the international level, a distinction should also be made between global estimates such as those published by the Internal Displacement Monitoring Centre, which draw on various statistics and data sources, and any international statistics which are produced in line with recommendations in the IRRS/IRIS, in accordance the quality standards set for UN and other international agencies, or published by any future custodian of IDP and refugee statistics.

18. The Inter-Agency and Expert Group on SDG Indicators accepted the proposal of EGRIS members to recommend disaggregation by forced displacement categories for 12 priority SDG indicators, as well as further broken down by age and sex whenever possible.³ In order to ensure availability of minimum quality statistics on refugees and IDPs, it is recommended that national statistics providers include statistics on these priority indicators in the reporting plans for the 2030 Agenda.

0.2 Coordination of International Partners

19. Some statistical series, particularly those used for monitoring the achievement of the SDGs, have been allocated an international custodian agency⁴, usually a UN agency. This custodian agency, promotes adherence to international statistical principles for the thematic area under its control; it collects and publishes national official statistics and potentially anonymised micro-data; and provides technical support and capacity-building to NSOs and/or other national bodies responsible for the production of official statistics as part of the NSS.

³ <https://unstats.un.org/sdgs/iaeg-sdgs/disaggregation/>

⁴ <https://unstats.un.org/sdgs/iaeg-sdgs/tier-classification/>

20. While UNHCR is generally accepted as the custodian agency for refugee statistics, the IRIS does not make recommendations on a specific custodian, although it outlines the role and responsibilities of any future custodian. It envisages that a custodian for IDP statistics could be appointed in the future to compiling international IDP statistics through a questionnaire, follow-up, validation, global publication and re-assessment of data collection processes.

21. International coordination is important to prevent duplication, fill existing gaps, and to develop national capacity.

Phase References and Links

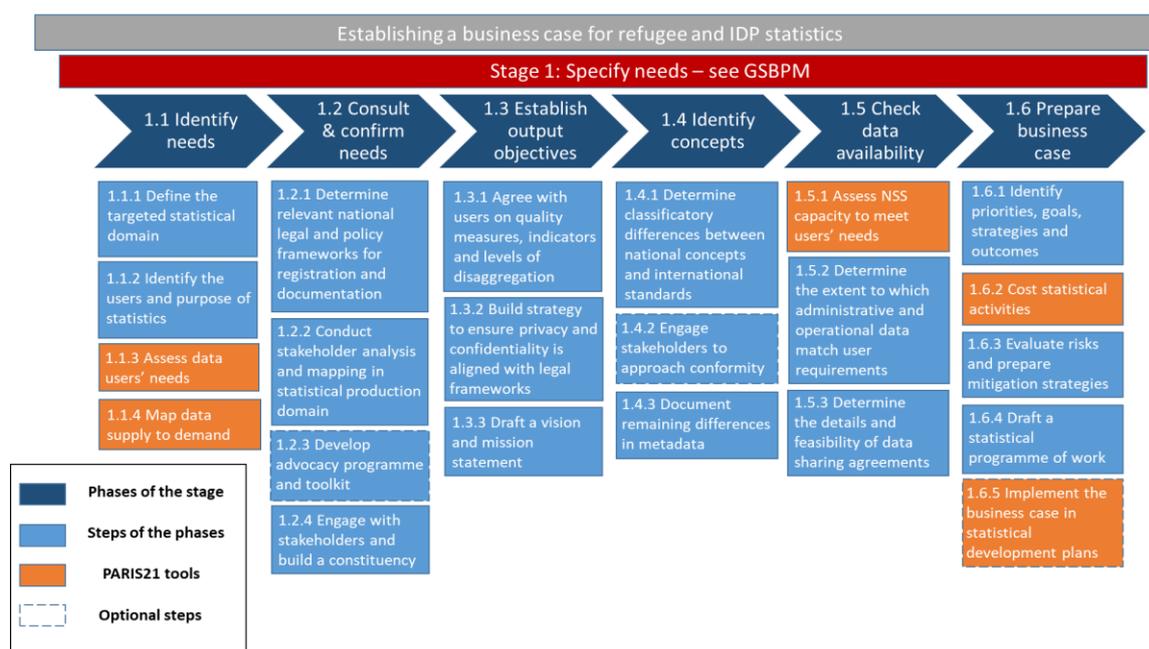
- Principles Governing International Statistical Activities:
https://unstats.un.org/unsd/accesub-public/principles_stat_activities.htm
- UN Quality standards for international statistics:
<https://unstats.un.org/unsd/unsystem/Documents-March2017/UNSystem-2017-3-QAF.pdf>

Phase 1. SPECIFY NEEDS

'This phase is triggered when a need for new statistics is identified, or feedback about current statistics initiates a review. It includes all activities associated with engaging customers to identify their detailed statistical needs, proposing high level solution options and preparing business cases to meet these needs. In this phase the organisation: identifies the need for the statistics; confirms, in more detail, the statistical needs of the stakeholders; establishes the high level objectives of the statistical outputs; identifies the relevant concepts and variables for which data are required; checks the extent to which current data sources can meet these needs; prepares the business case to get approval to produce the statistics' (UNECE, 2013)

22. A comprehensive approach to understanding and specifying statistical needs will help to make national statistical systems meet users' needs, or in the case of a new statistical series potential users' needs. The development of a roadmap for statistical development can be guided by several tools and guidelines developed by PARIS21 based at OECD. *Figure 1.1: Tools available to support establishing a business case* shows the sub-processes set out the GSBPM and indicates where tools are available for establishing a business case. These are discussed below.

Figure 1.1: Tools available to support establishing a business case



23. Advocacy from regional and international authorities can play a crucial part in raising the priority given to refugee and IDP statistics by governments, planning authorities and chief statisticians. Effective advocacy includes explicitly documenting and specifying the unaddressed needs, targeting a group within the national statistical system that can effect change, proposing a detailed business case for closing the data gap, and using hard evidence as

a basis for the business case. Joint approaches in strengthening national statistical systems, whether regional or international, have often been more effective than individual national efforts in capacity building and fundraising for emerging areas of official statistics, such as refugee and IDP statistics.

24. As the international statistical system derives its input from national statistical systems, establishing a business case for a new statistical series requires substantial planning and coordination at national and international levels alike. Nationally, the production of refugee and IDP statistics can require several ministries, departments or agencies to cooperate with the national statistics office (NSO) to make the resulting statistics fit for purpose and relevant to users (Eurostat, 2018).

1.1 Identify the Needs of Users and Potential Users

“1.1 - This sub-process includes the initial investigation and identification of what statistics are needed and what is needed of the statistics.” (UNECE, 2013)

25. The IRRS and IRIS recommendations propose a list of basic data tabulations (IRRS Chapter 3 and IRIS Chapter 5) and indicators (IRRS Chapter 5 and IRIS Chapter 4) to be produced at the national level. These recommendations serve as the core set of official statistics in the displacement domain which includes refugee and related persons; and internally displaced persons. However, a country's range of statistics will vary, depending on the national policy priorities and the reporting requirements of regional international users, and the capacity of the National Statistical System and what statistics are possible to produce with a reasonable effort

26. In respect of international reporting obligations, it should be noted that the Inter-Agency and Expert Group on SDG Indicators accepted the proposal of EGRIS members to recommend disaggregation by forced displacement categories for 12 priority SDG indicators, as well by age and sex whenever possible.⁵ In order to ensure availability of minimum quality statistics on refugees and IDPs, it is recommended that national statistics providers include statistics on these priority indicators in the reporting plans for the 2030 Agenda. However, governments should not feel limited to this priority list, and data collection on additional indicators is encouraged (IRIS, paragraph 130).

27. There are several 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,

⁵ <https://unstats.un.org/sdgs/iaeg-sdgs/disaggregation/>

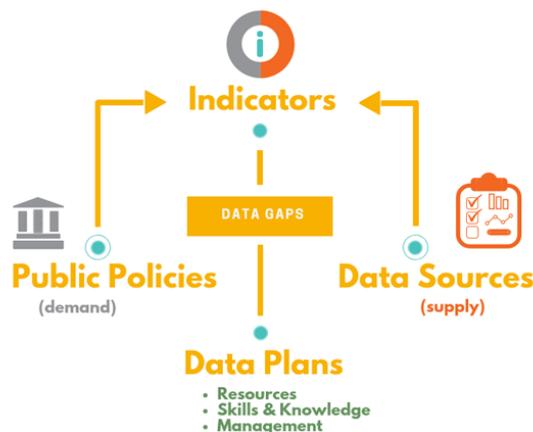
- to facilitate the allocation of resources,
- to enable regional/international comparison,
- to enable comparisons with the wider population,
- to assess progress towards durable solutions or integration with host communities,
- to inform the general public and enhance public debate.

28. Identifying important “needs gaps” and users’ priorities requires research. Which can involve a mixed-methods approach, using a combination of qualitative and quantitative research methods. A useful source of information on the needs of both displaced people and the agencies supporting them may be provided by needs assessments undertaken by or on behalf of operational agencies supporting refugees and IDPs. Needs assessments not only inform humanitarian partners but can inform those considering the data needs of those supporting the displaced. The UNHCR Needs Assessment Handbook provides guidance on their conduct (see also Phase 4 references Phase Links and Examples of Training Manuals and Data Collection Advice). These needs assessments were promoted by the World Humanitarian Summit in 2016 and established “commitments for aid organizations and donors to work together efficiently, transparently, and harmoniously in order to better deliver protection and assistance to the millions of people facing emergency needs. One of these commitments calls for the improvement of needs assessments, notably by:

1. Providing a single, comprehensive, cross-sectoral, methodologically sound, and impartial overall assessment of needs for each crisis to inform strategic decisions;
2. Coordinating and streamlining data collection to ensure compatibility, quality and comparability, and avoid over-assessment and duplication;
3. sharing needs assessment data in a timely manner, with the appropriate mitigation of protection and privacy risks; and
4. Prioritizing humanitarian response across sectors based on evidence established by the analysis.⁶

⁶ <http://needsassessment.unhcr.org>

Figure 1.2 Data gaps



29. PARIS21 has developed the [Advanced Data Planning Tool \(ADAPT\)](#)⁷ 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 website⁸.)

30. The results of ADAPT can contribute to the development of statistical plans, such as National Strategies for the Development of Statistics (NSDSs), and processes to strengthen the coordination of statistical planning at the national level. ADAPT offers three different options to identify indicators affiliated to a given statistical framework or specified topic, applicable to refugee and IDP statistics. Further details and an explanation of this tool can be found in PART III of the document (paragraph 240). Based on the information provided, the ADAPT tool can produce two main reports on the status of refugee and IDP statistics and identifies 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.

1.2 Consult and Confirm Needs

“1.2 - This sub-process focuses on consulting with the stakeholders and confirming in detail the needs for the statistics.” (UNECE, 2013)

31. 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

⁷ the Advanced Data Planning Tool (ADAPT) Manual and Capacity Development 4.0 (CD4.0) implementation guidelines provide resources to provide content and steps to the roadmap for building a sustainable business case.

⁸<https://unstats.un.org/sdgs/iaeg-sdgs/disaggregation/> and <https://unstats.un.org/sdgs/iaeg-sdgs/tier-classification/>.

policy priorities at international and national levels can attract the attendance of high-level political representatives and help to gain government support.

32. Many countries are mandated by their statistics law to set out a statistical work programme, an example of this is Turkstat whose Official Statistics Programme is based on the Statistics Law of Turkey No 5429. Many other countries have statistical development plans such as the NSDS described below.

33. [National Strategies for the Development of Statistics \(NSDSs\)](#) provide a framework for structuring support and developing statistical capacity in the national statistical system. These multi-annual strategies aim at mainstreaming sectors into the national statistical system (NSS), engaging with other data producers and coordinating the entire NSS, responding to data challenges, and building statistical capacity across the “the statistical value chain” (PARIS21, 2018). An NSDS can be an integral part of national ownership over statistical development, and secure funding at national and international level.

34. The implementation of refugee and IDP statistics in these statistical development plans, which cover all parts of official statistics, can be useful for mainstreaming these statistics into a wider national planning process and improving cooperation with the data producers involved. This is particularly important with displacement statistics as data is likely to be held by many agencies including non-traditional sources and the humanitarian sector.

35. The NSDS should be firmly linked to national development plans and involve several different sectors. Refugee and IDP statistics could be considered a sector. The identification of targeted sectors varies from country to country depending on national priorities and administrative structures. Usually, sectoral strategies correspond to line ministries, government departments or agencies with separate and well-defined areas of concern, mandate and budget. However, most countries do not have a corresponding government ministry or agency responsible for monitoring forced migration or displacement, and the level political sensitivity within these domains remains at a high level.

36. As part of the NSDS process, making the case for developing a sectoral strategy for refugee and IDP statistics is important. Traditionally in the NSDS process, technical committees are in charge of defining sectoral statistical needs, and the number of sectors involved is often constrained. Depending on its strategic importance to government, refugee and IDP statistics could face challenges in being “recognised” as a sectoral statistical area within the NSDS process and may require specific advocacy and the allocation of a lead government Ministry, Department or Agency.

37. Depending on whether the country has developed a NSDS or not, the mid-term evaluation or a new phase of the NSDS can be the right moment for integrating a new sectoral strategy 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. The process of developing a sectoral strategy for refugee and IDP statistics involves the following steps:

a. Establishing a Sector Statistics Committee for Displacement Statistics

38. Sector Statistics Committees (SSC) are responsible for defining and confirming the statistical needs for sectors. The Refugee and IDP Statistics Committee should be constituted by those leading of the sector at the request of the NSO, and a sector Coordinator should be

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

39. The responsibilities of a potential Refugee and IDP Statistics Committee would be similar to those of other sectors, including: a) proposing a vision for refugee and IDP statistics for approval, b) prioritising statistical activities over the lifespan of the Sectoral Strategy for the Development of Statistics, c) identifying major agencies currently collecting relevant statistics for refugees and IDPs, d) preparing an inventory of existent current data systems in existence operated by stakeholders, e) identifying major data needs related to the sector, f) developing strategic actions, including organisational and HR development, IT strategy, dissemination policy and other areas.

40. The role of the NSO is important for statistics coordination and to ensure that the potential data sources relevant for refugee and displacement statistics are able to be used, and that data confidentiality is adequately protected under the relevant statistics law.

41. The sector statistics design team should consist of staff from several ministries (e.g. Interior and External Relations Ministries, Migration agency) to ensure commitment and assurance of staff time through a more formal process, including memoranda of understanding between the agencies and participating ministries. As NSSs are not always able to capture forced migration and displacement within officially produced data sources, data may be needed from non-traditional data producing institutions, such as civil society and humanitarian agencies.

b. Using operational data from humanitarian agencies for official statistics

42. Where there is insufficient data from government sources for statistical purposes, operational data from humanitarian agencies present a possible resource for establishing or augmenting a new statistical series for refugee and IDP statistics. Operational data are produced during regular activities of humanitarian agencies are intended for internal use in planning operations only and do not usually meet the standards set for official statistics, before use they should be carefully assessed using a national or international statistical data (e.g. UN Statistics *Quality Assurance Framework quality assessment tool*).

43. The quality of any operational data identified for use in a new statistical series should be assessed focusing on standardisation, coverage, accuracy and confidentiality. Ultimately, the statistical authority (e.g. NSO, line ministry, specialised statistical unit within NSS) in charge will determine the usefulness of operational data as input for the new statistical series after assessing crucial quality issues and negotiating data sharing agreements (see IRRS, Chapter 6 pp 136-140).

44. One source of operational data is the IASC Common Operational Datasets, are developed and endorsed by the Inter-agency Standing Committee (IASC) and disseminated by OCHA. These are Common Operational Data sets (CODs) and fundamental operational datasets (FODs) used in humanitarian emergencies to support technical standards, improve the quality of data, and strengthen interoperability. In the IASC Common Operational Datasets

there are three levels of accountability: Guardian, Sponsor and Source. For more information see the [IASC Guidelines on Common Operational Datasets in Disaster Preparedness and Response](#).

45. The Protection Information Management for Quality Protection Outcomes (PIM) Initiative⁹ is a collaborative project, bringing together UN, NGO, and other protection and IM partners working to respond to protection needs in situations of displacement. The objective of PIM is to provide quality data and information on people in displacement situations in a safe, reliable, and meaningful way. Data on refugees and IDPs can be extremely sensitive and confidentiality and data protection are of paramount importance. Data collection should be consistent with national legislation on data protection and international obligations related to privacy. Data protection considerations may limit the access of statisticians to information in administrative systems, unless national laws establish the right of access to administrative data for statistical purposes. This project identifies three different roles for the use of PIM data;

1. **Guardian:** OCHA is the “Guardian” of the agreed upon datasets and will facilitate the distribution of the “best” available common operational and fundamental datasets in emergencies while managing forums for updates and distribution communication. If OCHA is unable to provide this service in a specific country, a suitable substitute Guardian should be identified by Humanitarian Country Teams or equivalent decision-making body during contingency planning. Quality assurance for compliance with minimum format and data characteristics in datasets will be conducted by OCHA, or the substitute Guardian, prior to distribution. This will include assigning a common GLIDE number to be associated with the dataset after the onset of a natural disaster¹⁰.
2. **Sponsor:** Each dataset has a designated “Sponsor” who is responsible for identifying and liaising with relevant “Sources” to analyse, collate, clean and achieve consensus around a specific operational dataset. If possible dataset Sponsors will proactively identify and collate information prior to emergencies in the data preparedness phase in support of the contingency planning process. If this is not possible, Sponsors will be identified in relation to their specific dataset early in the emergency and will assume responsibilities related to their thematic dataset. OCHA will maintain lists of dataset Sponsors at the country and global levels and coordinate between data Sponsors.
3. **Source:** Each dataset will have designated source(s) or owner(s), such as: national authority/agency, Cluster, NGO, UN agency, International Organization, International Red Cross/Red Crescent that agrees to be fully responsible for the development, maintenance and metadata associated with a dataset and control

⁹ <http://pim.guide/wp-content/uploads/2018/05/Framework-for-Data-Sharing-in-Practice.pdf> and <http://pim.guide/guidance-and-products/product/principles-protection-information-management-may-2015/>.

¹⁰ Accessing disaster information can be a time consuming and laborious task. Not only is data scattered but frequently identification of the disaster can be confusing in countries with many disaster events. To address both of these issues, **Asian Disaster Reduction Center (ADRC)** proposed a globally common Unique ID code for disasters. This idea was shared and promoted by the Centre for Research on the Epidemiology of Disasters (CRED) of the University of Louvain in Brussels (Belgium), OCHA/ReliefWeb, OCHA/FSCC, ISDR, UNDP, WMO, IFRC, OFDA-USAID, FAO, La Red and the World Bank and was jointly launched as a new initiative “GLIDE”. <http://glidenumbers.net/glide/public/about.jsp>

distribution restrictions.

46. 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. Part III of this manual includes an example of the data sharing Agreement on the Transfer of Personal Data of Refugees and Asylum-seekers to governments under certain conditions (see Part III UNHCR Agreements and Memoranda of Understanding). They acknowledge the importance of registration as a protection tool and expect States to take responsibility for registration as quickly as possible. The sharing of data needs to be “*subject to the data protection safeguards as contained in UNHCR’s Policy on the Protection of Persons of Concern¹¹*” The agreement allows for the ‘compiling statistical data’ among the allowable purposes for personal data transfer. Data may not be transferred to a third party except with the prior written authorisation of UNHCR and/or the consent of the data subject (the refugee or displaced person). Any third party must conform with the conditions set out in the UNHCR data sharing agreement. In the case of refugees, the data may not be shared with the refugee’s country of origin, except under the circumstances of voluntary repatriation under the auspices of UNHCR with the consent of the individual refugee.

47. The minimum set of indicators identified in the IRRS and IRIS recommendations should be considered as the core statistics to monitor internal and international displacement and for reporting to the global statistical system. However, it is recognised that most countries will not reach this minimum initially, and that time will be needed to build the capacity to produce these.

1.3 Establish Output Objectives

“This sub-process identifies the statistical outputs that are required to meet the user needs identified in the previous sub-processes.” (UNECE, 2013)

48. Once a need has been established, the practicalities of developing the statistical series will have to be defined. Establishing the practicalities include the agreement of users on the suitability of and quality measures for the proposed statistical outputs, and the recommendation of indicators and levels of disaggregation. Furthermore, this sub-process can be used to address any special concerns of refugees and IDP through the statistical series that could be relevant in the national context (trauma, sensitivity, confidentiality, migrating children etc.).

49. There are inherent potential risks of producing displacement statistics, as data use would include politically sensitive issues such as refugee support, population displacement, integration, disaster response and crossing borders. Any data collection effort in these domains needs to observe people’s rights to consent, privacy, security and ownership, and consider legal frameworks relating to confidentiality, which are both likely to be constraints when defining the output objectives.

1.4 Identify Concepts

“This sub-process clarifies the required concepts to be measured by the business process from the point of view of the user
.” (UNECE, 2013)

¹¹ <https://www.unhcr.org/uk/4b040a429.pdf>

50. An internationally agreed refugee and related persons classification was agreed in the IRRS 2018 (see Chapter 3) and a similar classification for IDPs is set out in IRIS 2020 (see Chapter 3). When referring to the refugee population within a given country, the IRRS include persons who are “primarily classified in legal terms by their claim to refugee status, or their descent from those with or seeking international protection” (IRRS, 2018). There are three major categories; a) Persons in need of international protection which includes refugees, asylum seekers and others granted temporary or subsidiary forms of protection; b) Persons with a refugee background which includes naturalized former refugees, the children and other family members of refugees, and c) Persons returned from abroad after seeking international protection. (see IRRS Figure 3.1 page 30).

51. For IDPs and related populations there are four major categories; a) Persons who have displacement-related protection needs (subdivided into IDPs in locations of displacement, IDPs in location of return, and IDPs in other settlement locations); b) IDP-related populations who are the children born after displacement to at least one IDP parent; c) Other non-displaced family members of IDPs; and d) Those who have overcome key displacement-related vulnerabilities. (See IRIS Figure 3.1 page 26).

52. From the point of view of the statistics users, nationally and internationally, classificatory differences between national laws and international recommendations may be identified. These differences may reflect the specific needs of data users. If the concepts do not align, the stakeholder should agree on a process to approach conformity, and any remaining differences need to clearly be identified in the accompanying metadata.

Classificatory differences mainly may arise due to differences in national legislation and it may be required to add the reasons for granting asylum. The reference metadata for a new statistical series on refugee statistics should pay special attention to describing and defining these differences to achieve international comparability of refugee statistics. The structure of the reference metadata should follow established statistical standards, such as Statistical Data and Metadata eXchange (SDMX)¹².

1.5 Check Data Availability

“This sub-process checks whether current data sources could meet user requirements, and the conditions under which they would be available, including any restrictions on their use.”
(UNECE, 2013)

53. 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. As national statistical systems differ in their level of centralization, the effort needed to foster consultation and collaboration between statistical producers varies accordingly. Therefore, establishing a statistical programme of work, developed jointly by the statistical producers, requires an initial mapping of strategic stakeholders to be consulted in the consultation for and the confirmation of the business case.

54. Accessing data produced by the agencies responsible for displaced people for statistical use can be problematic. There are a number of reasons why data sharing between agencies for statistical purposes can be difficult. These involve legal restrictions and issues of trust between

¹² <https://sdmx.org> and <https://ec.europa.eu/eurostat/web/sdmx-web-services/sdmx>

officials. While the statistics law is usually clear about data confidentiality and the uses to which data can be put, the agency collecting the data will also have its own laws and regulations which may prohibit data sharing, and the issue becomes one of legal precedence. The situation in the European Statistical System as at 2015-2017 is set out below in *Table 1.2 National Legal Framework - Clarity and coverage of legislation; restrictions and limitations. (European Statistical System)*.

55. Member States of the European Union and EFTA have statistics laws which enables data sharing among government stakeholders. This legal framework authorises the NSOs to access administrative sources without imposing any restrictions with reference to the provisions laid down in the [Regulation \(EC\) 223/2009](#). In this regulation *Article 24, Access to administrative records* states;

“In order to reduce the burden on respondents, the NSIs and other national authorities and the Commission (Eurostat) shall have access to administrative data sources, from within their respective public administrative system, to the extent that these data are necessary for the development, production and dissemination of European statistics.”

56. The [ESS Vision 2020](#) highlights the importance of cooperation with stakeholders and of developing strategic alliances with private and public partners (key area "identifying user needs and cooperation with stakeholders"). A workshop was organised in 2016 in order to allow the National Statistical Institutes to identify and share best practices on cooperating with the owners of administrative sources in order to access to administrative data sources ([Workshop on Access to Administrative Data Sources](#), Brussels, 13-14 September 2016). Between 2015 and 2017, information was collected on the obstacles that the NSOs still face in accessing administrative data and how they have managed to overcome some of the difficulties. On this basis, two reports were finalised in 2018:

1. **Report on the legal and institutional environment in Member States:** [Final report](#)
2. **Good practices regarding the relation with data providers:** [Final report](#)

57. The above reports set out the situation in European countries which illustrate the differences which exist between member states.

Table 1.1 National Legal framework - Scope and extent of mandate to access administrative data (European Statistical System)

	NSI has the right to access administrative microdata	NSI is obliged to request authorisation from a third party on every instance of obtaining data	Data owners are obliged to consult the NSI when setting-up, discontinuing the data source	The NSI has the authority to influence the databases	NSI has the right to access data on a cost-free basis
BE	●	○	●	●	●
BG	●	×	●	●	●
CZ	●	×	×	×	●
DK	●	●	●	×	●
DE ⁽¹⁾	○	:	×	×	●
EE	●	×	●	●	●
IE	●	×	●	●	●
EL	●	×	●	×	●
ES	●	×	×	×	●
FR	●	○	×	×	●
HR	●	○	○	○	●
IT	●	●	○	○	●
CY	●	○	●	●	●
LV	●	×	●	●	●
LT	●	×	×	×	●
LU	●	○	×	×	●
HU	●	×	×	×	●
MT	●	○	●	●	●
NL	●	×	×	×	●
AT	●	×	●	×	●
PL	●	×	●	●	●
PT ⁽²⁾	○	○	×	×	●
RO	●	×	●	●	●
SI	●	×	●	×	●
SK	●	×	●	●	●
FI	●	×	●	●	●
SE ⁽³⁾	●	×	×	×	×
UK	○	×	×	×	○
IS	●	×	●	×	●
NO	●	×	●	×	●
CH	●	×	●	●	●
LI ⁽¹⁾	●	×	●	×	●

● = For all or most sources; ○ = For some sources; × = For none of the sources; : = Missing information

Source: Eurostat: [Analysis of the Legal and Institutional Environment in the ESS](#)

Table 1.2 National Legal Framework - Clarity and coverage of legislation; restrictions and limitations. (European Statistical System)

● = For all or most sources; ○ = For some sources; × = For none of the sources; : = Missing information

	Legislation applicable to data owners contradicts NSI's right to access the data	Additional legal acts are needed for accessing specific data sources	Reuse of data for other statistical purposes is restricted	Legal restrictions on maximum duration of retention of data	Legal restrictions on linking microdata with microdata from other sources
BE	×	×	○	○	○
BG	×	×	○	×	×
CZ	○	○	○	○	×
DK	×	○	×	×	×
DE (1)	×	●	○	○	○
EE	×	×	×	×	×
IE	×	×	×	×	×
EL	○	○	×	×	×
ES	×	×	●	●	×
FR	×	○	×	×	×
HR	×	×	×	×	×
IT	×	○	●	●	×
CY	○	×	×	×	×
LV	○	○	×	×	×
LT	×	×	×	○	×
LU	○	×	×	×	×

Source: Eurostat: [Analysis of the Legal and Institutional Environment in the ESS](#)

58. Examples of other legal provisions for the production of statistics about displaced people can be found in Part II Examples and Cases studies.

1. *The Bosnia and Herzegovina Case Study* cites several examples of enabling legal and policy frameworks for the production of displacement statistics (see paragraph 543).
2. *The Turkey Case Study* cites the ability of Turkstat to access administrative data collected under the *Foreigners and International Protection law* by the Ministry of the Interior, Director General of Migration Management (see Part II Case Studies Turkey) under its legally mandated 5-year statistical planning cycle (see paragraph 638 for reference to Law mandating statistical planning). The Statistics Law of Turkey No 5429 also mandates Turkstat for determining the basic principles and standards for national registration systems, and for supporting and monitoring the organisations involved in the statistical production process.

59. Where data is being integrated to combine various sources in order to increase the scope and coverage of a statistical series, the legal constraints are particularly important. Readers should see also *Phase 2D a. Legal environment and data protection legislation* which discusses data availability at length in respect of data integration.

Phase Links and References

Eurostat (2018), "International Recommendations on Refugee Statistics (IRRS)", Eurostat, Brussels, <https://ec.europa.eu/eurostat/documents/3859598/9315869/KS-GQ-18-004-EN-N.pdf/d331c9cc-1091-43c2-b589-2c250bccc281>.

Flowminder (2015), "Nepal Earthquake 2015 Case Study", Flowminder, Stockholm, <https://web.flowminder.org/case-studies/nepal-earthquake-2015>. PARIS21 (2018), "NSDS Guidelines", PARIS21, Paris, <https://nsdsguidelines.paris21.org/NSDS-GUIDELINES-full-lang-en.pdf>.

PARIS21 (2018), "Assessing the capacity of national statistical systems: a user's guide ", PARIS21, Paris, <https://paris21.org/sites/default/files/2018-08/NSS-Assessment-Guide%20-%20WEB.pdf>.

PARIS21 (2019), "The Advanced Data Planning Tool Version 1.2 Manual for Users", PARIS21, Paris, https://paris21.org/sites/default/files/inline-files/ADAPT_Manual_April%202019_Final_v1.2.pdf.

UNECE (2013), "Generic Statistical Business Process Model", UNECE, Geneva, <https://statswiki.unece.org/display/GSBPM/V.+Descriptions+of+Phases+and+Sub-processes#V.+DescriptionsofPhasesandSub-processes-Toc375051204>.

UNHCR (2017), "Global Trends Forced Displacement in 2017", UNHCR, Geneva, <https://www.unhcr.org/statistics/unherstats/5b27be547/unhcr-global-trends-2017.html>.

Phase 2. DESIGN

“This phase describes the development and design activities, and any associated practical research work needed to define the statistical outputs, concepts, methodologies, collection instruments and operational processes. It includes all the design elements needed to define or refine the statistical products or services identified in the business case. This phase specifies all relevant metadata, ready for use later in the statistical business process, as well as quality assurance procedures. For statistical outputs produced on a regular basis, this phase usually occurs for the first iteration, and whenever improvement actions are identified in the Evaluate phase of a previous iteration.” (UNECE, 2013)

60. This Phase is subdivided into four main data sources in sub-phases 2.3 to 2.5 in order to clarify and simplify the steps for dealing with the different sources. There are different sub-sections dealing with refugees and IDPs in some of the sections.

2A. Census

2B. Surveys

2C. Admin data

2D. Integrated Data

2.1. Design Outputs

“This includes the detailed design of the statistical outputs, products and services to be produced, including the related development work and preparation of the systems and tools used in the "Disseminate" phase. Disclosure control methods, as well as processes governing access to any confidential outputs are also designed here.

Outputs should be designed to follow existing standards wherever possible, so inputs to this process may include metadata from similar or previous collections, international standards, and information about practices in other statistical organisations from sub-process 1.1 (Identify needs).” (UNECE, 2013)

a. Refugee statistics outputs

61. Basic refugee tabulations are recommended as outputs in the guidelines for basic tabulations in Chapter 3 of IRRS. In addition, there are analytical recommendations in Chapter 5 of the IRRS which propose a list of indicators for measuring refugee integration and the measurement of their immediate and ongoing needs (see IRRS, Table 5.1), these are closely aligned with the SDGs. These recommendations may need to be modified in response to national needs as determined in Phase 1 Section (1.1 Identify the Needs of Users and Potential Users) of this manual.

1) REFUGEE STOCK TABULATIONS

Table 2.1: Basic Refugee Stock Tabulations

Basic Refugee Stock Tabulations – IRRS 2018 (para 114 to 125)

The most basic statistics to be derived are the counts of each type of refugee and refugee related populations identified earlier, (a) persons in need of international protection, (b) persons with a refugee background and (c) persons who have returned after seeking international protection. The numbers are cross-tabulated by sex and age or age group. If age groups are used, then it is important to distinguish children from adults aged over 18 years, and for some purposes it may also be helpful to identify those under 16 years of age. The country of citizenship could also be considered as a basic classificatory variable, and this should include stateless persons and those with undetermined citizenship status.

(a) Stocks of persons in need of international protection

- i. Total number of persons in a country for international protection, by sex and age
- ii. Total number of asylum seekers, by sex and age
- iii. Total number of refugees, by sex and age
- iv. Total number of persons admitted for subsidiary, complementary protection, by sex and age
- v. Total number of persons admitted for temporary protection, by sex and age
- vi. Total number of persons with refugee-like status, by sex and age
- vii. Total number of persons with refugee background, by sex and age
- viii. Total number of persons returned from abroad after seeking international protection, by sex and age

(b) Stocks of persons with a refugee background

- i. Total number of persons with a refugee background, by sex and age
- ii. Total number of naturalised former refugees, by sex and age
- iii. Total number of descendants of refugee parents who are not refugees themselves, by sex and age
- iv. Total number of persons in the country as a result of refugee family reunification, by sex and age
- v. Total number of other persons with a refugee background, by sex and age

(c) Stocks of persons returned after having sought international protection abroad

- i. Total number of persons returning to their country of habitual residence after having sought international protection abroad, by sex and age
- ii. Total number of repatriated refugees, by sex and age
- iii. Total number of repatriated asylum seekers, by sex and age
- iv. Total number of persons returned after having received international protection other than refugee status abroad, by sex and age
- v. Total number of other persons returned from seeking international protection abroad, by sex and age

2) REFUGEE FLOW TABULATIONS

Table 2.2: Basic Refugee Flow Tabulations

Basic Refugee Flow Tabulations – IRRS, 2018 (para 114 to 125)

The most basic flow statistics to be derived are the counts of each category of citizens who have returned to their country after having sought international protection.

- a) Total number of persons who returned to their country of habitual residence after having sought international protection abroad during a reference period, by sex and age.
- b) Total number of repatriated refugees during a period of time, by sex and age.
- c) Total number of repatriated asylum seekers during a reference period, by sex and age.
- d) Total number of persons returned after having received international protection other than refugee status abroad during a period of time, by sex and age.
- e) Total number of others returned after seeking international protections abroad during a period of time, by sex and age.

3) REFUGEE STOCK INDICATORS

62. In addition to tabulations key indicators for the refugee and refugee related populations are shown below.

Table 2.3: Refugee Stock indicators

- a) Stock of refugees in need of international protection (IRRS, 2018 para 115):
 - i. Percentage of the total population who are in a country for international protection.
 - ii. Proportion females among persons in a country for international protection.
 - iii. Percentage of persons present in the country for international protection who are asylum seekers.
 - iv. Percentage of persons present in the country for international protection who have determined status.
 - v. Total number of unaccompanied and separated children under 18 in a country for international protection, by sex and age.
 - vi. Proportion of persons in a country for international protection who have remained for 5 years.
 - vii. Proportion of persons in a country for international protection who have remained for 10 years.
 - viii. Proportion of persons in a country for international protection who have remained for over 10 years.

- b) Stocks of persons with a refugee background (IRRS, 2018 para 120):
 - i. Percentage of the total population who have a refugee background.

- c) Stocks of persons returned after having sought international protection abroad (IRRS, 2018 para 123):
 - i. Percentage of persons returned from having sought international protection abroad among all returned citizens
 - ii. Proportion females among persons returned from having sought international protection abroad

63. Stock indicators are recommended for all three categories of refugee and refugee related populations, whereas flow indicators are recommended only for those in need of international protection, and not for those with a refugee background or returned to their country of habitual residence.

Table 2.4: Refugee Flow Indicators

- a) Flows of refugees in need of international protection (IRRS, 2018 para 115):
 - 1. Total number of asylum seekers who received a decision during a period of time, by sex and age.
 - 2. Percentage of asylum applications decisions that are positive (or negative) during a period of time, by sex and age
 - 3. 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.
 - 4. Percentage of rejected asylum seekers who left the country during a period of time, by sex and age.
 - 5. 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].

a. IDP statistics outputs

64. Basic tabulations are recommended as outputs in the guidelines for basic stock and flow tabulations in Chapter 5 of IRIS. These recommendations may need to be modified in response to national needs as determined in Phase 1 (1.1 Identify the Needs of Users and Potential Users) of this Manual.

65. IDP statistics are separated into four population groups; a) Persons who have displacement-related protection needs (subdivided into IDPs in locations of displacement, IDPs in location of return, and IDPs in other settlement locations); b) IDP-related populations who are the children born after displacement to at least one IDP parent; c) Other non-displaced family members of IDPs; and d) Those who have overcome key displacement-related vulnerabilities. (See IRIS Figure 3.1 page 26).

66. The Section C of IRIS suggests basic flow tabulations divided into inflow, outflows and flows between sub-stocks, while Section D recommends basic stock tabulations.

1) IDP AND IDP-RELATED POPULATION FLOW TABULATIONS

67. The in-flow tabulations distinguish between a) IDPs with protection needs and b) IDP-related populations. Capturing flows may be difficult and challenging for national statistical systems depending on the data collection methodology used. The flows of IDPs are more complex to estimate than the flows of refugees, as the ending of IDP status is not legally defined in international law as is the case for refugees. IDP status ceases when a durable solution has been reached, and this can only be achieved if displacement related vulnerabilities have been overcome according to the IASC framework on durable solutions (see IRIS, 2020 Chapter 4).

68. The IRIS proposes a methodology to measure the progress that IDPs have made in overcoming key displacement-related vulnerabilities, and to assess statistically when IDPs have reached a durable solution and can be taken out of the IDP stock. The methodology is to be tested in a variety of contexts and will be amended to take account of country conditions.

Table 2.5: Flows of IDPs and IDP-Related Populations

Basic IDP Flow Tabulations IRIS, 2020 (paragraphs 179 to 182)

a. BASIC INFLOW STATISTICS FOR IDPS

- i. 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
- ii. 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
- iii. 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)
- iv. 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)

b. BASIC INFLOW STATISTICS FOR IDP-RELATED PERSONS

- i. 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
- ii. 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)
- iii. 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)

c. BASIC OUTFLOW STATISTICS FOR IDPS

- i. Total number of IDPs who have died or emigrated during a specified period of time, by sex and age.
- ii. 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.
- iii. 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.

d. BASIC OUTFLOW STATISTICS FOR IDP-RELATED PERSONS

- i. 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.
- ii. 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.
- iii. 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.

b. IDP and IDP-related population stocks tabulations

69. The most basic statistics are the counts of IDP and IDP-related persons, the tabulations recommended below represent the basis analysis of IDP stocks

Table 2.6: Basic IDP Stock Statistics

Basic IDP Stock Tabulations – IRIS, 2020 (para 183)

The numbers are cross tabulated by sex and age or age group. If age groups are used then it is important to distinguish children from adults aged over 18 years, and for some purposes it may also be helpful to identify those under 16 years of age. The reason for displacement could also be considered as a basic classificatory variable, as should the region in the country or other major administrative division, in respect of both their current usual and their habitual place of residence.

a. Basic stock statistics of Internally Displaced Persons (IDPs)

- i. Total number of Internally Displaced Persons in a country, by sex, age and current place of usual residence (province/district).
- ii. Total number of Internally Displaced Persons in a country, by sex and place of habitual residence (province/district).
- iii. Total number of Internally Displaced Persons in a country by reason for displacement and current place of usual residence (province/district).
- iv. Total number of Internally Displaced Persons in a country by years since initial displacement.

b. Basic statistics of IDP-related persons

- i. Total number of IDP-related persons, by sex and age.
- ii. Total number of IDP-related persons in a country, by sex and current place of usual residence (province/district).
- iii. Total number of IDP-related persons in a country, by sex and parents' place of habitual residence (province/district).
- iv. Total number of IDP-related unaccompanied and separated children under 18, by sex and age.

c. Basic stock statistics of persons who have overcome key displacement-related vulnerabilities

- i. Total number of persons, who have overcome key displacement-related vulnerabilities in a country, by sex and age.
- ii. Total number of persons, who have overcome key displacement-related vulnerabilities in a country, by sex and current place of usual residence (province/district).

70. Chapter 4, Section C of IRIS, 2020 recommends a measure of the progress made by IDPs in overcoming their displacement related vulnerabilities. Basic progress statistics for IDPs are recommended for capturing their progress towards reaching durable solutions, measured against a reference population, either the national or regional average, or the host

community.

71. population (see IRIS, 2020 paras 145 to 157).

Table 2.7: Basic tabulations for measuring progress towards durable solutions

Basic Progress Tabulations of IDPs – IRIS, 2020 (para 184 to 185)	
a)	Basic progress statistics for IDPs
i.	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).
ii.	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 place of habitual residence and current place of usual residence (district/province).
b)	Basic progress statistics for IDP-related persons (i.e. born to at least one IDP parent after the parents' last displacement)
i.	As above
ii.	As above
c)	Total number of Persons who have overcome key-displacement related vulnerabilities (as it is still relevant to measure their progress over all eight durable solutions criteria (See IRIS, 2020 Ch 4, C1. Para 134)
i.	As above

c. Key IDP Stock Indicators

72. In addition to the basic tabulations a set of key indicators for the stock estimates of IDPs and IDP-related persons have also been recommended. For people who have overcome key displacement vulnerabilities indicators i, ii and viii to x apply.

Table 2.8: Key stock indicators of IDP, IDP-related, and those who have overcome key displacement related vulnerabilities

Basic indicators for stock of IDPs include:

- 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

d. Examples of Outputs Illustrated in PART II

73. Examples of refugee tabulations can be found in Phase 6 of this manual
1. **Norway** refugee tabulations see paragraph 413 onwards for details of outputs.
 2. **Bosnia and Herzegovina** see paragraph 538 onwards and for examples of publications in the local language see; <http://www.mhrr.gov.ba/PDF/Izbjeglice/INFORMACIJA%20O%20POVRATKU%20DO%202010.pdf>
 3. **Georgia** case study includes IDP tabulations by a) age, sex and urban/rural and b) by sex and place of current usual residence and habitual place of residence, from the 2014 General Population Census, see paragraph 583.
 4. **Turkey** case study sets out the outputs published in their 5-year Official Statistics Programme of work, see paragraph 650.
 5. **MED-HIMS** Case Study includes recommended tabulations for forced migrants – see Part II paragraph 698.

2.2. Design variable descriptions

“This sub-process defines the statistical variables to be collected via the collection instrument, as well as any other variables that will be derived from them in sub-process 5.5 (Derive new variables and units), and any statistical classifications that will be used. It is expected that existing national and international standards will be followed wherever possible.

This sub-process may need to run in parallel with sub-process 2.3 (Design collection), as the definition of the variables to be collected, and the choice of collection instrument may be inter-dependent to some degree”¹³. (UNECE, 2013)

74. Preparation of metadata descriptions of both collected and derived variables and classifications is needed for subsequent phases of the statistical process. The lists of variables described below may be modified and supplemented by the users' needs as described in Phase 1 of this Manual.

- The variables and classifications recommended refer to guidelines in Chapter 3 of IRRS (2018) and Chapter 5 of IRIS (2020).

a. Variables for refugee statistics

75. Most of the variables to be used are common to other areas of social statistics, but several are unique to refugee statistics. The classifications of asylum seekers, refugees and those with a refugee background are unique, and more detail can be found in IRRS Chapter 3. These variables required are set out in below in *Table 2.9: Variables for refugee statistics*.

76. Variable xi in the table, *Legal residential/international protection status as applicable to the national context* in the table refers to the persons legal status in respect of international protection. The classification is set out in IRRS (2018) Figure 3.1 and reproduced here in *Table 2.10: Classification of legal protection status of refugees and refugee related populations*.

77. Another variable that is unique to displacement statistics is linked to the legal concept of place or country of habitual residence. This refers to the place or country where a person was usually resident at the time they were displaced. This concept occurs in variable viii listed below for refugee statistics and in IDP statistics.

78. The usual place or country of residence is where a person is living now, and has been or intends to continue living for one year. The habitual place of residence is the place or country from where the person was displaced. IRIS (2020) Box 3.3 describes the differences. The concepts are equally applicable to country and place of residence. The concept is used in the *Bosnia and Herzegovina Case Study* (see paragraph 538 onwards) and captures the address before the war and the current address in its Database of Displaced Persons (DDPR). Georgia asks in its census “Where are you a refugee from” (see paragraph 580178).

¹³ *ibid*

Table 2.9: Variables for refugee statistics

Basic classificatory variables for refugee statistics

- i. Age or date of birth
- ii. Sex
- iii. Country of birth
- iv. Country of citizenship (including stateless, undetermined status and multiple citizenship)
- v. Date of arrival in country
- vi. Reason for migration (harmonise responses as indicated in IRRS Chapter 4),
- vii. Country of previous or last residence (for both refugees in the country and refugees returning to the country of citizenship) (country of habitual residence).
- viii. Date of first displacement/leaving previous country of habitual residence
- ix. Parents' refugee statuses
 - x. If an unaccompanied child (under the age of 18 years and separated from both parents or legal guardian)
- xi. Legal residential/international protection status as applicable to the national context, (see box below for categories)

Box 2.1 Usual and Habitual Residence

'Place of usual residence' and 'place of habitual residence' are distinct terms and they must be clearly distinguished for statistical purposes.

For statistical purposes, the **usual place of residence in the country** is where the person is living at the time of data collection. The definition of usual place of residence as per the *UN Principles and Recommendations for Population and Housing Censuses* should be used, in order to comply with international standards. This definition is generally used for the purpose of distinguishing the usual residents of a household from visitors staying with the household at the time of enumeration.

"In general, usual residence is defined ... as the place at which the person lives ... and has been there for some time or intends to stay there for some time. It is recommended that countries apply a threshold of 12 months when considering place of usual residence according to one of the following two criteria:

The place at which the person has lived continuously for most of the last 12 months (that is, for at least six months and one day), not including temporary absences for holidays or work assignments, or intends to live for at least six months;

The place at which the person has lived continuously for at least the last 12 months, not including temporary absences for holidays or work assignments, or intends to live for at least 12 months."

The definition of the **habitual place of residence** is discussed in Chapter 2 (paragraph 40). According to the legal definition, the place of **habitual residence** is where the person was residing at the time of their initial displacement (i.e. their usual place of residence pre-displacement). The place of habitual residence for IDPs is therefore static, whereas the place of usual residence, as for other population groups, is subject to change.

Source: IRIS (2018) Box 3.3

Table 2.10: Classification of legal protection status of refugees and refugee related populations

Classification of Legal international protection status (variable xi)

a) Persons in need of international protection

- i. prospective asylum seeker
- ii. asylum seeker
- iii. individual refugee status
- iv. prima facie refugee status
- v. subsidiary or complementary protection status
- vi. temporary protection status
- vii. refugee resettled in a third country
- viii. refugee-like (state form of protection granted)

b) Persons with a refugee background

- ix. naturalised former refugee
- x. child born of refugee parents without own refugee status
- xi. reunified family member
- xii. others with refugee background

c) Persons returned from abroad after seeking international protection

- xiii. repatriated refugee
- xiv. repatriated asylum seeker
- xv. returning from other form of international protection
- xvi. returning from international protection other than categories xiii-xvi.

b. Variables for IDP statistics

79. The list of variables for IDP statistics is very similar to those of refugees but omits the international migration related questions and includes the date of first and last displacement. The reason for displacement is required in a more extensive form, and the type of dwelling or habitation they are currently living in is also required. In order to identify the IDP-related population the IDP status of parents is also required.

Table 2.11: Variables for IDP statistics

Basic IDP variables	
i.	Age or date of birth
ii.	Sex
iii.	Place of birth
iv.	Date of first displacement
v.	Date of last displacement
vi.	Main reason for forced displacement;
	a. Armed conflict,
	b. Generalised violence,
	c. Violations of human rights,
	d. Natural or man-made disasters, or
	e. Other forced evictions or displacements
vii.	District/Administrative area of place of habitual residence
viii.	District/Administrative area of place of current usual residence
ix.	Parents being IDP or not
x.	Type of habitation (camp/settlement/own home/hosted by friend/relatives)

80. Variables vii and viii are used to assess whether or not individuals have returned to their habitual place of residence. This return enables the IDP to be classified according to the three sub-stocks (see IRIS, 2020 Figure 3.1 - i. IDPs in locations of displacement, ii. IDPs in locations of return, and iii. IDPs in other settlement locations) and to be counted as flow within the stock.

81. As a simple return does not mean that IDPs have overcome all their displacement related vulnerabilities, additional variables will be needed to assess the IDPs progress and to contribute to the composite measure to assess overcoming 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 Figure 4.3 reproduced below. The same chapter discusses the choice of the variables to be used, and future progress on recommendations made will be included in future versions of this Manual.

82. Therefore, in addition to classificatory variables described above, a set of variables should be chosen to assess the IDPs progress towards overcoming their displacement related vulnerabilities. *Table 2.12: Criteria and sub-criteria included in the composite measure of overcoming displacement-related vulnerabilities* lists the criteria.

83. 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.

Table 2.12: Criteria and sub-criteria included in the composite measure of overcoming displacement-related vulnerabilities

Criteria		Sub-criteria
1. Safety and security	1	1.1 Victims of violence (protection from)
	2	1.2 Freedom of movement
2. Adequate standard of living	3	2.1 Food security
	4	2.2 Shelter and housing
	5	2.3 Medical services
	6	2.4 Education
3. Access to livelihoods	7	3.1 Employment and livelihoods
	8	3.2 Economic security
4. Restoration of housing, land and property	9	4.1 Property restitution and compensation
5. Access to documentation	10	5.1 Documentation

Source: IRIS, 2018 Table 4.3

84. For each sub-criterion, there are many different indicators which can be chosen, although a final list has not yet been agreed. The indicators selected in each national context should, as far as possible, be aligned with already tested and standardized indicators. The Durable Solutions indicator library¹⁴ provides a thorough overview of potential options (see IRIS,2020 paragraph 127-128). Where relevant, they should also mirror the SDG indicators recommended to be disaggregated by different categories of displacement. 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.

85. For the purposes of statistical measurement, to make it as unbiased and cost efficient as possible, the following aspects should be taken into account when selecting the indicators:

1. Commonly used: It is an advantage if the indicator in question is collected for other purposes as well to facilitate population group comparison and to incorporate displacement analysis into existing larger data collection processes. SDG indicators are a good example here as most surveys and censuses will cover a number of these at least up until 2030.

¹⁴ In 2015, an interagency process was established to operationalise the IASC Framework. Under the leadership of the Mandate of the UN Special Rapporteur on the Human Rights of IDPs, a group of development, humanitarian and peacebuilding actors started work on developing and testing indicators and guidance for comprehensive durable solutions analysis in internal displacement situations and to measure progress over time. The work, coordinated by the Joint IDP Profiling Service, resulted in a library of standardised indicators and guidance for the operationalization of the eight IASC criteria and can serve as a unified starting point for statistical analysis on various characteristics of displacement related vulnerabilities. For more information on this process, please see: <https://www.jips.org/tools-and-guidance/durable-solutions-indicators-guide/>. For results of the work see: <http://inform-durablesolutions-idp.org/>

2. Tested and evaluated. Linked to the previous point, it is also an advantage to select indicators or questions which have been tested and satisfy quality criteria. This is the case for most of the indicators and related questions asked in standard household surveys, for example on education, food security or employment.
3. Covering the population in question: There are several potential indicators which may be relevant in the context, but do not cover all of the population in question.

86. This manual discusses four types of data sources: A. Census of Population and Housing, B. Sample Surveys, C. Administrative Data and D. Integrated Data Sources. Each of these data sources is described in turn, and subprocesses 2.3, 2.4 and 2.5 are separately described for each of the sources where relevant.

87. The design of the data collection should at a minimum aim to collect data to populate the variables described above in sub-phase 2.2. Design variable descriptions. In many cases the data collection exercise will need to identify both the refugee and related populations, and the IDP and related populations. Compromises may need to be made to keep the questioning to manageable proportions.

Data Source A: Censuses of Population and Housing

2.3. Design collection

“This sub-process determines the most appropriate collection method(s) and instrument(s). The actual activities in this sub-process will vary according to the type of collection instruments required, which can include computer assisted interviewing, paper questionnaires, administrative data interfaces and data integration techniques. This sub-process includes the design of collection instruments, questions and response templates (in conjunction with the variables and statistical classifications designed in sub-process 2.2 (Design variable descriptions))”. (UNECE, 2013)

88. In designing the Census of Population Housing there are two questions to consider:

1. Are displaced people included in the census coverage? If so, is it possible to identify them as such in the data?
2. How coherent are the results when compared to other sources such as surveys and administrative data?

a. EGRIS recommendations

89. The IRRS, 2018 includes a recommendation on identifying refugees in the population census, by asking about the main reason for the immigration to the country and allowing for an option on forced displacement. If internal migration is also included in a census, then there is the potential to screen for IDPs as well as refugees by asking similar questions on the reason for internal moves within a country. Typically, a census asks about a person's location a year ago, and in some cases 5 years ago. Where there have been large displacements in a country the date selected often 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.

90. The IRIS, 2020 also make census recommendations. Recognising that relying on internal migration questions in a census is likely to lead to an under-estimation of IDP stocks because the last move of an IDP is not always the most relevant one. The limitations are as follows;

1. Place of habitual residence will usually not be identified,
2. Those who have returned to their habitual place of residence within the period specified by the census,
3. Those who within the period specified by the census have moved one or several times after their initial displacement, and
4. Those who have moved within the same administrative area used in census geographical classifications.

91. Ideally every census would ask questions about the reasons for both internal and international migration and include a response on forced displacement. This would capture

IDPs, refugees, asylum seekers, those under temporary protection and those returning to their country of habitual residence. The reason should refer to the main reason for the most recent move. The question should be asked of all individuals, not just adults and also be asked to those living in camps and other temporary accommodation. Camps are often treated as institutional buildings and respondents are asked a brief subset of questions which are unsuitable for identifying displaced people.

Table 2.13: Census questions for the identification of refugees

<p>IRRS Recommended topics for Census to support identification of refugees and related categories</p> <p>Core questions from <i>Principles and Recommendations for Population and Housing Censuses, Revision 3</i>.</p> <ul style="list-style-type: none">i. Country of birthii. Country of citizenshipiii. Acquisition of citizenshipiv. Year or period of arrival in the country <p>In addition, it is recommended to include as a core census question:</p> <ul style="list-style-type: none">v. Reason for (international) migration, with response categories:<ul style="list-style-type: none">a. Employment (including military service)b. Education and trainingc. Marriage, family reunification or family formationd. Forced displacement (refugees, asylum seekers, temporary protected status, others)e. Other

1) INSTITUTIONAL QUESTIONNAIRES

92. In many censuses there are several questionnaire forms, there is usually one form for a private household and another for 'institutional establishments'. These institutions often comprise schools, hospitals, hotels, workers camps and so on. They 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.

93. It may also be possible to distinguish the type of camp and to record that on the questionnaire forms. The type of living quarter as recommended by the UN/ECE manual is a single response question including alternatives for "Refugee camp" and "IDP" camp. Use of this question makes it possible to distinguish between IDPs and Refugees, and "camp dwellers".

2) QUESTIONS FOR IDP POPULATIONS WHERE DISPLACEMENT IS A PRIORITY TOPIC FOR

USERS

94. If it is important to capture the entire IDP population a specific set of questions on forced displacement could be asked, but this is only likely to be justifiable in cases where there have been major national displacements and where;

1. IDPs have started to return to their places of habitual residence and it is expected that they still have key displacement related vulnerabilities,
2. IDPs are expected to have settled elsewhere after their initial displacement, but still have key displacement related vulnerabilities,
3. Multiple forced displacements have occurred to the same population,
4. Displacements are expected to have happened within the smallest administrative area proposed by the census, or
5. Displacement might be for a combination of these reasons.

95. The IRIS, 2020 recommends a set of questions for identifying IDPs and/or refugees in contexts where users require a more complete estimate of stocks and flows than is likely to be obtained from using regular internal migration census questions.

Has *(NAME) ever been forced to flee? (adding reason)

If yes;

- When was this?
- Where did you move from?
- Where did you reside before you were forcibly displaced?
- Where did you move to? (Within country borders/abroad)?

b. UNECE Recommendations for the 2020 Census Round

96. The UNECE makes recommendations for the 2020 Population Census round which cover many of the core variables required to screen the population for Refugees (♥) and IDPs (♣), although several of the items are non-core topics which are at the discretion of countries to add. The reason for migration, both international and internal, would need to include an answer category which includes the concept of forced migration. The UNECE recommendation on the reason for international migration is '*admittance for humanitarian or political reasons*'.

97. The UNECE document also provides a definition of the population group "*Population with a refugee background*". The definition of this population group includes:

“(a) Persons who declared that their main reason for migration international migration was ‘forced migration.’ Some examples of categories that would qualify as ‘forced migration’ are;

armed conflict, situations of general violence, violations of human rights, natural or human-made disasters.

(b) Foreign-born persons who declared that their main reason for migration was 'Family' and are members of the same family nucleus of a person in group (a).

(c) Native-born children members of the same family nucleus of the parents and having both parents in group (a) or one parent in the group (a) and the other parent of the group (b)."

Table 2.14: UNECE recommendations in the 2020 Population Census round

Most/all core elements needed for identification of internal migrants and with some limitations also for international migration (immigrants) are already recommended by UN/ECE 2015 (ECE/CES/41) for inclusion in the 2020 population census as follows:

Topics relevant to international migration (immigration only)

- Country of birth (core topic) ♥
- Country of births of parents (non-core topic) ♥
- Country of citizenship (core topic) ♥
- Citizenship acquisition (non-core topic) ♥
- Ever resided abroad and year of arrival in the country (core topic) ♥
- Country of previous usual residence abroad (non-core topic) ♥
- Total duration of residence in the country (non-core topic) ♥
- Reason for migration (non-core topic) ♥
- Population with refugee background (derived non-core topic) ♥

Topics relevant to internal migration

- Place of birth (core topic) ♣
- Previous place of usual residence and the date of arrival at the current place (core topic) ♣
- Place of usual residence five years prior to the census (non-core topic)
- Internally displaced persons (derived non-core topic) ♣

Legend: Supports recommended variables on Refugees (♥) and IDPs (♣)

98. The UNECE recommendations cover IDPs as a non-core topic in the internal migration section, which is defined as “*Internal migrants, as identified in the census, are, therefore, specifically defined as those who are usually resident in a civil division at the census reference time and who have previously been resident in another civil division within the country*”. This is slightly problematic for IDPs who need not cross an administrative boundary in order to qualify as an IDP (IRIS, 2020 paragraph 115). To capture these IDPs who do not move administrative areas a direct question about displacement may be required for all individuals (see above 49).

99. In paragraph 689 of the UNECE Recommendations¹⁵ “Internally Displaced Persons (IDPs) are defined as nationals or citizens who were ‘forced migrants’ within their country of residence together with their dependents living in the same household at the census reference

¹⁵ http://www.unecce.org/fileadmin/DAM/stats/publications/2015/ECECES41_EN.pdf

time, including children born after the forced migration took place". This differs from these recommendations, as family members who were not themselves displaced are not treated as IDPs, although children with at least one IDP parent are allocated to the category 'IDP-related'.

c. Country Examples of Census Questions for Identifying Refugees and IDPs

100. Many countries have already taken steps towards identifying refugees. In the 2010 census round, there were 14 countries that asked questions on residents' status, including refugee status. Details of the types on questions asked can be found in *Table 2.15: Refugee questions in selected censuses of population and housing* below, and in Table 4.1 in IRRS.

- Morocco and Algeria had special questionnaires for their nomadic population.
- 35 countries asked the question 'reason for migration' and 14 countries provided response options referring to "involuntary return", or "displacement".
- Census data from Kenya 2009 census include conventional households, unconventional households (i.e. group quarters and those included in special populations), and households in refugee camps.
- Bosnia and Herzegovina (see Case Study Part II of this Manual, paragraph 170) includes questions about citizenship, displacement and return. It asks about residence at the time of the 1991 census which took place a year before the armed conflict in the country.
- Cameroon collected information on refugees and IDPs in its fourth General census of Population and Housing (see Case Study Part II of this Manual, paragraph 554 onwards)
- Georgia during its 2014 General Population Census asks directly about IDP or refugee status and asks about previous place or country or residence (see Case Study Part II of this Manual, paragraph 580 onwards).

Table 2.15: Refugee questions in selected censuses of population and housing

Country	Year	Question	Responses
Belarus	2009	Why did you arrive to the Republic of Belarus for the permanent abiding place? (Indicate only one main reason)	6. Asylum seeking
Bosnia and Herzegovina	2013	Was the person a refugee from BIG (after 30 April 1991)? Has the person returned from refuge? Is the person still formally-legally considered a displaced person in BIH?	Yes/No
Burkina Faso	2006	Motif de la venue ou du retour: Pourquoi (NOM) est-il venu ou revenu pour la première fois au Burkina?	3. Expulse situation irreguliere
Burundi	2008	Former household member leaving Burundi since 1/10/1993 – Country of destination, reason for move.	4. Family reunification 6. War, crisis, conflict or insecurity

Compilers' Manual Part 1 - Phase 2A Design Population Census

		Has anyone returned to the household from abroad or another settlement before August 2000? Reason for displacement, year of return, place returned from	1. Repatriation 2. Voluntary repatriation 3. Refoulé (repression) 4. Returned to a camp 5. Resume studies 6. Other
Cayman Islands	2010	Which one of the following best describes ...'s status in the Cayman Islands?	8. Non-Caymanian – Asylum holder/seekers
Cote d'Ivoire	2014	<i>Residents born before September 2002</i> Were you ever forced to leave your place of habitual residence because of war or armed conflict? When did you leave? Place of habitual residence. Do you intend to or have you returned to your place of habitual residence?	Yes/No 2002, 2010/11, or 2002 & 2010/11 Returned (Year), intend to return, stay here, go elsewhere.
Djibouti	2005	Years at place of residence. Last place of residence, reason for move.	Professional reasons (hiring, transfer, establishment of business), urgent reasons (drought, flooding, food shortages, war), personal reasons (family reunification, health reasons), school reasons, seeking amenities
Ethiopia	2018	Reason for migration	1. Search for job 2. Join family 3. Education 4. Marriage/divorce 5. Drought/environmental degradation 6. Dispute/conflict 7. Health 8. Other
Georgia	2014	Are you an IDP or refugee? If yes, where are you IDP or refugee from?	Yes/No; free response
Greece	2018	Reason for settlement in Greece	Work, Repatriation family, Reunification, Studies, Asylum seeking, Refugee, Other reason (specify)
Jordan	2015	Asylum status (for non-Jordanian persons): Is (name of person) registered as a refugee in records of the government of Jordan or UNHCR?	1. Registered and has document; 2. Apply an application; 3. Unregistered; 8. Don't know
Kazakhstan	2009	Purpose to visit to Kazakhstan Refugee status	5. Refugee Yes/No
Kenya	2019	Why did <NAME> move to current place of residence?	8. Refugee; 9. Asylum Seeker

Liberia	2008	Displacement. Has...been displaced by war since 1990? Hasbeen resettled	Yes/No Yes/No
Mauritania	2013	What is the main cause of the arrival of (name) in Mauritania?	Refugee/Asylum
Sudan	2008	Population Group/Type of Household	Private Household, Nomads, Internally Displaced, Institutional Household, Homeless, Refugees, Cattle Camp, Overnight Travelers
Somalia (population survey)	2013/14	What was the main reason the household left its place of origin?	Insecurity, drought, floods, loss of livelihood, economic opportunities, access to services, other
Zambia	2010	What was the main purpose of your stay in Zambia?	5. Refugee/Asylum

101. There are several countries who have introduced specific IDP question content into their Census of Population and Housing. Several examples can be found in the *Table 2.16: IDP questions in selected censuses of population and housing* below.

Table 2.16: IDP questions in selected censuses of population and housing

Country & Year	Year	Census Question/Topic	Response Category
Azerbaijan	2009	For those who changed the former place of residence: refugee (forced migrant); Since when has (...) been living here? Place of former residence? (state, region)	Year; State, Region;
Bosnia and Herzegovina	2013	Was the person displaced in B&H (after 30 April 1991)? Has the person returned to the settlement he/she was displaced from? Is the person still formally-legally considered displaced person?	State settlement and municipality the person was displaced from; Yes/No; Yes/No;
Colombia	2018	The main cause for changing your place of residence on that occasion was:	Difficulty in finding a job or shortages in your means of subsistence, Risk or as a consequence of natural disaster (flood, landslide, earthquake), Threat or risk to your life, freedom of integrity, due to violence, Education needs, Health reason, Family reasons, Member of a nomadic community, other reason;
Côte d'Ivoire	2014	Have you been forced to leave your place of usual residence due to war or armed conflict? When did you leave your place of usual residence? Where was your place of usual residence? Have you returned to or do you have the intention to return	Yes/No 2002, 2010/2011, 2002 and 2010/2011 Name of S/P or commune Returned already (place),

Country & Year	Year	Census Question/Topic	Response Category
		to your usual place of residence before the war/crisis?	Intend to return, Intend to remain here, Intention to move elsewhere
Kosovo	2011	What was the main reason why you moved to your current place of residence?	Employment reasons, Education or training, Family reasons, 1998-1999 conflict reasons, Other reasons
Philippines	2007 (not in 2010)	Questions for households residing in temporary relocation or settlement area: Current residence in temporary relocation area; Date moved to current residence; Previous residence; Intention to return to previous residence within one year?	Yes/No Month/Year Barangay, City/Municipality/Province Yes/No
Somalia (Population Estimation Survey, not a full census)	2014	What was the main reason why this household left its place of origin? What are your plans regarding the place of origin for the coming year?	Insecurity; drought, floods; total loss of livelihood (destitution); better economic opportunities elsewhere; lack of access to services; other Stay here in this current location; return to place of origin; temporarily return to place of origin; relocate to another place permanently; have no plan

d. Hybrid census techniques

102. A hybrid census produces population estimates for small areas or for uniform, detailed grids in the absence of a traditional national census. Hybrid censuses rely on complete counts of population within small, defined areas, through 'micro-census surveys' selected across an area. These data are collected relatively rapidly and at a fraction of the cost of a full national census. Statistical models are then used to link these micro-census data to spatial data with full coverage to predict population numbers in the unsampled locations. Population totals can also be produced for administrative units or for the national level by aggregating these high-resolution predictions.

103. This approach can never replace the data generated by a traditional population and housing census, which provides full coverage of data on individuals, families, households and communities. However, where a traditional census cannot be fully executed in all locations of a country due to insecurity or other concerns, then this hybrid approach can produce population estimates for small areas in the absence of traditional census data.

104. Population data for a sample of areas across the area or country are needed as a primary input for a hybrid census. These data may come from a partial census, a census-like population survey or a specifically designed micro-census survey. Robust geo-referencing of the geographical areas where population data has been acquired is also a requirement.

105. The design must capture the range of population densities, demographics and

environmental dimensions that exist across the area. Enumeration can be conducted within administrative units or within other arbitrarily designed polygons, as long as the population data is explicitly linked to the correct geographical area. Global Positioning Systems (GPS) can be used to ensure 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. Depending on the questions included in the micro-census, disaggregation for certain population groups can be achieved.

106. There are several limitations to the approach:

1. Micro-census data must be collected with the same care and rigor as a full census.
2. Population data are highly political and contentious, affecting all per capita rate estimates, shifting political representation and changing claims to power or resources. Conflicts, environmental hazards, or large population displacements make it difficult to accurately collect information on the dynamic population, and misrepresenting populations at risk should be avoided.

Box 2.2: Afghanistan hybrid census

Afghanistan Census:

The last population census in Afghanistan was conducted in 1979 and a new census is not foreseen until after 2020.

To fill this gap, UNFPA, Flowminder/WorldPop and the Afghanistan National Statistical Office have worked together to develop an innovative methodology to estimate subnational population counts.

By combining survey, GIS and satellite data into a statistical model to predict numbers in areas where no population data were available, it was possible to generate high resolution maps of population estimates disaggregated by age and sex, together with uncertainty metrics. The data will be used to support decision making and the measurement of the 98 SDG indicators that require population data. The collaboration has also contributed to further strengthening capacities of the Central Statistics Office and other related government agencies to generate population estimates based on the integration of geospatial data and ground surveys.

2.4 Design Frame

a. Dynamic population mapping

107. In order to ensure coverage of refugee and IDP population in a census it is essential to have an image of the places where the people of interest, the refugees and/or IDPs are currently concentrated as mapping exercises can become outdated very quickly in times of displacement. A very useful method is to collect data from mobile phones using call connection records. From there we can have up to date information about sudden changes of population density in small areas. With this method it is possible to have also a picture for the size of the new refugee or IDP settlement which can help in planning census operations.

2.5 Design Processing and Analysis

a. Quality Assessment – the Post Enumeration Survey

108. Assessing the quality of population and housing census data is an integral part of the census operation. UNFPA advocates for more 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 with regard to potential underestimation or overestimation of the size of the permanent population as well as to the capturing of the characteristics of the population.

109. Of the 134 countries or territories responding to a UN Statistics Division (UNSD) questionnaire on census implementation in the 2010 census round, 89 had conducted a PES to evaluate the coverage of the census; among them 75% used the PES to evaluate content errors. In Africa, Asia and South America, 78% of countries undertook a PES, compared to only 40% of countries in Oceania. One third of the countries implementing a PES used the results to adjust census figures.

110. Any census Post Enumeration Survey should look at the coverage of refugees and IDPs 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 institutional or other places where refugees and IDPs may be staying. Special attention must be paid for the demographic characteristics and the refugee status of the persons concerned. This may need a special design component within the PES

111. Such assessments need to consider two kinds of quality measures, namely:

1. Census coverage: This is the percentage of actual population/households enumerated. The most common problem is under-coverage. Although with populations on the move over-coverage is also possible. Over-coverage also happens, if local authorities, with an interest in exaggerating their populations, are able to influence the data collection. An aggregate undercount of 3 to 6% is considered acceptable, but some population groups tend to be disproportionately affected, e.g. children under low age, young mobile men and undocumented migrants.
2. Content errors: This is the extent of incorrect answers for some key variables. Individual data are often provided by an informant (the head of household or other adult household member) who may not know all the details regarding each person. The household members themselves may be unsure or misunderstand the question. They may also hide or distort information on purpose. As a result, data quality may be poor, despite excellent coverage.

112. There are several methods for quality assessment. Content errors can often be detected through internal consistency analysis. Inconsistencies (e.g. between children's ages and educational status, or young girls' ages and number of children, etc.) indicate poor data quality. But not all errors lead to inconsistencies.

b. PES Design

113. Depending on the design of the PES sample, the household identification can be more or less rigorous. The following designs can be considered:

- i. The PES sample consists of census enumeration areas that are completely re-enumerated in the PES. This design requires sufficient cases to provide results for

relevant sub-samples including displaced people). It is easier to administer than the other options and reduces the travel time taken by enumerators. There are several requirements for this kind of PES; the enumeration areas must be well defined, with explicit geographical boundaries, based on good census cartography; out-of-area enumeration in the census should be almost non-existent; and the interval between the census and the PES should be very short (1-3 months).

- ii. The PES sample consists of a larger number of census enumeration areas in each of which a certain fraction of households is sampled.
- iii. The PES sample contains no clusters, but simply re-enumerates every n-th household from a newly produced national household listing. This design is efficient from a sampling viewpoint, but it is difficult to administer.

114. To apply designs ii) or iii) it is essential to use a rigorous household identification system. In its absence, design i) is the only viable option and even this may be challenging if the conditions listed above are not satisfied.

c. PES Analysis

115. 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.

116. 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.

117. If the census is taken as a base line, the entire PES questionnaire has to be filled out according to the situation at census time and any changes have to be justified in terms of births, deaths and migrations, or changes of physical dwelling.

118. Although the number of changes should be small, especially if the PES is carried out very soon after the census, correctly classifying them can be time-consuming, especially if the information is found to be inconsistent at the time of matching. Take the case of a household member who is declared in the PES as having been temporarily absent from the household at census time, but who is present in the household at the time of the PES (alternative ii). The following possibilities exist:

- i. The person was declared as member of the household at census time and the census was organized according to the de jure criterion. This would be correct. The best practice (which is not attainable in countries without well-defined postal addresses) is to verify the census records of the place where this person was at census time to check if (s)he was enumerated there, to detect any potential double counting.
- ii. The person was declared as member of the household at census time and the census was de facto. This would be wrong. If the person was also enumerated at the place where (s)he spent census night, it would be a case of double counting. If not, it would be a classification error.
- iii. The person was not declared as a household member at census time and the census was organized according to the de jure criterion. Depending on the person's enumeration status in the place where (s)he was at census time, it would be a classification error or an omission.
- iv. The person was not declared as a household member at census time and the census was de facto. This would be correct. But in this case, it would be particularly important to verify if (s)he was enumerated elsewhere.

119. Deciding whether a person was enumerated elsewhere in the census also arises if the person permanently resided elsewhere at census time. The best information obtainable in this case may be based on his/her answer to the question, but verification may be impossible in all but the most sophisticated PES operations.

120. A slightly different situation presents itself in the case of persons who moved out of the household and dwelling after the census. In theory the PES should try to locate these individuals, to obtain their answers to the PES, to compare it to the census. Countries with a long history and high level of technical expertise in PES execution, such as Australia, actually collect this information. But in the PES processes of most developing countries this is impossible, and one has to assume that the PES data of such persons coincide with what they declared in the census.

Data Source B: Sample Surveys

2.3 Design Collection

121. This section includes recommended modules and modes of collecting data from probability samples, most often household surveys. The type and cause of displacement should be considered in designing a survey. This section covers additional questions which could be added to existing household surveys in order to identify the displaced, and a displacement module which could be added to surveys to collect information specifically about the displaced.

a. Refugee questionnaires and methodology

122. There are several examples of methodologies for collecting refugee statistics from household surveys. There are two possible methodologies, the first is to design a specialist refugee survey, which could be part of a migrant survey and example of this is the [MED-HIMS survey](#) (see also Case Study in Part II of this Manual) designed for Mediterranean countries; the second is to include refugee screening questions in a more general household survey such as a Labour Force or Living Conditions Survey. The [European Union Labour Force Survey had a special module in 2014](#) designed to collect data on the labour market situation of migrants and their descendants and includes questions to identify and describe the situation of refugees and asylum seekers. It asks if respondents migrated as a *refugee or seeking international protection* in the previous 10 years and about obstacles to participating and integrating in the labour market.

123. Where data on stocks and flows, and comparisons with the living conditions of other groups in the population are required, then screening questions on a regular survey may be the most cost-effective methodology. This will depend on the incidence and distribution of the refugee and refugee-related population in the country and will require the input of a specialist sampling expert to determine if adequate samples of the target population of refugees are likely to be obtained.

124. Where detailed information about refugees and related populations is required, then a specialist survey may be the preferred option, but this likely to expensive and carried out rather infrequently. It also depends on having a good quality sampling frame from which to draw the sample.

Box 2.3: Jordan: The living conditions of Syrian refugees in Jordan: Results from the 2017-2018 survey of Syrian refugees inside and outside camps

Sampling Summary

Drawing on a survey implemented by the Department of Statistics (DoS) between November 2017 and January 2018, the report presents recent statistics on Syrian refugees residing in Jordan. It presents findings for six geographic localities: Amman; Zarqa; Irbid; Mafraq; the other governorates taken together; and the refugee camps.

The report is based on information from 7,632 households and 40,950 individuals. Sampling was based on the DoS sampling frame constructed on the 2015 population census. A total of 1,121 clusters (locations) outside camps and 82 clusters inside camps were randomly selected. The sampling design was not geared towards estimating the number of Syrian refugees in the Hashemite Kingdom, but instead aimed for efficiency with regard to describing the Syrian refugee population in accordance with project objectives.

A household is defined as a unit that pools its resources together, and whose members usually sleep and eat (most meals) together. Usually, the household members are immediate or more distant relatives, but they do not have to be related. This report defines a refugee household as one where the head of household is a Syrian refugee. It understands a Syrian refugee to be any Syrian individual who fled to Jordan from Syria as a consequence of the crisis and war there and arrived in Jordan after 15 March 2011, and who acknowledges or defines him- or herself as a refugee; any Syrian national who resided in Jordan before 15 March 2011 and became a refugee because he or she could not return to Syria; or any children of these two categories of people born after 15 March 2011. The vast majority of these self-ascribed Syrian refugees, 97 per cent, have formalized their refugee status by registering with the UNHCR.

Source: <https://reliefweb.int/report/jordan/living-conditions-syrian-refugees-jordan-results-2017-2018-survey-syrian-refugees>

b. IDP Survey questionnaires

125. IRIS, 2020 recommends adding a set of screening questions to national household surveys similar to those recommended for population census (see paragraph 94 to 95) 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. Suitable questioning will be developed and included in future editions of this Manual.

126. To measure IDPs' living conditions globally, and particularly whether they have overcome key displacement related vulnerabilities, a core module will be developed and proposed as a standard approach for periodic use in existing national multi-topic household surveys such as the Living Standards Measurement Survey (LSMS), Demographic Health Survey (DHS), Multiple Indicators Cluster Surveys (MICS) or other multi-purpose surveys.

127. In order to maximize the quality, utility and comparability of data collected globally, it

is recommended that the model questionnaire to be recommended is used to develop, design and test national survey instruments, particularly related to questions on key displacement related vulnerabilities. Maintaining consistency in questions over time also allows for the creation of time-series for IDP statistics.

128. There is some recent research work on methodologies improving the estimates of consumption obtained from displaced households and avoiding misreporting.

“Indeed, survey respondents in IDP camps may believe that their responses will influence the provision of humanitarian aid and will thus misreport consumption in an attempt to influence its distribution. If survey respondents are underreporting, the inaccuracies generated in the data are highly problematic. At best, it makes the data spurious and unusable. At worst, it could lead to misallocations of aid, from more vulnerable areas to less vulnerable areas, or from solutions emphasizing sustainability to immediate relief when immediate relief is unnecessary. Given this context, light touch adaptations to the design of the survey that prime the idea of honesty offer to make big improvements to the quality of the data and support provisions the data informs.¹⁶

c. Country Examples of Questionnaires found in PART II

129. Cameroon example of survey questions for identifying refugees and IDPs (See Part II of this Manual paragraph 558)

130. An example of this modular design is the *2017 South Sudan High Frequency Survey and Conflict Recovery Survey* which has a number of modules. The Module I has 4 subsections, two of which are reproduced in The South Sudan Country Case Study in Part II paragraph 617) of this report.

MODULE A: Interview and Household Information - Questions: 4

MODULE B: Household Roster Questions: 99

MODULE C: Household Characteristics - Questions: 132

MODULE D: Food Consumption - Questions: 108

MODULE E: Non-Food Consumption - Questions: 13

MODULE F: Livestock - Questions: 13

MODULE G: Durable Goods - Questions: 24

MODULE H: Wellbeing and Opinions - Questions: 41

MODULE I: Conflict and Displacement - Sub-sections: 4, Questions: 69

MODULE J: End of Interview - Questions: 9

MODULE K: Enumerator Feedback - Questions: 12.

¹⁶ https://link.springer.com/chapter/10.1007/978-3-030-25120-8_11#Fn3. This chapter is a summary of Kaplan, Pape, and Walsh (2018, forthcoming), “Eliciting Accurate Responses to Consumption Questions Among IDPs in South Sudan Using “Honesty Primes”, Policy Research Working Paper Series. See also open source ebook The World Bank and Johannes Hoogeveen, Utz Pape Editors. *Data Collection in Fragile States Innovations from Africa and Beyond* <https://link.springer.com/content/pdf/10.1007%2F978-3-030-25120-8.pdf>

2.4 Design Frame

“This only applies to processes which involve data collection based on sampling, such as statistical surveys. It identifies and specifies the population of interest, defines a sampling frame (and, where necessary, the register from which it is derived), and determines the most appropriate sampling criteria and methodology (which could include complete enumeration).

Common sources for a sampling frame are administrative and statistical registers, censuses and information from other sample surveys. This sub-process describes how these sources can be combined if needed. Analysis of whether the frame covers the target population should be performed. A sampling plan should be made: The actual sample is created in sub-process 4.1 (Create frame and select sample), using the methodology, specified in this sub-process”. (UNECE, 2013)

a. Sampling Frames

131. A sample is a subset of units drawn from the target population of interest. It is not usually convenient, efficient or possible to examine every member of an entire population. Sampling reduces the resources needed and, if done properly, provides estimates of the population characteristics with known accuracy. Sampling is the process of selecting units (e.g., households or individuals) from a population of interest so that by studying the sample we may generalize our results back to the population from which units were selected.

132. Sampling design describes how the final sampling units are selected and reached for the study. It is determined by the substantive goals of the study, and by resource and logistical constraints. There are different forms of designs in sampling.

133. Probability-based sampling ensures that every unit in the population has a known chance (greater than zero) of being selected in 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.

134. Non-probability-based sampling is any sampling method where we do not know the chance respondents have to be selected to the sample. In most cases some elements of the population have no chance of selection because the sampling strategy is based on decisions related to convenience, quotas or deliberate choice of elements of the population. In these cases, the probability of selection cannot be determined. Because the selection of elements is non-random, non-probability-based sampling does not allow the estimation of sampling errors, and it is not possible to assume that biases cancel each other out. When using non-probability based sampling the ability of making inferences about the population is limited, and the information generated needs to be treated carefully. Some adaptations of non-probability-based sampling have been made to allow for limited inferences. An example of this is Respondent-Driven Sampling, (RDS) or Snowball Sampling that propose ways of estimating weights to try to account for biases. See the example below from Turkey of a two-stage sampling methodology combining spatial and respondent driven sampling.

135. There are sampling constraints which should be considered when designing a sample survey of displaced persons. Displaced people may be a small proportion of the population, and it is likely that they live in specific locations and are not evenly distributed around the country or region. Any sample must be of a sufficient size and design to yield reliable results. Each survey has its own sampling design, even if it is similar from most national household surveys. Using the sampling design for an existing national survey will introduce a selection bias, if it is not amended to take account of the living circumstances of displaced people.

Refugees and IDPs can live either in camps specially designated for displaced people or they can live in houses among the general population, either as part of another household or as an independent household.

136. In Somalia where security considerations prevented the usual sampling techniques to be carried out an innovative technique was used¹⁷, (see Utz Pape and Philip Wollburg). To address security concerns, the Somalia High Frequency Survey adapted logistical arrangements, sampling strategy, and questionnaire design to limit time on the ground. In logistical arrangements, a detailed and timely security assessment ensured that the enumeration areas to-be-visited were safe on the day of fieldwork. The fieldwork protocol was designed such that teams would spend as little time as possible in any given region and draw little attention, ensuring enumerator and respondent safety. Concerning sampling strategy, 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 are small enough for enumerators to list and select households immediately before conducting the interview. (see also paragraph 351 and The High Frequency Survey in South Sudan¹⁸ also described in Part II of this Manual - South Sudan.

¹⁷ Utz Pape and Philip Wollburg. "Estimation of Poverty in Somalia Using Innovative Methodologies." World Bank Policy Research Working Paper 8735, February 2019

¹⁸ <https://microdata.worldbank.org/index.php/catalog/2914>

Box 2.4: Turkey example: Two-Stage Sampling Methodology Combining Spatial Survey and Respondent-Driven (Snowball) Sampling

Context: WFP was interested in a vulnerability assessment of the refugee population in Turkey.

Challenge: Difficult to identify population (in this case, unregistered refugees), that are likely not randomly distributed geographically and have no comprehensive sampling frame.

Methodology: The study used a two-stage sampling methodology.

The first stage was a Simple Spatial Survey Method (S3M) that sought to select settlements evenly (rather than randomly) across Turkey's geography. Using R's spatial-sampler function, the researchers superimposed a hexagonal grid over the country, and then selected the refugee settlement that was closest to the nodes of the grid. It also stratified by urban/rural sites, although this stratification was later abandoned due to difficulty of reaching rural respondents. This grid-based sampling method lacks the benefits of true randomness but is acceptable if there is a belief that there are significant differences in geography that might not be adequately captured by a random sample.

The second stage was Respondent-Driven Sampling, which combined a snowball sample and probability weighting based on network theory. First researchers selected 2-3 seed households close to each of the geographic points (seeds) selected by the S3M model. The seed households were selected based on having strong social networks and willingness to participate, but researchers also tried to select households of different ages, genders, aid eligibility, and socioeconomic status. While this likely resulted in a more diverse set of initial households, if this stratification was conducted by field researchers in an ad hoc manner it may have unwittingly biased the sample. The researchers then used a snowball sampling method – after interviewing the first round of seed households, they asked each household to identify 2-3 additional refugee households who they could speak to. This process repeated at each subsequent step until the researchers had selected 25 respondent households from each seed. Based on the original 52 sampling sites, this generated a sample size of approximately 1,300 households.

Researchers then used the RDSAT application to analyse information about the respondent's social network in order to generate an individual weighting value for each respondent that adjusts for their likelihood of being selected. This weighting process uses network theory to adjust for the non-random nature of the households who are more likely to be selected by this sampling method, and to arrive at a post-weighting sample that should be more representative of the general population. Analysis can then be performed on the post-weighting sample.

Conclusion: While this methodology can be used in situations in which there is no available frame and populations are hidden, it has several sources of bias that make it inferior to true probability sampling. For instance, there is a tendency for respondents to refer households who are similar to them in non-representative ways (e.g. select households from the same ethnic group). While the RDS application can help correct for some of this bias, it is important to ensure that the original seeds are representative of the population in a rigorous manner, and that this method not be used when more comprehensive sampling frames are available.

Box 2.5: Enumeration and Profiling of IDPs, refugees and migrants in Hargeisa, Somaliland (December 2015)

Context: Somaliland had developed a policy framework on internal displacement in 2014, but still needed comprehensive and reliable data in order to inform its implementation by stakeholders.

Challenge: Study was starting from a very low-information baseline, and had difficulties getting access to a representative sample of the population due to resistance within the camp populations and the dispersed nature of the urban population.

Methodology:

Given the dearth of information available to researchers at the start of the study, the first stage of sampling was to develop a credible map of the population in the various settlements through information-gathering meetings with community focal points (imams, elders, teachers, community organizations). The next stage, enumeration, sought to reach all households in the camps with a brief 12-question census to establish membership to person of concern group (IDPs from Somaliland, Refugee returnees, Economic migrants, etc.). This census succeeded in reaching nearly 12-thousand households, and provided them each with a barcode that, along with phone numbers, was later used for randomization. Thus, the listing exercise resulted in multiple sampling frames of household as the final sampling units for the different populations of concern.

From the enumerated list of 11,962 households, 2,510 were selected to be surveyed. The randomization was stratified by person of concern category, so that the selected sample was proportionate to the population totals. Non-probabilistic sampling was applied to groups that did not have a population size large enough to support randomization (asylum seekers and IDPs from South Central Somalia).

As a result of the population size of some groups being small, as mentioned in the previous paragraph, the study also used non-probabilistic snowball sampling methods for the IDPs from South Central Somalia who lived in urban areas outside the camp, asking each respondent to refer additional households for the study. Due to the non-random nature of this selection mechanism, this cannot be considered a representative sample, although it may give a qualitative sense of overall IDP sentiments.

Limitations:

Due to community resistance to being categorized as IDPs, the enumeration stage was asked to conclude before it was able to reach all households living in the settlements. Thus there is a disconnect between the sample frame and the population of interest. There were also challenges with accessing refugees and asylum seekers directly – many declined to be interviewed, and others had had their enumeration barcodes taken by their landlords. While both of these elements contribute to a higher risk of bias within the sample, they speak more to the logistical difficulties inherent in working with IDP and refugee populations rather than any avoidable methodological problem with the study design. All possible mitigation steps should be taken to avoid these types of issues, but only if they can be addressed without jeopardizing the safety of researchers, the well-being of respondents, and the goodwill of communities.

There were also some challenges identifying the categories that were most appropriate for households, particularly between economic migrants and IDPs. If possible, it is important to establish clear and consistent guidelines whenever a categorization methodology is to be used.

137. Several types of frames could be considered:

- i. Using list of displaced people drawn from administrative sources.
- ii. Those drawn from a population census, a surveys or satellite imagery, or a combination of all three.
- iii. Those drawn from operational data sources such as the Displacement Tracking

Matrix (IOM-DTM) (e.g. “Comprehensive Food Security and Vulnerabilities Analysis” (CFSVA) 2016 in Kurdistan Region of Iraq.

- iv. Innovative sampling methodology (e.g. example above of Turkey Two-Stage Sampling Method applied by WFP for a study of their Emergency Social Safety Net programme in 2018¹⁹).

138. The selection of a sampling methodology depends on a number of factors:

1. The type of population to be covered and their mobility within the population. IDPs are usually more mobile and often not registered with the authorities, versus refugees who are likely to be more stable and often registered with the authorities once an application for asylum has been made and processed.
2. Their living arrangements, those living in camps will be easier to find, versus those living outside of camp, and where these are located.
3. The temporal aspect of displacement event (recent versus protracted)
4. Size of displacement population in comparison to resident population

139. The geographical scope and characteristics of the target population should be determined. This involves describing where the members of the target population are located, defining the boundaries of the area they are in (e.g. administrative or natural boundaries), and, if relevant to the study, adding any other information that characterises the physical space of the target population such as natural hazards or political situation.

140. The target population should further be narrowed down by determining what differentiates them from the general resident population of the study area (unless of course the latter is the target population). This can for example be descriptions such as “IDPs in urban settings in Colombia” or “refugees living in a camp”. Vulnerabilities, specific age groups, socio-economic characteristics, displacement status and legal immigration status are examples for such a description.

141. Especially in a volatile and fast-changing setting like forced displacement, a clear definition of the time frame or point in time covered by the study with respect to the target population is important in order to develop a well-defined sampling scheme.

b. Challenges in defining the target population and constructing a sampling frame and sample size

142. **Ineligibles:** Ineligibles are elements of the general population who are not part of the target population of the study. These may be non-IDPs who are erroneously selected in a sample of IDPs

¹⁹ <https://www.wfp.org/publications/turkey-emergency-social-safety-net-programme-vulnerability-profiling-2018-analysis-results>

143. **Inaccuracies:** Sampling frames obtained from previous surveys, censuses or administrative data might contain outdated information. For example, when using 5-year old census data of urban and rural areas to draw a representative sample of the population, we need to carefully assess whether population growth rates have been different for rural and urban areas during these 5 years, or whether there have been large population movements caused by displacements. If there have been more people have been moving from the countryside to cities than the other way around, but our sampling frame does not account for it, then using the old census data would lead to under-sampling of urban areas.

144. **Missing information:** In the context of forced displacement, it can be difficult to obtain information to define target population and construct sampling frames, especially for unregistered out-of-camp populations. In such situations, alternative information sources should be considered. These could include but are not limited to; previous studies, scoping exercises, qualitative assessments, key information interviews, desk review, community-based methods to build sampling frames.

145. **Multiple listing:** Multiple listing denotes the situation where a final sampling unit such as a household gets listed more than once in the sampling frame. This can be a problem for example in respondent-driven non-probability sampling frames where listed households are asked if they know of other households that fit the target population. One household can easily be mentioned more than once. Sampling frames need to be carefully assessed and cleaned to ensure no final sampling unit appears more than once.

Sampling Examples and References

- Afghanistan Hybrid Census (See *Box 2.2: Afghanistan hybrid census*).
- Jordan survey of Syrian refugees (see <https://datacatalog.worldbank.org/dataset/jordan-survey-syrian-refugees-and-host-communities-2015-2016>)
- MED-HIMS Surveys – see Part II of this Manual Paragraph 698
- Enumeration and Profiling of IDPs, refugees and migrants in Hargeisa, Somaliland (December 2015) (see *Box 2.5: Enumeration and Profiling of IDPs, refugees and migrants in Hargeisa, Somaliland (December 2015)*)
- South Sudan High Frequency Survey – sampling methodology and Conflict and Displacement Module Questionnaires, see Part II of this Manual Paragraph 617
- Turkey example of Snowball sampling (See *Box 2.4: Turkey example: Two-Stage Sampling Methodology Combining Spatial Survey and Respondent-Driven (Snowball) Sampling*).
- *Data Collection in Fragile States, Innovations from Africa and Beyond*; Editors Johannes Hoozevee and Utz Pape <https://link.springer.com/book/10.1007%2F978-3-030-25120-8>

Other Links to survey design:

- i. UNHCR Sampling Decision Assistant: <https://unhcr-sampling-assistant.firebaseio.com/#/home>
- ii. UNHCR Integrated Framework for Household Survey (IFHS): A toolkit to facilitate design, collection & analysis: <https://unhcr.github.io/Integrated-framework-household-survey/>
- iii. UNSD Household Sample Surveys in Developing and Transition Countries
https://unstats.un.org/unsd/publication/seriesf/Seriesf_96e.pdf
<https://unstats.un.org/unsd/demographic/sources/surveys/Handbook23June05.pdf>
- iv. UNSD Designing Household Survey Samples: Practical Guidelines:
<https://unstats.un.org/unsd/demographic/sources/surveys/handbook23june05.pdf>
- v. [JIPS Profiling Methodology:](#)
- vi. JIPS Essential Toolkit: <https://jet.jips.org/phase/designing-the-methodology/>
- vii. Durable Solutions: <https://inform-durablesolutions-idp.org/>.
- viii. IHSN Question Bank: <http://dev.ihsn.org/qbank-sandbox/web/>
- ix. MICS Surveys: <http://mics.unicef.org/tools>
- x. LSMS Surveys: Measuring Conflict Exposure in Micro-Level Surveys
http://siteresources.worldbank.org/INTSURAGRI/Resources/7420178-1294259038276/Measuring_Conflict_in_Micro_Level_Surveys.pdf
- xi. LSMS Surveys: <http://surveys.worldbank.org/lms/c4d2/refugees-migrants>
- xii. DHS Surveys: <https://dhsprogram.com/What-We-Do/index.cfm>, [DHS Methodology](#), [DHS Capacity Strengthening](#)

Data Source C: Administrative Data Including Population Registers

2.3. Design Collection

146. There are several potential sources of non-survey data on displaced people, the first is national population registers, the second is administrative data maintained by governments for the purpose of administering a range of services, controlling migration or issuing documentation, and the third is operational data kept by humanitarian agencies and governments for the purposes of supporting displaced people. This section gives examples of population registers, supplemented administrative records; and describes the operational data maintained by UNHCR about displaced populations and the way it is used.

147. Eurostat, in its Reference and Management of Nomenclatures system (RAMON) defines administrative data as a “Database which is updated continuously (often for administrative purposes, such as population registers or building registers) and from which statistics can be extracted / aggregated / computed”.²⁰

148. Administrative data are a useful source of information for measuring both the stock of displaced populations, and potentially the flows. 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 provide potential for integrating a number of sources (See next section Data Source D: Integrated Data). This can potentially be a rich source of statistics.

149. Very diverse types of primary administrative data may be of interest from a statistical point of view, these include:

1. Specific registers maintained by administrations or organisations responsible for refugees and IDPs (including registration of displaced people in camps and/or elsewhere).
2. Databases maintained by humanitarian organisations supporting refugees in informal settlements. These are often linked to the beneficiaries accessing displacement specific services and benefits.
3. General population registers of the total resident population of a country and/or registers of foreign citizens.
4. Other administrative registers (residence permit registers, tax registers, social security registers, register of border crossings etc.).

150. Such administrative registers are, in particular, 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,

²⁰

https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=DSP_GLOSSARY_NOM_DTL_VIEW&StrNom=CODED2&StrLanguageCode=EN&IntKey=16702785&RdoSearch=BEGIN&TxtSearch=register&CboTheme=&IsTer=&IntCurrentPage=1&ter_valid=0

they are also the main source of statistics on the number of refugees living in the country through specific sub-categories of reasons to stay in the residence permits statistics (refugees legal status, subsidiary protection, humanitarian reasons) and statistics on persons who are the subject of pending applications for international protection (asylum seekers whose application is still being examined).

151. In many countries without the current capacity to maintain administrative registers, UNHCR maintains registers on behalf of the government until the necessary capacity is developed. Arrangements can be made for accessing these registers for statistical purposes.

152. A number of countries keep population registers, and several examples of these are described in this manual. 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, and if they are then they may not be identifiable as IDPs because their movements have not been captured.

153. According to these recommendations, two types of administrative sources should be considered for the collection of migration statistics:

1. Administrative registers, defined as "a data system providing for the continuous recording of selected information pertaining to each member of the target population" and including "population registers, registers of foreigners and other special type of registers covering particular groups of persons, such as registers of asylum seekers" (United Nations, 1998, p. 17).
2. Other administrative sources derived "from the operation of administrative procedures designed to control international migration" including sources related to "the issuance of residence permits", "the issuance of work permits", "applications for asylum", "tax or social security", and "border collections" (United Nations, 1998, pp. 20, 23).

154. To be useful for IDP statistics the system must be accessible to those who have been displaced and be regularly updated to reflect recent movements, changes of circumstances and to reflect flows in the stock by either adding new cases or removing those who have exited from the stock.

155. For refugees and refugee related populations asylum seekers are often not treated as part of the usually resident population and are therefore omitted from the population register. Registers of asylum seekers or foreigners may also be required to complete the statistical estimates, (see Part II of this Manual: Hungary see para. 590590, and Turkey Case Studies see para. 631 and 655).

156. 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.

157. Two types of units are used to record applications, "cases or applications" and "persons [covered by these cases or applications]" and consequently to be registered.

a. Population Registers

158. Part 2 of this Manual provides a number of Case Studies, and several case studies of Population Registers as the basis for displacement statistics are described briefly below, more information can be found in Part II.

1) HUNGARY (SEE PART II HUNGARY)

159. In Hungary both the National Directorate-General For Aliens Policing (NDGAP) and the Hungarian Central Statistical Office (HCSO) publish official statistics on persons involved in forced migration. National Directorate-General for Aliens Policing (NDGAP) is the authority which is responsible for the asylum procedure in Hungary. The NDGAP produces statistics on the asylum procedure, on numbers of persons taken under international protection status (refugees under the Geneva Convention and persons taken under subsidiary protection) and the figures of the persons who receive national humanitarian status (persons who are taken under the principle of non- refoulement).

160. The NDGAP decides on the asylum applications²¹, recognising refugees under the Geneva Convention and persons who are under subsidiary protection. Data is forwarded to the Ministry of Interior which runs the Population and Address Register (PAR). In Hungary the persons taken under international protection are issued with an identity card (and have similar rights to Hungarian citizens) and therefore their data is stored in the population register.

161. In contrast to asylum seekers who are not considered to be part of the usually resident population and are not included on the PAR, refugees and subsidiary protected persons are part of the usual residence population in Hungary and included in the PAR.

162. The Hungarian Central Statistics Office (HCSO) receives the aggregated statistical data on asylum applicants from the NDGAP on a monthly basis and the aggregated data on positive decisions (refugees, subsidiary protected, and persons taken under subsidiary protection) on a quarter-year basis.

2) NORWAY (SEE ALSO PART II NORWAY)

163. Refugee statistics in Norway are produced from administrative data. All asylum seekers must register with the Norwegian immigration authorities, and their case data and personal data are stored in a database maintained by the Norwegian Directorate of Immigration (UDI). The database contains data on UN convention refugees as well as refugees who enter Norway on their own and apply for asylum.

164. The UDI statistics division has access to copies of raw data from the database and has the ability to extract, aggregate and process the data in whatever way it chooses. At the same time, selected personal and case data are transferred directly from the case processing system to the Central Population Register (CPR), and from there transferred to Statistics Norway through daily batch jobs. UDI also provides supplemental, processed data to Statistics Norway

²¹ Since the establishment of the so called 'transit-zones' the asylum procedure takes approximately 30 days but even before the existence of the 'transit-zones' the asylum procedure did not take longer than 60 days which indicates that this population subgroup is not part of the usual residence population of Hungary.

in a data set that is compiled every year. The most important variable in this data set is each immigrant's reason for immigration.

165. An obvious advantage of this system is that the producers of official statistics on refugees – Statistics Norway and UDI – have excellent access to data as well as control of how the data is processed, analysed and disseminated.

166. A major challenge is that data collection first and foremost is designed and built for case processing, not to produce statistics. Hence, the main purpose of the data collection is to facilitate the flow of applications through the administrative systems, while keeping track of the applicants' legal status and rights. Many people are involved in the handling of each application, and the system is built to enable each case handler to finish tasks and transfer the case to the next person, through all the steps from application to final decision. For this purpose, it may not even be necessary to access or collect structured data, as many case processing tasks can be resolved simply based on case documents.

3) TURKEY REFUGEE STATISTICS (SEE PART II TURKEY)

167. All addresses within the boundaries of the country were registered in the National Address Database (NAD) in 2006, and by linking addresses of Turkish citizens living within the boundaries of the country with the Turkish identification number, Address Based Population Registration System (ABPRS) was established in 2007. Subsequently, in order to obtain information which was not available in the ABPRS at provincial level, Population and Housing Survey (PHS) was conducted using a combined methodology (by linking the ABPRS to a large-scale sample survey) in 2011. According to the Official Statistics Programme, annual information on population size by administrative division (province, district, town, village and quarter), and its basic characteristics (age-sex structure, place of registration, nationality, literacy and educational attainment, legal marital status, place of birth, type of households, and internal migration in province level) are produced from ABPRS and announced to the public in January of the following year.

168. TurkStat is responsible for producing stock and flow statistics on international migration by demographic and other related indicators of migrants. For foreigners residing in Turkey, citizenship and country of birth statistics are produced from ABPRS. In addition, TurkStat produced information on immigration and emigration, namely annual migration flow statistics, for the first time based on administrative registers in 2018 for the years 2016 and 2017. Similar to the statistics on foreign stock population, migration flow statistics covering both Turkish citizens and foreigners are produced in "country of birth" and "country of citizenship" breakdowns (no disaggregation by refugee or refugee like populations). In order to obtain the statistics on foreign population residing in Turkey, foreign registers held by the General Directorate of Civil Registration and Nationality (GDCRN) and the residence-work permit registers are mainly used.

169. Along with the Göç-Net (migration management system), registers of residence permits, international protection (applications, refugee, conditional refugee, subsidiary protection, and temporary protection under international protection), illegal migration and other events related to foreigners are kept in connection to each other. Directorate General of Migration Management (DGMM) is responsible for collection, analysis and dissemination of the data on international protection, residence permits and irregular migration.

170. Principal variables defined concerning applicants of international protection are age, gender, educational status, nationality and marital status etc. Statistical figures on international protection applications are shared on DGMM's official website and published in Annual Migration Reports of DGMM on a yearly basis.

171. With the "*Temporary Protection Regulation*" issued in 2014, Syrians seeking international protection were given temporary protection status in Turkey. Syrians are granted a foreign identification number to benefit from health, education, labour market access and social support services. Based on the legal (de-jure) population definition, Syrians under temporary protection are not covered in the Address Based Population Registration System (ABPRS). Thus, this population group is not officially accepted as "resident population". Only the Syrians who entered the country legally (with the required official documents such as a passport) are included in ABPRS population like other foreigners holding a residence permit/work permit. Briefly, while approximately 3.5 million Syrians under temporary protection are not covered in ABPRS, the number of Syrians in the ABPRS was just 87,955 as of December 31, 2018.

172. The number of persons under temporary protection in Turkey was 14,000 in 2012 reached 225,000 thousand in 2013, 1.5 million in 2014, 2.5 million in 2015, 3.5 million in 2017 and almost 3.6 million as of June 2018.

b. Refugee and IDP Registers

173. Registration data provides valuable information humanitarian agencies about the overall demographic and other characteristics of the population of concern, including population figures, levels, rates, averages and other patterns. When disaggregated, it also facilitates the identification of individuals and groups for support programmes and interventions, for designing, planning, targeting, delivering, tracking and monitoring protection and assistance interventions. The data is potentially available for statistical analysis if the appropriate inter-agency agreements are in place (see Phase 1.5 Check Data Availability in this Manual and Part II for a specimen 'Agreement on the Transfer of Personal Data of Refugees and Asylum-seekers' between UNHCR and a host government)

174. Registration activities can be a major source of population data, which are essential for humanitarian agencies programming purposes and of universal relevance to the entire humanitarian community. This section provides examples of registers maintained by both governments and the UNHCR registration system.

4) BOSNIA AND HERZEGOVINA (SEE PART II BOSNIA AND HERZEGOVINA)

175. Statistics on IDPs and refugees 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 from other entities and districts, and for producing statistical reports. The country's municipalities and cantons are responsible for collecting the data and for delivering it to the Ministry, together with The Ministry of Refugees and DPs of the Republika Srpska Entity; and the Government of Brčko District B&H, Department of Displaced Persons, Refugees, and Housing.

176. Producing statistics on refugees and IDPs was initiated by the government at the beginning of the war in the 1990s. When 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 completed on 31/03/2005.

177. The administrative records in the Database of Displaced Persons (DDPR) are the main source of statistics on refugees and IDPs, supplemented by census data. Details of the variables found in the DDPR can be found in the Part II Case Study, as can the census form which included specific questions on persons displaced since 1991. These questions include their habitual place of residence place of residence and returns and legal status.

5) COLOMBIA (SEE PART II COLOMBIA)

178. The National Planning Department and the National Department for Social Prosperity collect and produce official statistics on the country's most vulnerable population. These datasets account for most of the administrative records used by the Victims' Unit to update socio economic variables of IDPs at an individual level. Other relevant datasets are the official education enrolment system, the social security system, the registers' office, the housing and agricultural subsidies registers, among others.

179. 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. 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 has to declare within two years after the event of displacement took place. 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.

180. 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 is 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. Additionally, IDPs update their location directly to the Victims' Unit through services points located throughout the country.

181. 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.

182. 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.

183. For statistical purposes, an IDP is taken out of the stock when they overcome displacement related vulnerabilities.

6) UKRAINE (SEE PART II UKRAINE)

184. The Ministry of Social Policy in Ukraine is responsible for ensuring the creation and maintenance of a Unified Information Database (UIB) on Internally Displaced Persons. Registration of the place of residence of the IDP is carried out in accordance with the laws of 2014 and 2015. However, the State Statistics Committee of Ukraine receives only very aggregated information from the UIB, and is unable to use that data in its demographic calculations.

185. In addition, several studies have been made of IDPs in Ukraine. Most of them by humanitarian organisations providing assistance to IDPs. The studies do not generate official statistics. The International Organization for Migration regularly produce statistics on IDPs in Ukraine through their National Monitoring System, based on their Displacement Tracking Matrix approach. It collects and analyses information on the socio-economic characteristics of IDPs and IDP households. One challenge the IOM faces, is to have an updated sample for their survey.

7) UNHCR REGISTRATION SYSTEMS (SEE PART III FOR MORE INFORMATION AND PART II FOR A SPECIMEN DATA SHARING AGREEMENT BETWEEN UNHCR AND NATIONAL GOVERNMENTS (PAGE 224))

186. Registration data for 'Operational Planning and Programming' is required for designing, planning and delivering suitable programmes for the population of concern. Registration data analysis strengthens the overall operational intelligence for protection and solutions, helping to facilitate early identification of protection risks and highlights protection gaps as well as target resources for the greatest protection impact.

187. Registration data can only be usefully analysed if the data collected at registration is complete, of a good quality and regularly updated. A good registration system should also include mechanisms to manage its data, making it reliable and effective for planning and programming. There are many reasons why registration data may be of poor quality: data entry error, inadequate training for staff, ill-designed processes and insufficient oversight, cultural misunderstanding or weak standard operating procedures.

188. The personal data of persons of concern must be managed in a way that protects confidentiality and is consistent with privacy and data protection principles set out in UNHCR's [Data Protection policy](#). To this end, data repositories (databases) should be designed so that access rights for each content material are determined individually per user, based on a "need-to-know" and "need-to-use" basis. A framework on the segregation of duties aims to clearly define and grant the level of access that corresponds with the user's specific role and responsibilities, in line with the [Data Protection policy](#) (para. 4.2) and related [Guidance](#) (para. 6.3.8).

189. UNHCR's population, Registration and Identity Management Eco-System (PRIMES)²² contains a Data Port element that uses registration data to generate reports, statistics and trends analysis based on global aggregate data. The Data Port was developed to maximize the potential of the data collected and contribute to global data analysis on displacement. The Data Port can be used by UNHCR and Partner Users of PRIMES tools; and will eventually be

²² Suite of UNHCR interconnected (registration, identity management, case management, assistance management) tools that forms the basis of UNHCR's work on digital identity.

accessible to persons of concern as well as others looking for statistics on forced displacement. The tool is simple enough to be used by regular users and not exclusively technical users.

190. Registration data also provides the basis for identity management during assistance distribution. Individuals are verified against their registration record (biographic and biometric) at the point of distribution, and records of those receiving assistance updated after each distribution, ensuring accountability and reconciliation.

191. For assistance activities related to food security for persons of concern, the [2011 MoU between UNHCR and WFP](#), complemented by the [Cash Addendum \(May 2017\)](#) the [Joint Principles on Targeting \(December 2017\)](#) and the [Data Sharing Addendum \(September 2018\)](#)²³ establish areas of cooperation between UNHCR and WFP, namely in the areas of preparedness planning, joint assessment of the eligibility of refugees and returnees for food assistance, the composition of the food basket or cash equivalent and the modality of the assistance to be distributed.

192. The Joint Data Centre (JDC), a collaboration between the World Bank and UNHCR, supports the alignment of data and information between humanitarian and development actors, pursuant to the Global Compact on Refugees. It also supports the World Bank's IDA 18 programme, which dedicates World Bank grants to refugee populations and their host communities.

193. Registration data in proGres is 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). When complemented with additional socio-economic household surveys and assessments, this data can provide UNHCR and its partners, as well as the World Bank and its partners, with demographic and welfare information on persons of concern to support tailored programming and planning.

194. UNHCR and the World Bank also collaborate on the anonymization of registration data for open data access. Open data access via [UNHCR's Microdata Library](#) allows for high-quality analysis and research on refugee and other populations of concern, including for the purposes of identifying gaps in the implementation of services to populations of concern as well as informing broader national development and humanitarian planning.

195. The data available in UNHCR registration systems is set out in **Error! Reference source not found.** to be found in Part III of this Manual.

²³ The 2018 Addendum on Data Sharing sets out a global framework for the sharing of personal and non-personal data between the two agencies. It simplifies the procedure for requesting data on persons of concern and specifies the data elements that can be shared via this simplified procedure. The Addendum aims to facilitate timely provision of data for programme efficiency and protection, and to reduce duplication in data collection activities.

Data Source D: Integrated Data, Combining Data from Different Sources

2.3 Design Collection

196. Because of the variety and the varying coverage of the possible sources containing statistical information on refugee and internally displaced persons (IDPs), 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 ESS 2020 vision project in the area of data integration witness of the growing interest for data integration.

197. In particular, data integration can answer several challenges raised when compiling official statistics on refugee and IDPs by:

1. Improving the coverage of target groups,
2. Improving data availability, especially socio-economic variables related to refugee and IDP populations,
3. Improving accuracy through the crossing of different sources in order to assess refugee and IDP status.

198. The objective of this part of the Manual is to present the main steps related to data integration as well as an overview on the possible data integration techniques which can help and guide practitioners in NSOs interested in applying data integration to refugee and IDPs statistics. As far as possible, the provided information will be illustrated by existing good practices in NSOs and will be completed by more detailed technical information and references about the presented methods.

199. This section is an edited version of the full text on data integration. A much fuller description of options and techniques can be found in Part 3 of this report PART III – 2D Data Integration Techniques.

200. The first step is to clearly define the output objectives of integration and to agree this with the providers of the data sources to be used. This is followed by a discussion of the several issues that NSOs will face if they want to apply data integration techniques; these issues relate to the legal environment and data protection legislation, the relationships with external data providers, the required IT resources and the skills needed. The next section addresses issues pertaining to the main sources which could be the object of data integration and is followed by the main data integration statistical methods. These methods are illustrated by examples of good practices. Since any data integration process actually covers all the phases of the GSBPM, practitioners and data compilers can refer to the other parts of the Compilers' Manual for more detailed information on the statistical tasks common to all statistical processes, such as the quality assessment or the design of the process.

2.3.1 Output objectives of data integration

201. The output objectives of data integration should be defined at the very first stage. The feasibility and quality of the foreseen outputs need also to be assessed during this phase. When

establishing contacts with possible data providers establishing outputs has to be undertaken simultaneously, in order to have a clear vision and understanding of the data requirements between all stakeholders. Among possible output objectives which could be achieved through data integration, the following list shows the main objectives:

1. Production of new variables of interest (socio-economic variables related to refugees and IDPs)
2. Improve statistical classification (profiling)
3. Editing and imputation
4. Improve the coverage (detection of duplicates, use of capture/recapture methods)
5. Produce geographically very detailed data (integration with GIS)
6. Improve sampling frame for surveys (Administrative sources, New digital sources)
7. Quality assessment

202. The advantages and limitations of integrating data from different sources are set out below in **Error! Not a valid bookmark self-reference.** The starting point is the administrative register of individuals, to which can be added other data. These other data sources can improve quality, add variables and improve coverage and where new digital sources are used add geographic and movement details.

Table 2.17: Advantages and disadvantages of integrating data from different data sources

Sources	Administrative registers	General population surveys	Specific surveys (Refugees, IDPs)	Operational data	New digital sources (Satellite images, mobile phones, GIS)
Administrative registers	Determination of status Add variables Improve coverage Improve quality	Add variables	Add variables Improve coverage Improve quality	Improve coverage Determination of status	Geographic context, e.g. proximity to nearest school or hospital Movement patterns, e.g. daily commutes Detailed geographical breakdown Improve Coverage Improve timeliness
General population surveys			Add variables	Add variables	

Sources	Administrative registers	General population surveys	Specific surveys (Refugees, IDPs)	Operational data	New digital sources (Satellite images, mobile phones, GIS)
Specific surveys (Refugees, IDPs)			Improve coverage Improve quality	Improve sampling frame Improve coverage Improve quality	Improve sampling frame Detailed geographical breakdown
Operational data					Estimation
New digital sources (Satellite images, mobile phones, GIS)					Estimation Improve coverage Geographical information New indicators Sustainability and accessibility

2.3.2 Data availability, accessibility and other requirements

203. The design of the integration methodology constitutes one of the main tasks to be realised in data integration. Besides the technical tasks, good collaboration with external data providers and other involved stakeholders is of high importance when designing the data collection, production system and workflows. Very often, the sources that are considered for data integration are not under the direct responsibility of statistical authorities. The first requirement is to ensure the data's availability and accessibility, as well as to have in place the capacities to use them (see also Phase 1.5 Check Data Availability).

204. This process is obviously strongly dependent on the national context, in particular the legal environment as well as the status of the various statistical authorities. The following paragraphs will describe in more detail the main tasks to achieve taking into account the potential national context. Assessment of the points set out below should be undertaken at the very first stage and help to construct a business plan for data integration.

a. Legal environment and data protection legislation

205. 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.

206. A guarantee of data confidentiality, ethical standards, as well as the legal environment obliges NSOs to guarantee the confidentiality of data, either for dissemination or when sharing data with other providers. In particular, NSOs shall avoid direct identification or indirect identification of the respondent (statistical unit). Direct identification means identification of the respondent from their formal identifiers (name, address, identification number), whereas indirect identification means inferring a respondent's identity by a combination of variables or characteristics (e.g. age, gender, education etc). Concerning dissemination, statistical disclosure control can use methods like tabular data protection for aggregate information on respondents presented in tables (use of suppression, rounding, randomisation and/or interval publication) and micro-data protection (using local suppression, sampling, global recoding, top and bottom coding, rounding, rank swapping and micro-aggregation). Concerning the storage of data, physical protection guarantee that the data is securely stored and not accessible to anyone without explicit authorisation. When coming to data integration (micro-data sharing with external data providers), the use of anonymisation techniques (see also g. Security environment discussed below) shall guarantee the confidentiality of the data shared.

b. Cooperation with external data providers

207. Political willingness and technical readiness are essential. Considerations of data security tend to dominate, and a data sharing agreement is a requirement, this should identify the data points to be shared, and the modalities of data transfer. The following issues are recommended;

1. Building trust, depoliticizing the discourse.
2. Define cooperation with data providers and other involved stakeholders (administration, NGOs, other) through clear MOUs (“data used for this and only this”, safeguard of confidentiality by statistical authorities). What are the precise requirement and limits of use of the data, access to micro or macro data?
3. Undertaking a comprehensive mapping of the systems and of the data.
4. Ownership of the results by statistical authorities.

c. IT Resource requirements

208. Data integration has specific IT requirements which are as follows;

1. Necessary software capable of integrating data sources.
2. Compatible data format for receiving and sharing data, for extraction.
3. Secure connections to allow data sharing (VPN, API...).

d. Skills needed

209. In addition to IT data integration needs high level skills which go beyond the statistical;

1. Software architect and engineer to look at integration part,
2. Database developer to then look at structure,
3. Data management expert,
4. Legal advisor,
5. Statisticians/data scientists for building and implementing integration methods.

e. Statistical requirements

210. The availability of unique identifiers (UID) for statistical units or the possibility to build them is necessary. These may be obtained from

- 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
- Building of a synthetic UID based on data matching
- Probabilistic data matching
- Respect of the legislation on data protection

211. There should be a quality assessment of the chosen data integration methods including;

- Assessment of the data linkage quality.
- Assessment of the estimation quality when integrating various data sets.
- Assessment of the applicability and relevancy of the chosen method.
- Documentation on the methods used in order to guarantee transparency and clarity of produced statistics.

f. Collaboration with external data providers and other stakeholders

212. External data providers should be included in the business process in the following areas;

- Presentation to external stakeholders of the foreseen output (data to be collected, flows, stocks and other variables to be produced, analytical purposes),

- Mapping in collaboration with data providers between the information contained in various sources and the official statistics needs,
- Design of the data collection process in cooperation with external data providers (including metadata),
- Design production systems and workflows in cooperation with external data providers (software, connection between various stakeholders, IT resources, etc), and,
- Sustainability of the envisaged process.

g. Capacity requirements

213. Data integration requires considerable statistical capacity including;

- In house skills which vary depending on the adopted systems but in general which include data science and statistical modelling skills, necessary IT skills to exchange and store data in a secure and efficient way, as well as database administration skills including data mapping documentation, extraction, transformation and load (ETL), database performance tuning for load and retrieval of data.
- Capacity building
- Costs.

Table 2.18: Access requirement according to various sources

Sources/Access requirements	Legal environment	Availability	Sustainability	Statistical requirements	Required resources (Cost, IT environment, technical skills)
Censuses	Statistical law	Micro/macro data	Depends on statistical authorities	Low Frequency UID	Costly
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
General population surveys	Statistical law	Micro/macro data	Depends on statistical authorities	Frequency UID Size of the sample of the	Compatibility between IT stems

				population of interest	Data modelling
Specific surveys (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

g. Security environment

214. When exchanging or sharing data from multiple data providers, this process shall be in conformity with the data protection legislation. It means in particular that any data exchange is secure, and that micro-data confidentiality is respected. In addition, confidentiality of data disseminated must also be respected. Secure data exchange can be realised through data encryption, whereas hashing techniques can be used for ensuring micro-data confidentiality. Lastly, confidentiality of data for processing and analysis can be done with pseudonymisation techniques. Part III Section 2.3.2 of this Manual *e*. Security environment provides a short presentation of data encryption, pseudonymisation and anonymisation techniques.

2.3.3 Possible sources of data for integration

215. The possible sources for refugee and IDPs statistics are various and extensive. Before considering any kind of data integration, a review of the possible sources of information should be undertaken. This review should focus on checking their availability, defining the possible outputs that can be derived through combination/integration of several sources as well as assessing the quality of those outputs.

216. In the field of statistics on refugee and IDP statistics, an indicative list of main possible sources is presented below:

- Administrative registers on refugees and IDPs: asylum seekers registers (applications), registers on decisions granting international protection, resettled persons register, residence permits registers, population registers, IDP registers,
- Other administrative registers: social and health registers, tax registers, social security registers, electricity subscribers, etc ...
- Census
- General population surveys: general sample survey (labour force, education, income, living conditions, ...)
- Dedicated surveys on refugees and IDPs
- Operational data collected by UNHCR or NGOs, e.g. movement tracking systems (Displacement Tracking Matrix)
- Geographical Information system (GIS)
- New digital sources: Passively collected information from satellite images, mobile phones, web-scraping and web-crawling of internet platforms or social media content?

Table 2.19: Main characteristics of potential data sources for integration

Sources	Content	Statistical unit	UID*	Limitations
Administrative registers on refugees and IDPs	Information on the number of refugees and IDPs. Identification of the legal refugee/IDPs status	Individual, Household	Yes	Presence of duplicates Time lag in registering Update of the registers Quality considerations
Other administrative registers	Socio-economic information Stock estimation	Individual	Yes	Accessibility Identification of refugees/IDPs Update of the registers Quality
Census	Socio-economic information Stock estimation	Individual or household	Yes	Frequency Cost Identification of refugee or /IDP status
General population surveys	Socio-economic information	Individual, household	Yes or Composite UID	Sampling design for targeted sub-population Identification of refugees/IDPs status Cost
Surveys on refugees and IDPs	Socio-economic information Identification of the self-perceived refugee/IDPs status	Individual household	Yes or Composite UID	Sampling design Cost Coverage

Sources	Content	Statistical unit	UID*	Limitations
Operational data	Number of refugees/IDPs in camp	Individual	Yes or Composite UID	Coverage and quality concerns/ suitability for official statistics
GIS	Geographical breakdown	Individual household	Yes	Availability
New digital sources	Detection of population inflows Definition of areas to be surveyed	Individual household	Yes or Composite UID	Accessibility Need resources and robust methods Sustainability Cost?

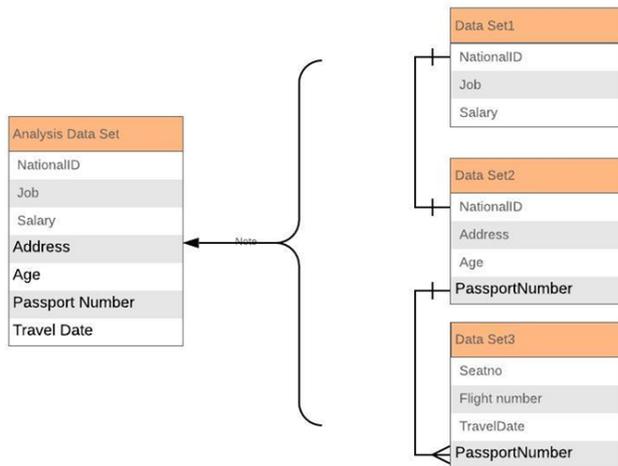
*UID = Unique identification number

217. As the list of possible sources presented in *Table 2.19: Main characteristics of potential data sources for integration* shows, a unique source does not exist that is capable of drawing a complete statistical picture of refugees and IDPs. Either in term of coverage or data availability, usually only integration of several sources can lead to the production of the necessary statistics for the users. Most often, integration of several sources, requires the linking of several data sources at the micro-level by using an identifier common to all the involved sources. Sometimes, data integration can also consist in using several sources at a more aggregated level in order to derive the best estimate of a target value, in particular the total number of a population. In both cases, micro and macro data integration, output objectives of the foreseen process must be well defined.

2.3.4 Overview on linking several data sources

218. One of the main challenges in any data integration project is how to establish the link between the different data sets, and the fact that those data sources usually have different structures. Depending on how the data need to be integrated, this can be generalized into 2 main categories;

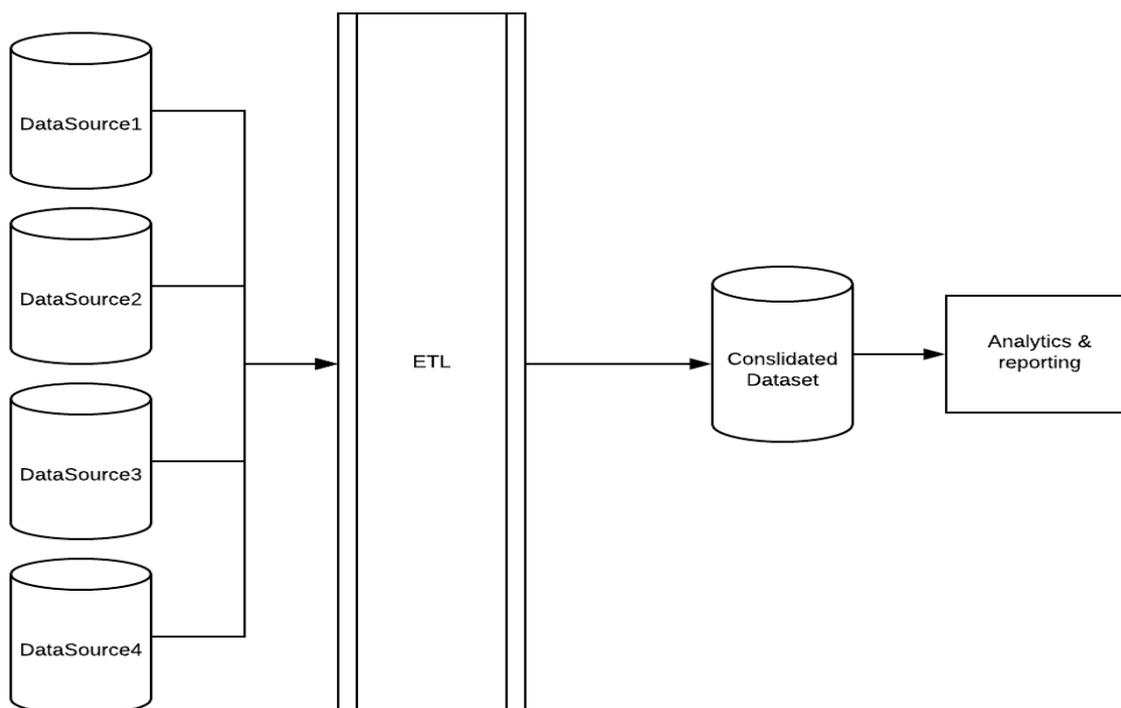
1. **Data linking (Join):** normally a common variable is required in multiple data sets to establish a link. This common variable or variables is called a Join Key. It could be one unique identifier used in 2 data sets such as a national ID number, or it could be set of quasi identifiers that exists in both data sets. The figure below illustrates an example of what needs to be done to expand the data set horizontally or by adding more variables to analysed data set.



2. **Data consolidation (Union, or data alignment):** when data sets come from multiple sources and the aim is to create one unified dataset with all records. As data sets might have different structure and variable scopes, there might be a need to transform the variables before adding them to the consolidated dataset. For example, in one data set there might be date of birth while the target dataset uses the year of birth. In that case one needs to transform date of birth before loading the data into the target dataset. As under the previous example, one can think of what needs to be done to expand the data set vertically or by adding more observations to analysed dataset.

Birthdate		Birthyear
31/12/1990	----->	90's
15/10/1978	----->	70's
19/04/2000	----->	00's
01/11/1997	----->	90's
15/03/2000	----->	00's
01/12/1983	----->	80's
05/05/1995	----->	90's
03/08/1948	----->	40's
05/09/1992	----->	90's
14/11/1992	----->	90's

219. In most projects both are needed i.e. relevant variables have to be selected from different data sets and transformed before injecting them to target database. This process is known as ETL (Extract, Transform, Load).



2.5 Design, processing and analysis

220. This section describes the different tasks required by data integration and provides a description and references on the methods that could be applied. This part is mainly based on the work done within the framework of the ESS Vision 2020 ADMIN (Administrative data sources), in particular the one on data integration and estimation methods²⁴. Nevertheless, this section shows as far as possible how integration and estimation methods could be applied to the field of refugee and IDP statistics.

2.5.1 Data editing and imputation

221. When using multisource data, inconsistencies at micro level are frequently encountered. Integration of data sources at micro-level may give rise to composite records that consist of a combination of values obtained from several different sources. For instance, a register could be combined with values obtained from a survey for the same units (obtained by record linkage); or from an integration of several surveys with non-overlapping units, in which case a unit from one source is matched with a similar (but not identical) unit from another source. Records with values obtained from different sources can also arise as a consequence of item non-response and subsequent imputation, in which case the two sources are the directly observed values versus the values generated by the imputation method. In all these cases the composition of a record by combining information obtained from different sources may lead to consistency problems because the information is conflicting, in the sense that edit rules that involve variables obtained from the different sources are violated. The purpose of editing conflicting micro-data is to achieve numerical consistency in the first instance and, ultimately,

²⁴ https://ec.europa.eu/eurostat/cros/content/wp2-statistical-methods_en

statistical consistency of the resulting aggregates.

222. For example, a survey provides an aggregated value A for the total income received from social allowances received by person x; this value A should be equal to the sum of two types of allowances B (health) and C (education) collected through administrative sources. If A is not equal to the sum of B and C, this inconsistency should be removed by using micro integration methods²⁵ like prorating, minimum adjustment methods, or generalised Ratio Adjustments. In the case of data on refugee and internally displaced persons, such aggregation rules when integrating several sources should be rather rare, but their existence should be checked on a case by case basis, and data compilers should take them into account if necessary. Micro integration or editing conflicting micro-data needs to be distinguished from variable harmonisation (later on described) which applies to situations where there exist similar versions of the same target variable, for instance due to different definitions across the sources.

2.5.2 Creation of joint statistical micro data

223. A full description of a) data linkage: combining data from several sources belonging to the same group, and b) data matching: inference of joint distribution from marginal observations, can be found in greater detail in Part III of this Manual Creation of joint statistical micro data.

224. The integration of data at micro level with the main purpose of accurately reflecting the individual's real-world circumstances at individual micro level, even when the data are stored in different sources of various types, is known as record linkage. Ultimately, the record linkage process creates a micro dataset where the corresponding separate observations are combined into joint statistical data drawn from information which originate in multiple sources. Record linkage is also referred to as object identification, record matching, entity matching, entity resolution, or reference reconciliation. In the case of the National Statistical Institutes (NSIs) the joint use of statistical and administrative sources is a product of a rationalization of all the available sources in order to reduce costs, response burden and, most of all, to enrich the information collected in order to produce high quality statistics.

225. An important distinction between record linkage and statistical matching needs to be made, the former regards the "fusion" of sources composed mainly of the same units, partially or completely overlapping, e.g., in the case of integration of administrative registers and sample surveys, while the latter concerns the integration of different units, e.g. derived from different sample surveys.

226. The record linkage problem requires statistical estimation methods when unique identifiers are not available for all the units in the sources, or the unique keys are affected by errors, or a unique composite key can be derived from a combination of error-prone variables. The presence of errors in the unit identifiers and the use of statistical estimation methods for linking data files, may introduce linkage errors in the integrated data. This point needs to be taken into consideration by data compilers' during the design of the data linkage process. Actually, a standard statistical estimation approach can produce inaccurate results in case of linkage errors, which requires the use of statistical methods in order to evaluate and adjust for

²⁵ Pannekoek, J. and Zhang, L.-C. (2015) Optimal adjustments for inconsistency in imputed data. *Survey Methodology*, vol. 41, pp. 127-144.

linkage errors.

227. In **statistical matching**, also referred to as data fusion, the ultimate aim is to provide estimates on variables observed in distinct data sources by exploiting the available common information (variables observed in both data sources). The data to be joined is characterized by representing the same population but not necessarily having common units. Thus, statistical matching can be framed as an imputation problem in which a missing target variable for units in a recipient data set is predicted based on information on similar units (“statistical twins”) observed in a donor data set.

228. A statistical matching problem can be defined according to its objective and data availability:

1. (objective) knowledge on a (possibly conditional) joint distribution function of a couple of random variables, or on some of the parameters representing the (conditional) joint relationship between the two random variables, or on a data set representative of the (conditional) joint distribution of the two variables, is requested;
2. (available data) the two random variables of interest are not available jointly in any datasets, but it is possible to use two random samples, each one observing one of the two random variables, both observing all the conditioning variables. Furthermore, the two samples should be representative of the same population, but the overlap between the two observed sets of units is either null or consists of subsets of units that are not representative of the population itself.

229. The two most important outputs that statistical matching can produce are: either macro (estimation of a joint distribution or of a relationship parameter) or micro (a data set with complete observations on the variables of interest, i.e. both the couple of random variables and the conditioning variables). It is worthwhile noting that the macro approach to statistical matching may be also be considered as a ‘Multisource estimation at aggregated level’. In the field of refugee and IDP statistics, the micro-objective (fusion of two datasets) is the most frequent possible use of data matching.

230. A full description of deterministic and probabilistic linking, with examples, can be found in Part III) Creation of joint statistical micro data.

2.5.3 Alignment of populations and measurements

a. Alignment of populations: Unit harmonisation

231. When data originating from multiple sources are used, one needs to transform objects into statistical units referring to the target population. This is a mandatory requirement valid both for usages based on a single data source, and for cases based on multisource data. After a preliminary analysis aimed at verifying such a consistency, it is sometimes necessary to start a process of harmonisation or alignment of statistical units. For example, when combining survey data with administrative registers, to create a living household as the ideal unit for refugee or IDP statistics, one needs to make use of units such as person, family, residence address, study or workplace address, etc. The different units need first to be aligned with each other, in order to improve the accuracy of the living household created based on them.

232. Possible estimation methods: No general statistical methods are available, while ad-hoc deterministic methods are generally used. For instance, statistical units are derived using certain deterministic derivation rules based on available information.

b. Alignment of measurements: Variable harmonisation

233. When administrative data are used, one needs to transform attributes of the objects into measurements referring to the concepts to be measured, that is the target variables. After a preliminary analysis aimed at verifying such a consistency, it is sometimes necessary to start a process of harmonisation or alignment of measurements.

234. When data from multiple sources are combined, differences in definition can occur between variables in different sources. In particular, variables in an administrative data source are defined according to the administrative purposes of the register owner. These definitions may differ from those of the target variables for statistical purposes. For example, when the variable 'age-group' is observed in two data sources with different groups. A technique for obtaining a unique variable 'age' is needed. This apparently deterministic task can involve statistical estimation method (e.g., classification techniques) whenever a well-defined mapping is not known.

235. In case differences in variable definitions occur between data sources, these variables need to be harmonised during data integration. That is to say, for each unit in the integrated data set, the values of the target variable according to the desired definition need to be estimated from the observed values that are available.

2.5.4 Multisource estimation at aggregated level

236. This task refers to the phase of production of estimates by using integrated multisource administrative data. Estimation methods in this case must deal with the problem that data are not obtained according to a random sampling design, and that observations may refer to specific subsets of units of the population (under-coverage), or out-of-scope units (over-coverage). In addition, when integrated multisource data are used, specific problems may arise. They are concerned with the problem of consistency and coherence of estimates.

237. Part III of this Manual 2.3.5 Multisource estimation at aggregated level describes a number of estimation and reconciliation techniques in much further detail including;

a. Population size estimation: multiple lists with imperfect coverage of target population.

238. When each one of the multiple sources has imperfect coverage of the target population, including both under- and over-coverage, statistical methods for population size estimation are needed. The most common approaches relate to capture-recapture (CRC) methods. In estimating a population size based on several sources, the misalignment between the scope of the administrative data and that of the statistician poses several methodological challenges and sets us apart from a classical capture-recapture setting. For instance, it is often useful to develop methods taking into account the dependence among data sources, the fact that some data sources refer to a specific subpopulation, and that data may contain units out-of-the target population that are not deterministically identifiable.

239. 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. Difficult to reach refugees or IDPs through surveys will most likely not be recorded in administrative registers, which can break the third assumption and imply a problem of under-coverage. 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, applying the CRC method to each identified specific group before proceeding to an aggregation of the results found for each group. Lastly, the fifth hypothesis could be assumed to be verified in most of the cases.

240. The violation of these assumptions can lead to serious bias in the CRC-estimation of the population size. CRC is in particular sensitive to violation of the assumption in the case of a low implied coverage, i.e. the second register overlaps greatly with the first register and adds relatively few new records to it. So, several extensions of the CRC method were proposed in order to face problems connected to violation of the basic assumptions, we can divide them into two groups: methods aiming at improving the CRC-method; alternatives methods.

b. Reconciliation

241. Different estimates for the same phenomenon could lead to confusion among users of these figures. Many NSIs have therefore adopted a one-figure policy. According to this one-figure policy, estimates for the same phenomenon in different tables should be reconciled, meaning that they should be equal to each other, even if these estimates are based on different underlying data sources.

242. When using a mix of administrative data sources and surveys on which to base estimates upon, obtaining one estimate for the same phenomenon may become problematic as for different (combinations of) variables data on different units, e.g. different persons, may be available.

243. This means that different estimates concerning the same variable may yield different results, if one does not take special precautions. For instance, if one uses a standard weighting approach to produce estimates, where one multiplies observed counts or values with surveys weights, one may get different estimates based on two samples, because different units and hence different survey weights are used in the two samples.

244. In principle, these differences are merely caused by “noise” in the data, such as sampling errors. So, in a strictly statistical sense, different estimates concerning the same variables are to be expected and are not a problem. However, different estimates would violate the one-figure policy and form a problem for the users.

Examples of good practices

- 1) WORK PACKAGE 1 OF THE ESS VISION 2020 ADMIN: ACCESS TO AND DEVELOPMENT OF ADMINISTRATIVE DATA SOURCES

245. The European Statistical System [ESS Vision 2020](#) highlights the importance of cooperation with stakeholders and of developing strategic alliances with private and public

partners (key area "identifying user needs and cooperation with stakeholders"). In particular, the ESS Vision 2020 ADMIN (Administrative data sources) project carried out several tasks related to the access to administrative sources. The goal of this project is to help ESS Member States to make wider and better use of administrative sources in the production of official statistics. This means addressing the most typical challenges faced in the use of these sources: limited access to the data, the lack of quality of the sources, methodological issues related to the processing of the data and the integration of several sources. It is also important to ensure that the European statistics produced using administrative data are comparable across Member States and are of sufficient quality. The first work package²⁶ of ESS Vision 2020 ADMIN answers this by aiming to identify and share best practices on cooperating with the owners of administrative sources. (see Phase 1.5 Check Data Availability for reports and more information)²⁷.

2) THE POPULATION REGISTER AND LINKS TO OTHER ADMINISTRATIVE REGISTERS IN NORWAY

246. The Central Population Register (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. Persons who die or emigrate are not deleted from the register, but a code for their status is changed.

247. The most important stock (or status) variables in the CPR are: 11-digit PIN (which includes date of birth and sex), residence status (resident, deceased, emigrated, no permanent address, disappeared), address, municipality, dwelling number, place of birth (municipality or country), name (incl. first and middle names), citizenship, country of immigration, country of emigration, marital status, PIN of spouse, mother and father.

248. Persons who have come to Norway as refugees or asylum seekers are included in the CPR once they are recognized as legal residents. Refugees who arrive via UNHCR are given a PIN on arrival. 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

249. All vital events and migrations and address changes are registered in the CPR. The most important flow variables are births, deaths, marital changes (incl. same-sex marriages, separations, divorces and annulments), emigrations and immigrations, internal moves in Norway, address changes, name changes, citizenship changes, gender changes and PIN changes. When a report of a change is received by the CPR the information is checked, for

²⁶ https://ec.europa.eu/eurostat/cros/content/wp1-access-administrative-data_en

²⁷ A workshop was organised in 2016 in order to allow the National Statistical Institutes to exchange best practices concerning access to administrative data sources ([Workshop on Access to Administrative Data Sources](#), Brussels, 13-14 September 2016). Between 2015 and 2017, information was collected on the obstacles that the NSIs still face in accessing administrative data and how they have managed to overcome some of the difficulties. On this basis, two reports were finalised in 2018:

Task 1.1 Report on the legal and institutional environment in Member States: [Final report](#)

Task 1.2 Best practices regarding the relation with data providers: [Final report](#)

example, that the spouses of a new marriage are not already married and that the parents of a new-born baby are residents.

250. 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 not any longer necessary to conduct traditional population and housing censuses. The last regular census was conducted in 2001, whereas the 2011 census was completely based on register data.

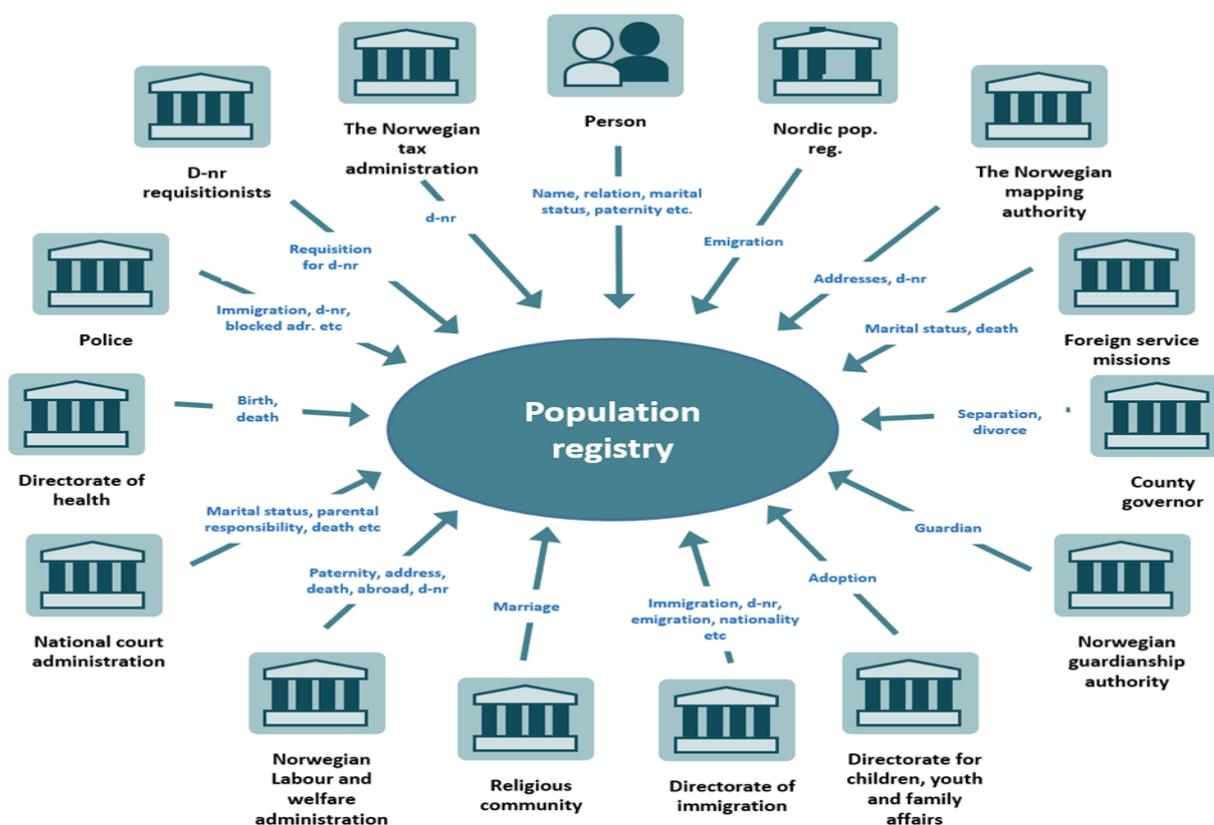
251. The two figures below show the flows of data on individuals (microdata) to and from the CPR. The entity at the middle of both figures, "Population register", functions both as the central register of civil events in Norway and as the country's population register. Thus, civil registration and national identity management are fully integrated. Births to residents of Norway and new immigrants are assigned personal identification numbers. Residents who die or emigrate are not physically removed from the register, but their codes are changed from "Resident" to "Dead" or "Emigrated".

252. The first figure shows that the population register receives microdata *from* many public (and some private) institutions on births, deaths, many external migrations, marriages and divorces, adoptions, address changes, name changes, and other events. It also receives information directly from the individuals themselves on internal migration, many external migrations, and name changes.

253. 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. The second figure shows the public and private institutions *to* which the Norwegian Tax Administration distributes microdata. Statistics Norway is one of the most important recipients and users of these data, together with health and educational institutions, the police, the army, the electoral register, banks, and other bodies.

254. The task of sharing the microdata is not performed by Statistics Norway, but by the owner of the administrative register, the Norwegian Tax Administration. Statistics Norway may, however, share anonymized data on individuals if this is approved by the data inspector.

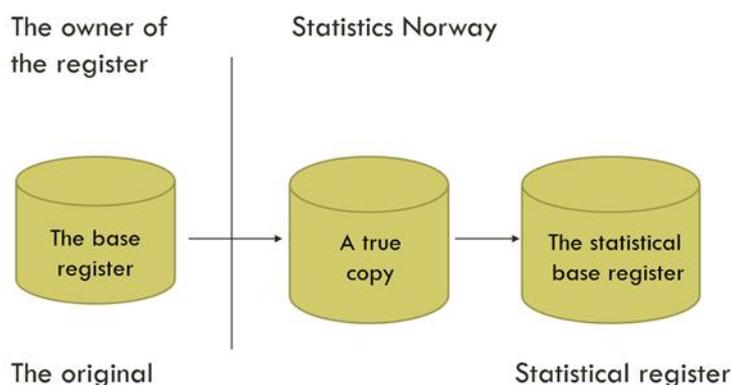
Figure 2.1: Data flows to the Norway Population Register



255. Statistics Norway receives updates on vital events and other changes in the register every night, five days a week, following the same routine as many other users of the population register that need daily updates for administrative purposes. The updates are verified every morning. The diagram below shows the relationship between the administrative and statistical population registers. Statistics Norway maintains two versions of the population register:

- A true copy of the administrative register;
- A separate statistical population register, in which internal adjustments and amendments are made and saved.

Figure 2.2: Copies of registers in Norway



256. The key reason for this approach is that the administrative register is incomplete and may include errors and other shortcomings, such as missing data. 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. This information cannot be shared with the Tax Administration without permission from the other sources for confidentiality reasons, but it still has an impact on the quality of the statistics produced. Statistics Norway may, however, report faulty microdata (including the PIN) back to the register owner, including the Tax Administration, as this may be considered a complaint regarding the quality of the data received. These error correction procedures must be in line with the principle that data relating to individuals should not be returned from a statistical agency to a supplier. The quality improvements mean that, in most cases, the statistical population register is a better source for statistics production than the administrative population register (CPR) would have been.

257. It is not possible to identify refugees/(former) asylum seekers directly in the CPR as information about reason for migration/residence permit is not one of the variables that is included in the original transactions from the CPR. However, this information is collected on a yearly basis where Statistics Norway receives data from the Directorate of Migration (this procedure is described more in detail in box 4.6 in the IRRS) and link these data using the PIN to match the statistical version with the CPR. Based on this information, Statistics Norway produces both flow and stock data on persons with a refugee background, see: <https://www.ssb.no/en/befolkning/statistikker/flyktninger/aar>.

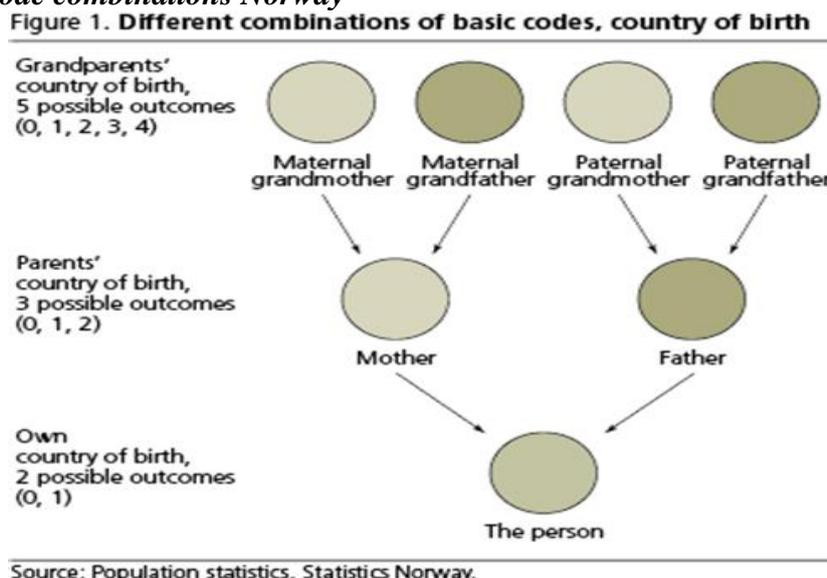
3) MAINTENANCE OF A REGISTER FOR STATISTICAL PURPOSES – CONSTRUCTION OF THE VARIABLE MIGRATION CATEGORY IN NORWAY:

258. In Norway, population, migration and refugee statistics – both on stocks and on flows - is primarily based on register data from the Central Population Register (CPR), a register situated in the Norwegian Tax Administration. Statistics Norway (SSB) maintains a copy of the CPR – kept for statistical purposes - and receives electronic transaction data from the CPR every night.

259. The same definitions are used in both the CPR and in the SSB copy of the register and routines for updating correspond closely. The result is a good correspondence between the two registers, even after many years with physically independent updating routines. However, in addition to the variables received from the CPR, SSB constructs several variables that are needed for statistical purposes.

260. One of these is the variable 'migration category' which is used for determining the population with immigrant background. See Figure 2.3:.

Figure 2.3: Code combinations Norway



261. The variable is based on the person’s own country of birth, parent’s country of birth and grandparent’s country of birth. All these are derived from the CPR. From this variable SSB constructs the following categories used for its migration statistics.

- Immigrants - persons born abroad of two foreign-born parents and four foreign-born grandparents.
- Norwegian-born to Immigrant Parents – persons born in Norway of two parents born abroad and in addition have four grandparents born abroad. This category was earlier labelled “Second generation migrants”.

262. In the statistics on ‘persons with a refugee background’ (see other reference in this document) only those who themselves have immigrated are included. However, through using the variable ‘migration category’ it is possible to follow the children of refugees and study their integration into Norwegian society.

4) FEDERAL STATISTICAL OFFICE OF GERMANY ON DATA INTEGRATION IN THE ABSENCE OF UNIQUE IDENTIFIERS - USING MACHINE LEARNING TO IMPUTE REFUGEE STATUS IN SURVEY DATA

263. 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.

264. 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.

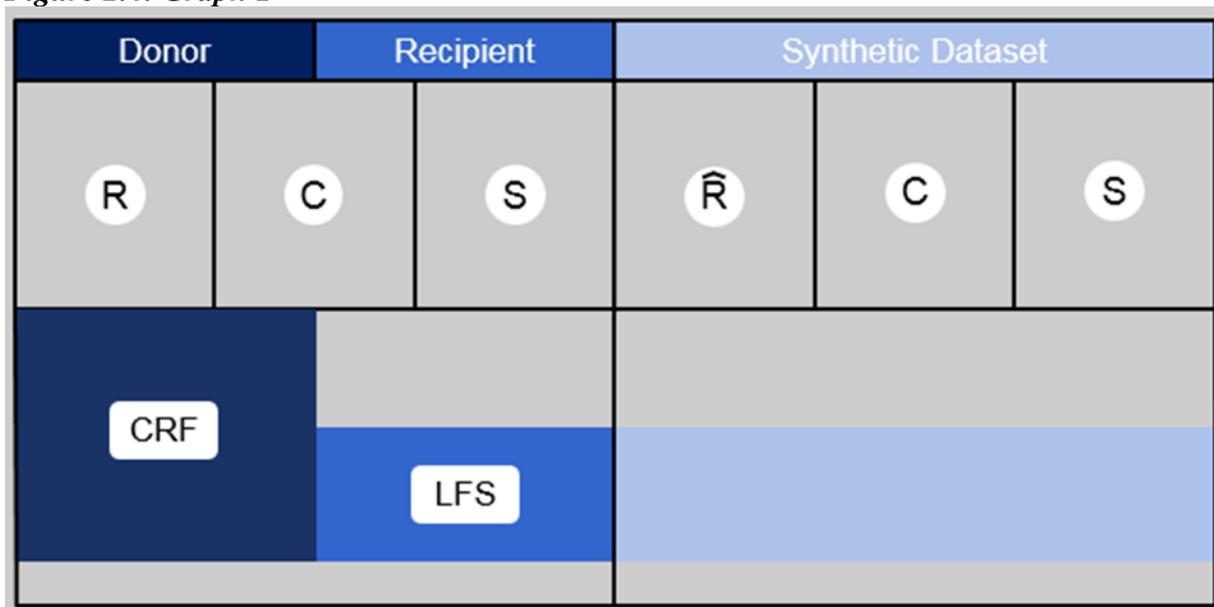
265. The data environment in which statistics on refugees and people in refugee-like situations are produced in Germany is typical for many countries: Germany uses administrative data on residence permits from the Central Register of Foreigners (CRF) to identify refugees and foreigners in refugee like situations. While the CRF confidently identifies and comprehensively covers refugees, it does not provide 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.

266. 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 $\hat{R}_i = f(C_i)$
- (4) Synthetic data set with C_i, S_i, \hat{R}_i

Graph 1:

Figure 2.4: Graph 1



267. This case-study aims at 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, the case-study emphasizes the importance of carefully evaluating model performance and guides the reader through a framework of possible evaluation techniques.

268. In summary, this case-study adds to the existing literature in two ways: First, statistical matching techniques have not yet been applied to the field of refugee statistics – a field in which they might be especially beneficial since they offer a quick and cheap alternative compared to traditional forms of data collection for refugee hosting countries. Secondly, the field of data science has developed rapidly over the last decade. Powerful classification algorithms have emerged that have not yet been exhaustively tested in the statistical matching environment. As a convenient feature, those algorithms are implemented in open source software²⁸ making this study reproducible for a broad audience.

²⁸ Insert link

2.6. Design production systems and workflow

“This sub-process determines the workflow from data collection to dissemination, taking an overview of all the processes required within the whole statistical production process, and ensuring that they fit together efficiently with no gaps or redundancies. Various systems and databases are needed throughout the process. A general principle is to reuse processes and technology across many statistical business processes, so existing production solutions (e.g. services, systems and databases) should be examined first, to determine whether they are fit for purpose for this specific process, then, if any gaps are identified, new solutions should be designed. This sub-process also considers how staff will interact with systems, and who will be responsible for what and when”. (UNECE, 2013)

2.6.1. Introduction

269. This section deals with fitting together the various components of the design into a coherent whole. It also includes the design of the quality assurance processes through which statistics should be evaluated for quality.

270. The gathering and harmonisation of data on displaced people are the key to international comparative analysis and policy work on forced displacement. The timely, accurate statistical information on IDP, asylum and refugee statistics enables a wide range of policy answer in the rapidly evolving forced displacement caused by humanitarian crises.

271. The objectives of this IDP, asylum and refugee-related statistical production system is to improve efficiency of data and metadata collection, validation, processing, storage and dissemination. It is also a target of a well-functioning statistical production system to improve quality by eliminating errors and incoherencies, and by shortening the statistical publication cycles. The accessibility and visibility of the asylum related statistical outputs should also be a key objective.

272. In order to achieve these targets a modern statistical production system should utilize of advances of modern information technologies and standards.

2.6.2. The statistical production environment

273. Before developing a statistical production system, the data sources of this system should be determined. Based on their supervising body in the national data production system, there are two types of data sources which can feed these systems.

274. The first type is part of the national statistical system, this is the administrative data source where the data owners are an integrated part of the national public administration. In this case the country's government with its public administration is present and a national legislation is available which regulates the asylum procedure or IDP registration. Parallel to the national legislation are specialized administrative bodies or custodian agencies responsible for the IDP or asylum procedure and for the collection of the administrative data related to the procedures.

275. In this case the statistical data are by-products of the administrative procedure. In an optimal situation the IDP or asylum procedures are supported by IT-systems, and case-management systems where the data are stored and can be queried. This case-management

system is interlinked with a statistical sub-system, it means that the case-management system should have an interface where the statistical queries can be set. In this production system chain the case-management system is the input for the statistical subsystem, it is not part of the statistical data production system.

276. The statistical sub-system should ensure the ideal technical environment for data production and for the metadata production environment as well. The statistical production system should have also a link to the data storage warehouse, which is a dedicated part of the statistical production system. In the storage system the data are stored, if possible, in so called cubes, these cubes should have a link to the dissemination system which enables the statistical data publication via user interfaces. According to the methodology of the data inquiry we can differentiate between static systems where the data can be requested via static tables e.g. Excel tables or html tables. The other way is to enable the compilation of user-friendly, interactive statistical tables.

277. The second type of data source describes an IDP or asylum system where the national legislation is not so well developed as in the first case, and a dedicated governmental custodian body for the asylum procedure is not functional. Refugee status-determinations can be undertaken by UNHCR in contrast to status determination carried out by a government. In the latter case the responsible agency for the asylum procedure is an international UN organisation such as UNWRA or UNHCR. The data source is the administrative system of this international organization which collects data related to the IDP or asylum procedure and also disseminates statistics at the international level.

278. If such an asylum related statistical system is available, the international organization serves as the producer of the input for the national public administration. A memorandum of agreement could enhance the cooperation between organizations of public administration and the UNHCR which is directly involved in providing shelter for asylum seekers and refugees and in data collection (see Part 2 *UNHCR Agreements and Memoranda of Understanding* for an example of this).

*'UNHCR collects data on persons in concern i.e. refugees, asylum-seekers, internally displaced persons (IDPs), returnees and stateless persons. The data collection includes information on country of origin and of asylum, sex age and location or settlement such as camp or centre. Furthermore, information about protection and living conditions is also gathered with variables such as education, health, nutrition as well as information regarding sexual and gender-based violence.'*²⁹

279. The main obstacle to data collection is the lack of harmonised definitions and measurement criteria which can occur in cases when the country did not comply with the 1951 Geneva Convention relating to the Status of Refugees (See IRRS Chapter 2), or in the case of IDPs observe the *Guiding Principles on Internal Displacement*³⁰ The lack of harmonisation can be especially problematic if the translation of definitions of the convention into national law is not done in a systematic and harmonised way, and not in observance of the

²⁹SIGMA0108, The UN Refugee Agency – global provider of asylum data, <https://www.unhcr.org/en-my/4868e61e4.pdf>, downloaded: 29th May 2019

³⁰ E/CN.4/1998/53/Add.2, 11 February 1998, UN Commission on Human Rights, Report of the Representative of the Secretary-General, Mr. Francis M. Deng, submitted pursuant to Commission resolution 1997/39. Addendum: Guiding Principles on Internal Displacement, available at: <https://goo.gl/GJU2Sz>

Recommendations set out in the IRRS or IRIS. Other factors which can make the data collection difficult are differences in quality and periodicity.

2.6.3. *The statistical database environment*

280. When establishing a statistical production system there are also some criteria related to the technical details and functionalities of the system which should be analysed more detailed. It is recommended to develop a common hosting environment for production databases, the idea behind the common production system is that it strengthens standardisation of the statistical data production and it minimises the number of tools used in statistical activities and reduces training and support requirements.

281. It is recommended that the statistical data production system should include a toolkit which enables the management of statistical data, this toolkit should include different facilities, such as:

- data migration³¹,
- security management,
- database administration (management of user access rights)
- data collection,
- data validation,
- data-query toolkit,
- metadata management system.

282. When planning a statistical data production system, it is essential to design a centralized metadata storage system which enhances the efficiency of metadata preparation, storage, access, management and dissemination. The centralized metadata system addresses the problems of fragmented metadata located in numerous databases. The metadata sub-system of the statistical production system should meet some basic expectations, such as:

- the metadata system should have a common structure, defining common metadata items,
- all statistical data should have appropriate metadata,
- it should cover a broad group of user needs,
- the metadata must be consistent across the different subject matter areas.
- The metadata system should accommodate any kind of metadata related to the corresponding statistical data.

³¹ Source: OECD Statistical Information System (SIS) <https://www.oecd.org/sdd/oecdstatisticalinformationsystems.htm>, downloaded: 29th May 2019

Eurostat's Single Integrated Metadata Structure (SIMS)³²

The Single Integrated Metadata Structure is the dynamic inventory of statistical concepts used for quality and metadata reporting in the European Statistical System.

The Single Integrated Metadata Structure streamlines and harmonises metadata and quality reporting in the European Statistical System (ESS) and it decreases the reporting burden on the statistical authorities by creating the framework for "once for all purposes". It is an integrated quality and metadata reporting framework, which ensures a flexible and up to date system where future extensions are possible.

The system enables that the same concept names and the same quality indicators are used in the different ESS metadata structures, it also enhances the consistency with the SDMX statistical standards as it stands in the SDMX Content-oriented Guidelines.

It has to be underlined that data users and data producers have different needs on statistical information and this has to be reflected by the quality reports that are addressed to them. The distinction between the user and producer oriented quality reporting is assured through the Single Integrated Metadata Structure.

Based on this approach there are two different quality reporting systems available within the SIMS, the user-(U) and producer (P) –oriented quality reporting. The short user-oriented quality report is implemented through the improved visibility and readability of the quality related concepts and the producer-oriented quality report (P) is implemented via the ESQRS³³ report structure.

2.6.4. Statistical dissemination process

283. It is essential that the design of the statistical production system (the statistical production database) should interlinked with the dissemination system, this system enables the establishment of a modernised tool for traditional statistical publications and interactive data products.

284. For the dissemination system the data source is the statistical production system, the production database, it increases the efficiency of statistical dissemination and reduces the risk of mistakes, enables more timely publication of statistics.

285. The improved statistical dissemination system ensures a common layout and reduces the number of different software applications in use. The recommended statistical publication tool should have an interface for managing publication contents definition and it is also necessary to have an interface for generating XML output file in order to combine statistical data and metadata.

286. A workflow model should be developed which monitors the complex statistical data production from the production of statistical data to the production of metadata, through the storage of the statistical data. This complex model also monitors the statistical channels used for monitors.

2.6.6 Indicator system for measuring the design production systems and workflow³⁴

³² Source: European Commission, Technical Manual of the Single Integrated Metadata Structure (SIMS): <https://ec.europa.eu/eurostat/documents/64157/4373903/03-Single-Integrated-Metadata-Structure-and-its-Technical-Manual.pdf>, downloaded: 03rd June 2019

³³ ESS Standard for Quality Reports Structure

³⁴ Source: Modernstats; Quality Indicators for the Generic Statistical Business Process Model (GSBPM) - For

287. In order to measure the effectiveness of the statistical production system and workflow an indicator system should be established which can measure the effectiveness of the statistical production with indicators like, soundness of the implementation, the cost effectiveness, accuracy and reliability, timeliness and punctuality and last but not least accessibility and clarity.

288. A suitable indicator system can be seen in the table below, in the following this indicator system is presented: the indicator soundness of implementation measures the quality of the workflow in the design and production system phase of the GSBPM structure. It identifies the different sub-processes being part of this phase. The soundness of implementation should take into consideration the types of the data which are processed (admin data or survey type data) in this phase, because it affects the specifications of the indicator system.

289. It should be also outlined that the cost of designing a statistical production system plays an important role, it is necessary to reduce the costs of the different software tools. In order to achieve this goal, the best solution is the use of common applications. The products should be disseminated (see also Phase 7 *Dissemination*) on an appropriate platform, e.g. on the home page of the body responsible for statistical data production which can also increase the costs of the statistical production but using optimal software and technical solutions this cost can be reduced.

290. The indicator 'accuracy and reliability' measures the ratio of quality indicators used as performance indicators, the source data, integrated data, intermediate results and statistical outputs should be regularly assessed and validated.

291. The next indicator, 'timeliness and punctuality' defines the time frame which is available for the production in the different sub-processes.

292. According to the last indicator in the list, 'accessibility and clarity', the number of visitors of the home page or other social media products of the statistical data producer is calculated. During the performance of this statistical data production system a quality indicator system can ensure a permanent feedback on the production system.

Table 2.20: Quality indicators for measuring the design of production systems

Quality dimension	Indicator
Soundness of implementation	Percentage of identifies and documented GSBPM processes
	Specifications for production systems and workflow take into consideration the type of data being processed (survey data or data from administrative data sources)
Cost effectiveness	Percentage to which corporate solutions (e.g. tools, processes, technologies) are reused in subsequent phases
	Percentage to which responsibilities for subsequent phases and sub-processes have been set
	Estimated cost for producing and disseminate designed outputs
Accuracy and reliability	Percentage to which quality indicators are planned to be calculated for subsequent sub-processes of GSBPM
	Percentage of quality indicators used as performance indicators
Timeliness and punctuality	Planned time frame for subsequent phases and sub-processes
Accessibility and clarity	The number of visitors of the home page and on the social media
	Number of metadata consultations
	The number of consultations of data tables

Phase References and Links

- Online Template for metadata documentation (European SIMS system); <https://ec.europa.eu/eurostat/data/metadata/metadata-structure>
- Links to automated quality check for integrated data “Holoclean” software; <http://www.holoclean.io/>. See also Phase 5.1 of this manual.
- Metadata example: <https://www.ssb.no/en/befolkning/statistikker/flyktninger>
- Guidance for Data Integration for Measuring Migration; <http://www.unece.org/fileadmin/DAM/stats/publications/2018/ECECESSTAT20186.pdf>

Phase 3. BUILD A SYSTEM FOR COLLECTION OF DATA AND PRODUCTION OF STATISTICS

“This phase builds and tests the production solution to the point where it is ready for use in the "live" environment. The outputs of the "Design" phase direct the selection of reusable processes, instruments, information, and services that are assembled and configured in this phase to create the complete operational environment to run the process. New services are built by exception, created in response to gaps in the existing catalogue of services sourced from within the organisation and externally. These new services are constructed to be broadly reusable within the statistical production architecture” (UNECE, 2013)

293. Each type of data source has a defined scope, and the build phase must present recommendations to achieve their objectives as efficiently and comprehensively as possible. This phase deals with the steps needed to build or update the statistical systems that capture, in the most efficient way, the data required for producing accurate statistics on Refugee and IDP populations. This phase can be structured in 7 steps:

- 3.1 Build collection instruments
- 3.2 Enhance or build process components
- 3.3 Build or enhance dissemination components
- 3.4 Configure workflows
- 3.5 Test production systems
- 3.6 Test statistical business process
- 3.7 Finalise production systems

294. This part of the Compilers' Manual will only cover those parts of the Build Phase which apply specifically to refugee or IDP statistics. Descriptions of generic activities can be found elsewhere.

3.1 Build collection instrument

“The collection instrument is generated or built based on the design specifications created during the "Design" phase. Collection instruments may be from surveys, or from data extraction routines used to gather data from existing statistical or administrative data sets. This sub-process also includes preparing and testing the contents and functioning of that instrument (e.g. testing the questions in a questionnaire). It is recommended to consider the direct connection of collection instruments to the statistical metadata system, so that metadata can be more easily captured in the collection phase. Connection of metadata and data at the point of capture can save work in later phases. Capturing the metrics of data collection (paradata) is also an important consideration in this sub-process.” (UNECE,2013)

295. The internationally recommended variables, tabulations and analyses of Refugee and IDP Statistics are described in the previous Phase 2. The statistical requirements of any country will have been discussed with users or potential users in order to modify the international standards to meet national requirements and circumstances (see Phase 1).

3.1.1 Development of questionnaires

296. The development of questionnaires for surveys and censuses should follow an easy to follow format, which optimises the sequence of the questions and the selection of collection tools which meet national needs and circumstances. The design and development of the questionnaire should follow the best practice experience and international recommendations wherever possible

297. Paper questionnaire design must be driven by a planning process based upon dialogue between the NSO and the users and potential users of the results. The information to be collected must respond to user needs both at national and international levels and therefore user consultation, or in the case of new statistics potential users, is crucial in this regard. The recommendations of the IRIS and IRRS should be followed.

3.1.3 Development of the collection methodology

298. In undertaking a survey, the statistician in the build phase is faced with a decision about the use of paper or electronic format. The advantages and disadvantages of these are briefly discussed, but as the decision is not relevant specifically to displaced populations, except in so far as technological equipment may face particular challenges in areas affected by disaster or conflict.

Questionnaire with face-to-face interview (PAPI)

299. The face-to-face interview methods used paper questionnaire in most countries historically. However, as a result of advances in technology, new ways of enumerating populations have been introduced. The traditional method of enumerating the population with face-to-face interviews can now be applied with the use of handheld electronic devices or the telephone to automatically capture data during the interview. These approaches can either substitute or complement the traditional face-to-face, pen-and paper interviews (PAPI) and can also be used in a variety of combinations (multi-mode method).

300. Advantages of the PAPI method: This is the most common method, especially in developing countries and in population groups with significant illiteracy rates. Data obtained from this method are potentially more complete and accurate because of the interaction between the enumerator and the respondent, and the opportunity to clarify some of the questions and probe for more adequate answers;

301. Disadvantages of the PAPI method: The data collection phase of the survey may take longer (or require a larger interviewing force) compared to other modes of collection, because in-person interviewers typically operate alone with much less supervision and control, whereas CAP has some built controls. There may be greater interviewer variance in in-person interviews and there are also greater opportunities for interviewer falsification of some questionnaire items or for entire interviews.

Computer-assisted personal interviewing (CAPI)

302. In computer-assisted personal interviewing (CAPI), the enumerator obtains information from one or more household members through an interview using an electronic questionnaire (CAPI software application) loaded on a tablet or other handheld electronic device (e.g. PDAs and laptops). The enumerator asks the respondent questions in a face-to-face interview and the responses are recorded by the enumerator directly into the electronic device. Following the interview, the data are sent to a central computing network electronically via internet or other

means of data transfer.

303. Advantages of the CAPI method:

1. Cost savings in data processing;
2. Improved quality of data mainly due to the possibility of having electronic checks, managed skip patterns are managed and the possibility of tailoring questions.
3. Shorter timelines for completion of data processing;
4. Improve monitoring of fieldwork and enables more efficient management of interviewers, i.e. use of CAPI allows for a detailed analysis of the time duration for completing the questionnaire, as well as updating enumerators' assignments and checking of the completed questionnaires by the supervisors.

304. Disadvantages of the CAPI method: When considering the CAPI method, the cost of providing all interviewers with the electronic device used to administer the questionnaire must be incorporated into the project budget.

1. Each interviewer must have her or his own tablet computer for data collection, which can represent a substantial initial investment.
2. Cost of access to server hardware, software, server maintenance and technical support.
3. Development of parameters for monitoring and controlling data collection (paradata)

305. While converting a paper questionnaire into an electronic format, the following aspects should be kept in mind:

1. The wording and structure of some questions may need to be changed to make it easier for the enumerators to work quickly and accurately on the device;
2. Features which can be added: data validation, edits, and preloaded answers;
3. Questionnaire designers should consider how to take advantage of the added features of an electronic questionnaire, particularly the ability to: preload existing data; implement consistency checks, range checks, and edits; take GPS coordinates; and offer on-screen help features.

3.1.3 Development of training plan, evaluation and selection of human resources

306. This part includes the development of materials for data collection (reference material for enumerators, logistics, logistics management), preparation of the technical document for the recruitment of external staff (including quantities), development of the human resource allocation plan (based on workload, distances, isolation, language spoken, gender, security, etc.).

307. 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 who are fluent in the languages and dialects from the displaced persons place of origin. 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.

308. Many field staff are short-term temporary staff and maybe have limited experience or training in statistical data collection activities. The training should be delivered as close as possible to the time at which it is to be used in fieldwork. The training must follow a cascade principle: field work coordinator and supervisors must be trained by NSO staff and enumerators must be trained by supervisors.

309. It is important to be give staff training to understand the following, and above all to be motivated to collect good quality data:

1. The importance of their duties,
2. Confidentiality issues, particularly when dealing with vulnerable respondents,
3. The definitions and concepts used in statistics about displacement which may be unfamiliar to many experienced field staff (place of habitual residence, classification categories etc,)
4. The way they are expected to undertake those duties,
5. Administrative issues,
6. Definitions, mapping and other concepts,
7. Quality assurance.

3.1.3 Development of collaboration agreements

310. **A communication plan** must be developed in advance in order to explain the purpose of the survey to the respondents. The aim of a communication and publicity program is to support the operation of the survey and obtain a high-quality result through the following measures:

- Organizing a recruitment campaign for field staff, and in particular members of displaced groups to join the field force.
- Getting information about the reactions to the survey plans and activities from displaced people
- Organizing an awareness campaign to:
 - Maximize awareness of when the survey will be carried out,
 - Promote awareness of procedures and ways to get assistance with the survey topics,
 - Address any issues that need clarification, and in particular addressing the needs of displaced people,

- Contacting organisations and community groups relevant to displaced people in order to get the full cooperation of respondents, and in particular building trust in the data collection activities and the uses of the data.
- Encourage respondents to cooperate to the best of their ability, and to ensure the salience of the topics to the displaced people to be interviewed.

311. **Ethical aspects and vulnerable respondents.** Interviewing Refugees or IDPs presents sensitivities, as any IDP or Refugee who identifies themselves as such could face threats from doing so or be stigmatised. 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. All enumerators, coordinators, data managers, and anyone involved in the development of the official statistics must be aware and follows such principles. (See e.g. [IOM data protection manual](#), [IOM Data protection initiative](#), [Handbook on Data Protection in Humanitarian Action](#)).

3.4 Test production system

312. This section includes information on how to test the production system and the statistical business process is relevant when building new statistics whatever the covered statistical field. For example, providing list of the objectives as well as the types of tests to be used and best practices, even if it is generic, can be useful for compilers'.

313. **Test collection tools** on a number of units before the pilot survey. In a survey the collection tools consist of questionnaires and /or electronic data collection devices. This should be tested on a small number of units to make sure that the questions are well formulated, coherent and well sequenced. If using CAPI the collection program should be tested, and any bugs identified before the pilot survey. Some points to consider are:

1. How the respondents interpret the questions and are required to think to come up with an answer.
2. Clarity of the questions.
3. The processing capacity of the technology (long questionnaires may require a lot processor capacity and internal memory of the device used, especially as the interview often needs to run applications for maps and interview administration simultaneously).
4. Need to adapt the questions to the country context (i.e. taking into account ethno-cultural characteristics and potentially sensitive topics for vulnerable groups).
5. Sending out the final version of the questionnaire to the experts and line agencies/ministries for further recommendations and their comments.

314. Technical issues which can limit the efficiency of the data collection process or cause errors, these include;

1. Size of screen (may affect the ability of the enumerator to record responses correctly);

2. Length of battery life (considering daily workload of the field staff particularly in challenging environments with poor access to electricity);
3. Processing capacity;
4. Storage capacity.

315. Testing of the data transferring tools for transferring information from the data collection instruments to secure storage in databases. The transmission of the data must make use of secure channels, using IT encryption and state-of-the-art data security standards. This is particularly important when dealing with vulnerable respondents where data security breaches may impact on their safety and security.

316. **Testing of data treatment tools:** The data processing involves a series of steps to convert the responses to the electronic questionnaire from their raw format to a high-quality user-friendly database with a comprehensive set of variables for the analysis. The processing tools consist mainly of data capture and editing software including the SCPRO developed by US Census, SPSS, and of course Excel. After collecting and entering data, several test operations are performed to remove errors in the files, to check the data to assess whether or not errors are of sufficient magnitude to impact on the results, to ensure consistency, code open questions, create variables useful for the analysis of data. and, finally, systematize and document variables to facilitate their use for analytical purposes.

317. **Testing of dissemination tools:** All information gains value through its use. This is to test the media that will be used to ensure the dissemination of data. This may be a program designed to make the case of Cameroon accessible to the database or to the data file, such as ANADOC. Other media to be tested may be the capabilities and quality of the CD-room and USB key, the mock-ups of the leaflets and brochures as well as the report formats.

Phase references and links

- Reference on protecting respondents and their data; [IOM data protection manual](#), [IOM Data protection initiative](#), [Handbook on Data Protection in Humanitarian Action](#)).

Phase 4. DATA COLLECTION PHASE

'This phase collects or gathers all necessary information (data and metadata), using different collection modes (including extractions from statistical, administrative and other non-statistical registers and databases), and loads them into the appropriate environment for further processing. Whilst it can include validation of data set formats, it does not include any transformations of the data themselves, as these are all done in the "Process" phase. For statistical outputs produced regularly, this phase occurs in each iteration³⁵' (UNECE, 2013).

318. This section includes advice on collecting all four data source types (censuses, surveys, administrative data and integrated data). It includes the process of making sample selections from the frames developed in *Phase 2.4 Design Frame* and the training components of field data collections using the instruments designed in Phase 2 in A. Census, B, Surveys and C. Administrative Data, 2.3. *Design collection*.

Introduction

319. Data collection can include multiple different methodologies. This includes census, surveys, administrative data and for data integrated from several sources (See Phase 2D 2.3 *Design Collection*). This section focuses on surveys, administrative data collections are described in several of the examples in Part II of this manual - Case Studies (see Bosnia and Herzegovina, Norway, Turkey and Ukraine)

4.1 Create frame and select sample for surveys

320. This is covered in Phase 2, 2.4 *Design Frame*

4.2 Set up collection

321. Setting up data collection in situations involving refugees and internally displaced persons (IDPs) poses additional hurdles that one has to consider and try to meet before and during the data collection phase. The case studies for South Sudan and Yemen describes some of the difficulties involved in collecting data in dangerous environments (see Part II of this Manual; *South Sudan High Frequency Survey* and *Yemen IDP Example*).

322. The data collection phase encompasses the preparation for the actual field work activities. Set up of data collection can be divided in three main activities. These include: the survey instrument preparation, recruitment and training of field staff and actual data collection.

1) ENSURING COLLABORATION BY HOUSEHOLDS

323. The most important way to ensure collaboration by the households is to establish a good with local communities about the purpose and conduct of the survey. Efforts may be needed at the national, local and particularly at the community level. The survey should use polite, diligent, well trained interviewers and to ensure that the household is contacted and a time convenient for the interview arranged. Some additional measures may be needed. There are no fixed prescriptions that will work everywhere, but experience from various countries should be observed and evaluated.

³⁵ Ibid paragraph 65

324. **Use of Mass Media.** In general, the use of mass media is a waste of money because the mass media reach many people whom the survey does not. However, if mass media coverage can be obtained free it can be useful. Even if it is limited to a short newspaper story or some radio or TV briefs at the start of the survey, it can boost the field teams' morale and self-confidence at this critical time. (Occasionally Interviewers keep the old newspapers throughout the entire survey period, to show the households that they are official and serious.)

325. **Targeted Publicity.** This may include letters to the households and leaflets (preferably in colour, with graphs or other illustrations) that explain the purpose of the survey and the sampling methodology in simple terms.

326. **Material Incentives.** Sometimes a gift or payment is given to the households in return for their collaboration. There is some controversy about the quality and quantity of the material incentives that should be used to foster the household's collaboration. Some consider incentives to be standard procedure for all surveys. Other statistical agencies are reluctant to even consider the idea of rewarding the households in any way, to prevent households from becoming increasingly demanding and affecting all the household surveys conducted in the country. A relatively inexpensive alternative, likely to be cost-effective and be accepted in all countries, consists of giving away small presents for the interviewed households. Ideally, the giveaways should have little or no intrinsic value. This both ensures that they do not affect the household welfare measurement and reduces the accounting controls required.

327. **Community Level.** Publicity and motivation at the local community level are especially important in rural areas. Local authorities should be contacted and convinced of the usefulness of the survey. After this meeting the Interviewers can contact the selected households to introduce themselves and to make appointments for interviews. As an example, when doing a pilot survey in an IDP camp in Darfur, Al-Fasher, the field teams failed to approach the traditional leaders. This almost compromised the fieldwork operation. The problem was exacerbated as local IDPs were not approached or recruited to join the fieldwork team although this may have raised some problems over confidentiality, roles for local people in field operations may be advantageous in gaining local support .

2) SURVEY INSTRUMENT PREPARATION

328. The survey data collection phase requires the preparation of the survey instrument and the design of the questionnaire to be administered to respondents (see also *Phase 3 Build a system for collection of data and production of statistics*). There are two methods of administering household survey questionnaires: Paper and Pencil Interviewing (PAPI) and electronic data capture. PAPI uses paper questionnaires for interviews and the data can be recorded in two ways: i) data entered in a central location, and ii) data entered in the field, known as Computer Assisted Field Entry (CAFE). CAFE is preferable because it facilitates electronic data checks while data collection is ongoing, which could improve data quality. Data checks are performed during fieldwork, thus allowing for the possibility of correcting the data while Interviewers and respondents are still available. If any errors are flagged during field entry, the Interviewer can return to the household the next day to revise and/or make necessary corrections.

329. Electronic data capture can be done in a variety of ways: 1) Computer Assisted Personal Interview (CAPI), which uses tablets; 2) Computer Assisted Web Interview (CAWI), which is typically conducted via email and might not be feasible for certain populations because it is

self-administered; and 3) Computer Assisted Telephone Interview (CATI) in which the data is collected over the phone and therefore has a time constraint.

330. CAPI is increasingly becoming the survey instrument of choice as it allows for faster data collection and better-quality data, by allowing for immediate in situ data validity checks. The decision to use CAPI will depend on the capabilities of the organization that collects the data and the interests of the organization that finances the data collection. Such capabilities are measured in terms of: (a) technical ability, i.e. skilled data programmers and Interviewers; (b) the communications system and infrastructure; and (c) the ease of access to the equipment, given the procurement procedures.

331. Field work using either paper questionnaires and CAFE; or tablets and CAPI face constraints while in the field, and this can be particularly problematic in situations of disaster or displacement.

1. First, without access to electricity, it will be difficult to operate the data entry computer. Options to resolve this problem include providing each team with a set of solar panels, high-performance batteries, or other electronic equipment. In addition to the electronic equipment, the teams have to transport the computer itself, the printer and a sufficient supply of paper, diskettes, and so forth. The advantage is that this will allow proven field and data entry procedures to be used with only minor modification. The modifications that are being made relate mostly to reducing the amount of paper used in the supervision and data management process, in order to reduce the weight that has to be carried around. The main disadvantages of this approach to mobile data entry are the risk of something going wrong with such an elaborate setup and the weight the teams will have to carry.
2. Second, if there is no electricity, recharging the batteries for the tablets used for CAPI collection can be affected. A supply of extra batteries should be provided to each team, however in situations where this cannot be easily or safely done PAPI methods may be preferable
3. Finally, the lack of Wi-Fi access can have serious adverse effects on the data collection whether CAFÉ or CAPI is used. Sending data for central processing over the internet can help speed up overall completion of the survey implementation. If CAFÉ is used, the Data Entry Operator can send data to headquarters as soon as all the checks have been completed. If CAPI is used, it may, depending on the software used, require internet connectivity to implement the questionnaire. With either method, having WIFI connections allows the teams to stay in the field for longer periods of time without having to go to headquarters to deliver data.

3) PILOTING FIELD PROCEDURES

332. The pilot test for field procedures is less to determine whether they can work in general than to fine tune the details of how they are implemented in the specific country. Most problems found during the field test fall into three classes:

1. **Refining Logistics.** In spite of careful planning, some problems regarding the supply of one survey materials is always found.
2. **Debugging the Data Entry Programme.** The functioning of the data entry

program has to be tested. Despite extensive testing problems arise when data from numerous real households are entered. Additional consistency checks are likely to be required that were unforeseen by the survey Data Managers. These will become obvious after the first few days or hours of field work, early field tests are particularly useful.

3. **Statistical Quality Control of the Data.** Concurrent data entry makes it possible to conduct some preliminary statistical analysis of the data collected as the survey progresses. From an analytical standpoint, the data from just a portion of the field work has no statistical significance unless the total sample is exceptionally large, but it often can provide interesting insights into the quality of the field work. For instance, after the first period of field work, the frequency distribution of the last digit of the ages recorded in years may show too large a proportion of "zeroes" and "fives" (demographers always expect this to happen, but it should always be checked). The early detection of this problem allows for corrective re-training and can also give the data entry program credibility in the eyes of the field teams.

333. Regular supervision that includes practical suggestions to practical problems can be helpful in maintaining motivation, morale and professional standards. As discussed in the section on the Supervisor's duties, supervision should be extensive in any survey to ensure quality data collection.

4) RECRUITING AND TRAINING OF FIELD STAFF

334. Prior to data collection, the training plan for both the field staff and the management staff should be laid out. Normally this will include training in the techniques required to collect the data required or to compile the data. In addition, if an instrument will be used in the field, then specific training runs of deploying the instrument are necessary.

335. Hiring good field workers is often difficult. Because it is often difficult to assess all the required qualities in a short job interview, more field staff should be hired than will be required, usually 15 to 25 percent more. All the hires are trained. During the training period the candidates' work characteristics and ability to establish rapport with interviewees is displayed and can be assessed more accurately. Then, after the training period, a final selection is made.

336. Good training contributes greatly to the quality of the data collection effort. Training is one of the hardest aspects of field work to convince people to do well. Training is expensive, takes time away from actual collection of data, and is often viewed as an unnecessarily complicated task. The most common complaint is that the Interviewers already know how to interview and there is little need for additional training.

337. Having Interviewers who are previously trained in how to interview can be an advantage. Interviewing techniques are difficult to learn - how to interact with the respondents, how to convince reluctant respondents to participate and actually complete the interview, how to get the attention of the respondents when they have so many other things to do - and are useful in almost all survey situations, regardless of the contents of the questionnaire.

338. However, initially the Interviewers do not know the reasons for this particular survey, the contents of this questionnaire, the analyses that are envisioned for the data or the policy impacts that these data could influence. It is extremely important that enough time is given to

training on the specifics of the survey so that all Interviewers understand the purpose of the survey and all of the concepts in the same way. This is particularly important for NSOs conducting surveys about displacement for the first time, as many of the concepts and definitions are new.

5) TRAINING FOR INTERVIEWING DISPLACED PEOPLE

339. Many of the concepts in the IRRS and IRIS will be new to statistical offices and their staff. All staff involved in any survey or data collection involving displaced people must be trained in these concepts and understand them. Of particular importance is training data collection staff on 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.

340. 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. Because of this, they may fail the interviewers usual understanding of 'usual members of the household' and can be treated as visitors and excluded from household rosters. These exclusions mean that they fail the test of probability sampling, that every person has a known chance of selection in the survey. Data collection staff must be carefully trained in recording all members of the household – e.g. temporary members who may be displaced.

341. Training in the interviewing of vulnerable persons such as children and disabled people. People who have been displaced have suffered trauma and any interview may ask people to relive painful experiences. A UNHCR commissioned survey of IDPs in Georgia in 2014 asking about their intentions was described as an emotional process³⁶.

Box 4.1: Interviewing Vulnerable IDPs in Georgia

Survey on Durable Solutions: Voices of Internally Displaced Persons in Georgia. UNHCR and ISSA, 2014

Two thousand and one (2,001) IDPs in Georgia have shared with the reader of this document 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.

The survey was carried out by the Institute of Social Studies and Analysis (ISSA)

6) TRAINING MANUALS

342. The main written materials used for training field teams (Supervisors, Interviewers, and Data Entry Operators) are the questionnaires and the field manuals. It is recommended to produce many more manuals than needed for training because apart from their obvious use as a support for field operations the manuals are also valuable tools for the survey analysts.

343. The basic contents of each kind of manual is described in much more detail in Part 3 of this Manual; *Training Needs Plan*.

4.3 Run Data Collection

344. This section discusses the main features of preparation for field work, other than designing the questionnaire or drawing the sample. Many of the data quality control mechanisms rely extensively on the preparation for the field work. Survey preparation is thus very important and should be given due time and attention. Often there is the temptation to skimp on preparation in order to move to the field too rapidly. This temptation should be avoided.

345. The survey's success relies on its staff. Here are some basic requirements that staff should have given their role in the data collection activities:

- **The Survey Manager** should be a social scientist or statistician with a grasp of the goals to be achieved by the survey. Often this person will have a graduate degree in statistics, economics, or demography. At a minimum the Survey Manager should have a lower university degree.
- **The Data Manager** can be a systems analyst or senior programmer with prior experience in statistical data management.
- **The Field Manager** should have substantial managerial skills and experience in conducting household surveys in the country. All members of the core staff (not just the Data Manager) should be familiar enough with personal computers to use word processing software and spreadsheets.
- **The Supervisors** should have completed secondary education. Experience in motivating people and managing resources; and the ability to foster teamwork are more important than credentials.
- **The Interviewers** should also have a good secondary education, but they do not need to have pursued further studies. In fact, that may be a disadvantage, as graduates tend to be more likely to abandon the survey halfway through if they are offered a better or more interesting job.
- Similarly, experience with computers is not an essential requisite for the **Data Entry Operators**, but it is helpful if they have keyboard skills. It is not difficult to learn to enter data and it is not necessary to understand how a program works in order to use it successfully. It is better, in fact, if the Data Entry Operators are interested in the survey rather than in computers; that way, when they note incorrect answers, they are more likely to use the same terms as the questionnaire instead of explaining the error in

computer terms.

346. It may also be desirable that team members, especially Interviewers, be fluent in two or more languages, and in when interviewing displaced people interviewers must be fluent in the languages and dialects that they are likely to encounter. Moreover, it is best if the Interviewers in each team have complementary language skills so that between them, so they can conduct all (or nearly all) the interviews themselves.³⁷

347. Unless specific cultural or religious conditions dictate, there seems to be no *a priori* reason for preferring male to female Interviewers or vice versa. However, this should be adapted to the cultural context, e.g. the NSO in Sudan aims to use male interviewers among the nomads, interviewers with military background when addressing military personnel, female interviewers when interviewing women on pregnancy, birth, breastfeeding etc. There is also some anecdotal evidence, that households are less likely to refuse women interviewers. Prior experience in the country may show that it is culturally more acceptable for female respondents to admit female interviewers to their homes when the men were absent. Even when male interviewers are used, anecdotal reports from several countries suggest that females may find it easier to establish rapport for the modules with sensitive content.

348. In some surveys, questions may be gendered with sensitive questions for males and females, an interviewer should be selected of the same gender to enable the interview to obtain the answers needed with as little social embarrassment as possible, and in privacy.

1) TEAM AND FIELDWORK LOGISTICS IN HAZARDOUS ENVIRONMENTS

349. Collecting survey data in hazardous or conflict environments is marked by specific situations that may be hard to plan for. Parts of a country may be inaccessible due to the presence of terrorist groups or violent clashes between various power factions may occur throughout the territory. In addition to conflict situations, the weather phenomena may cause severe droughts or severe floods that exacerbate pre-existing vulnerabilities in the population. Both conflict and weather conditions have led to large-scale internal displacement throughout the world.

350. Countries with conflict are often data deprived. Specifically, years of civil war or ongoing conflict will erode the statistical infrastructure and capacity, leading to the lack of key macro- and micro-economic indicators. The last full population census, needed to develop the sample frame, may be long out of date. In addition, risks to the safety of field staff may require spending as little time in enumeration areas as possible. While organizations operating within the country may implement a range of smaller surveys, these cannot meet the data needs for all policy interventions.

351. Several techniques have been developed to consider the special circumstances of trying to collect data in hazardous or conflict situations. To address security concerns, the Somalia High Frequency Survey adapted logistical arrangements, sampling strategy, and questionnaire design to limit time on the ground. In logistical arrangements, a detailed and timely security

³⁷ If interpreters are needed, they are usually other members of the household or community members acquainted with the respondent. Though both the presence of a third party and the fact that the interpreter is acquainted with the respondent violate some of the basic rules of interviewing, this may be the only possible option for some interviews. Interviewers should be aware - and be trained to avoid - a still more serious problem posed by a local interpreter, which is a tendency to answer for the respondent.

assessment ensured that the enumeration areas to-be-visited were safe on the day of fieldwork.³⁸ The fieldwork protocol was designed such that teams would spend as little time as possible in any given region and draw little attention, ensuring enumerator and respondent safety. Concerning sampling strategy, 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 are small enough for enumerators to list and select households immediately before conducting the interview.

352. Other techniques include enlisting the assistance of local community groups in conflict areas to assist in the collection of the data. Using local community groups who have the trust of the group of interest can bridge the communication gap between the data collection effort and the conflict area groups. When using this technique, it is important that the local community groups are contacted early in the survey design process and allowed to participate in the preparation of the survey. It is important that communities in the conflict areas feel that their concerns have been addressed.

353. A sufficient initial stock and a steady supply of survey materials and inputs should be ensured for all field teams throughout the whole survey period. The list of materials includes obvious items such as questionnaires/tablets, pencils, and erasers, less obvious ones such as USB keys, clipboards, and briefcases, and a myriad of country-specific items ranging from raincoats to camping stoves.

354. The elements that have proven hardest to manage in all countries are fuel, oil, and everything related to car maintenance. Probably the best solution is to provide a cash revolving fund and make the Supervisor accountable for it. This may be hard to do in more bureaucratic organisations.

2) CONDUCTING THE INTERVIEW

355. In each interview the Interviewer will go through all pertinent modules in the order they appear in the questionnaire. Interviews should be conducted with direct respondents as often as possible. For those respondents who are too young - children under 10 years of age - the adult who is most responsible for the children should respond on their behalf. Proxies will also be used for respondents who are temporarily absent or who are unable to respond for themselves due to illness or disability.

356. This practice enhances data quality in several ways. First, since the person best qualified to answer each part of the questionnaire is interviewed, inaccurate responses are avoided. Second, the whole survey, which can take considerable time in each household, is broken into a series of more manageable mini-interviews. Respondent fatigue is thus minimized. Third, since the mini-interviews are scheduled according to the respondents' convenience, the refusal rate is also minimized. However mini-interviews can also introduce logistical problems for both interviewers and respondents, and the context should be taken into account, particularly in hazardous environments where travel needs to be minimised.

357. The actual length of the interview varies widely from one household to another and

38 Utz Pape and Philip Wollburg. "Estimation of Poverty in Somalia Using Innovative Methodologies." World Bank Policy Research Working Paper 8735, February 2019.

between countries. The difference between countries depends on the length of the questionnaires used. Within countries, interview time varies depending on the number of people in the household. However, given that Interviewers generally need to visit each household several times, it is more useful to evaluate interview time (and consequently Interviewer productivity) in terms of "households per week" or "households per day" rather in "hours per household."

358. Field work with simultaneous data entry offers several powerful advantages. First and perhaps most important, it raises data quality. Simultaneous data entry, whether CAFÉ or CAPI makes it possible to correct mistakes while Interviewers are still in the household or at least in the area.

359. Secondly, simultaneous data entry can also shorten the time needed for data entry, and speed up the editing and analysis steps after the final interview. Thus, the goal of the timely availability of the data is accomplished.

360. There are a few disadvantages to this type of field organization. For one thing, the field workers need to be highly competent. Moreover, this strategy may also increase the survey costs.

3) DEVELOPING SUPERVISION FORMS

361. Three of the tasks of the team Supervisors should be supported by written documents, known as supervision forms. These are (1) Interviewer evaluation, (2) questionnaire verification, and (3) check-up interviews. The forms are intended to give these tasks formal definition, as opposed to loosely defined responsibilities left to the Supervisor's personal initiative, and to make it possible to supervise the Supervisors themselves (e.g., make supervision tasks verifiable by the survey core staff).

4) SCHEDULING FIELD WORK

362. The task assignment of each team should be done concurrently with the first stage of sample selection. The clusters are distributed among the field teams and the order in which each team will visit the clusters assigned to it is decided at random. The schedule of each team should then be made explicit, to indicate what each team will be supposed to be doing each week of the survey year.

363. Designing the actual schedule is very country-specific, as it is usually developed around national holidays and other significant dates, with the goal of either excluding or including them in the work period. In Muslim countries, for instance, the Ramadan month is particularly interesting to observe because of the differences in the household consumption patterns; Ramadan, however, is not a good month to train Interviewers or to initiate the survey.

5) COMPOSITION OF SURVEY STAFF

364. The key headquarters staff includes the Survey Manager, Data Manager, and Field Manager. These are the minimum requirements. Sometimes it is useful to employ a secretary and a bookkeeper. The need for a bookkeeper will be greater where financing is being provided by more than one agency or where substantial procurement will take place.

365. The core staff should be organized to work together as a team, with the Survey Manager

being the only head during the entire preparation stage instead of having individuals respond to separate divisions in the data collection agency. This is especially important - and sometimes hard to achieve - in large national statistical agencies that are organized under the traditional departmental structure with a census department, a household survey department, a data processing department, and so forth.

366. Each of the field teams is headed by a Supervisor and usually includes three to four Interviewers, a driver with a car, and a Data Entry Operator if CAPI is not used. This standard setup is used for most surveys. In some countries local conditions have dictated changes in the composition of the field teams:

367. When a team must work mainly in large urban areas, an additional Interviewer may sometimes be added. This allows for additional interviews to take place at a low marginal cost.

368. Cultural constraints in certain countries may require that an adult must be interviewed by someone of the same sex. Since it would likewise not be appropriate for a female to travel alone with several men, each team may need to have two female and two male Interviewers. In some cases, a relative of the female Interviewer(s) may need to travel with the team.

6) DUTIES OF SURVEY TEAM MEMBERS

369. The composition of a typical survey team is described in detail in Part 3 of this Manual. See - *Training Needs Plan*.

Phase Links and Examples of Training Manuals and Data Collection Advice

- **Training Manual for Civil Registration Functionaries in India, 2012** (Hindi and English); http://www.censusindia.gov.in/2011-Documents/CRS_Report/CRS_mannual_Hindi_English.pdf .
- 2020 Census Complete Count Committee - Training Manual, United States of America: <http://azmag.gov/Portals/0/Documents/MagContent/CCC-Training-Manual.pdf?ver=2018-04-25-145518-867>
- JIPS data collection advice. **Implementing Data Collection - Operational Plan for Data Collection**. This document outlines an operational plan for the data collection phase of the household survey during a profiling exercise. This document helps plan out how the fieldwork will actually be conducted. It includes the operational structure showing how the implementing/data collection partners fit in, the roles and responsibilities for the various actors involved and the work plan for the next steps. <https://jet.jips.org/phase/implementing-data-collection/>
- **UNHCR Needs Assessment Handbook**. The Handbook is structured in two parts. The first, which is recommended for all audiences defines need assessments and their different types; describes coordination modalities; outlines the roles and responsibilities of different actors in refugee situations, IDP situations, and mixed situations; provides an overview of the steps to conduct needs assessments and the principles that should guide them; and explains the relationship between needs assessments and other information systems. The second part of the Handbook provides detailed practical guidance on how to conduct needs assessments in the field. It can be used as a reference text, with readers referring to specific steps and sections as needed based on their role in the operation or the needs assessment, and the type of situation. <http://needsassessment.unhcr.org/>

Compilers' Manual Part 1 - Phase 4 Data Collection

- Johannes Hoogeveen, Emil Tesliuc, Renos Vakis with Stefan Dercon, A Guide to the Analysis of Risk, Vulnerability and Vulnerable Groups; <http://siteresources.worldbank.org/INTSRM/Publications/20316319/RVA.pdf>
- World Bank. *Informing Durable Solutions for Internal Displacement in Nigeria, Somalia, South Sudan, and Sudan: Volume C: Technical Aspects*. <http://documents.worldbank.org/curated/en/142811557466057440/pdf/Volume-C-Technical-Aspects.pdf> see also https://link.springer.com/chapter/10.1007/978-3-030-25120-8_11
 - Several data limitations prevent an accurate assessment of socioeconomic conditions among displaced people and hosts, which hinders efforts to design targeted policy interventions. This report addresses a key survey methodology question in the displacement context. In response to under-reporting of consumption patterns (in particular, food consumption), the study proposes the adoption of behavioral nudges in survey design. To mitigate potentially spurious responses, including ‘honesty primes’ in the consumption module of the survey’s questionnaire has led to more accurate reporting. The report also proposes clustering approaches to derive typologies of IDPs, to inform the required specificity of programs, and to find durable solutions.

Phase 5. PROCESS DATA

“This phase describes the cleaning of data and their preparation for analysis. It is made up of sub-processes that check, clean, and transform input data, so that they can be analysed and disseminated as statistical outputs. It may be repeated several times if necessary. For statistical outputs produced regularly, this phase occurs in each iteration. The sub-processes in this phase can apply to data from both statistical and non-statistical sources (with the possible exception of sub-process 5.6 (Calculate weights), which is usually specific to survey data)”. (UNECE, 2013)

370. This section of the Compilers' Manual will cover using the data from various sources and will in particular focus on processes for integrating and cleaning data from all four data sources.

371. This is the main step for preparing data for analysis. It is the step for describing the data and determining the type of variable; Numeric or string, length of the variable to cover the categories. Moreover, we need to check, clean and code the data, and check the consistency between variables. Then after this data is created to calculate the weights, if the data is derived from a survey. The steps would include;

5.1. Integrate data

372. In developing variables for use in the analysis of data, it will often prove necessary to recode and classify raw data in order to derive the variables which are required for analysis. Or to merge data from other relevant data sources. Data Integration is discussed in detail in Phase 2, Data Source D: Integrated Data. Readers should refer to this section for more information.

5.2 Classify and Code

373. Classification and coding should be carried in accordance with the classifications and variable descriptions described in Phase 2.2. *Design variable descriptions*.

1) EXAMPLE: CLASSIFYING AND RECODING ADMINISTRATIVE DATA - GERMAN FEDERAL STATISTICAL OFFICE

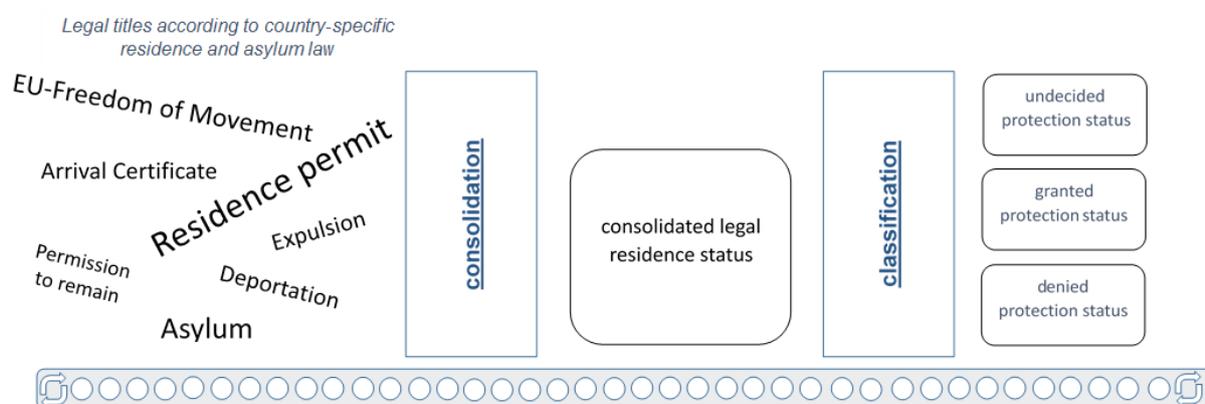
374. This example describes recoding data to produce statistics on refugees in accordance to the IRRS. Since 2017 the German Federal Statistical Office compiles statistics on foreigners seeking humanitarian protection based on the Central Register on Foreigners (CRF). In light of the unprecedented inflow of asylum seekers between 2014 and 2016, the Federal Statistical Office explored the feasibility of complementing existing statistics on the inflow of asylum seekers with figures about the stock of foreigners living in Germany for humanitarian reasons. The approach relies mainly on administrative residence permit data obtained from the CRF. In general, administrative residence status data allows for taking a comprehensive view on humanitarian immigration that not only includes foreigners with a granted protection status, but also those with an asylum application still pending, as well as foreigners who remain in the host country after having been denied protection. Thereby, administrative data allows for producing official statistics that follow the International Recommendations for Refugee Statistics (IRRS) that were adopted by the United Nations Statistical Commission in March 2018.

375. A necessary first step is to understand the scope of the administrative data at hand. In Germany, the CRF is the central database for all authorities responsible for administrative procedures concerning foreigners. It contains basic personal information on foreigners from

EU member countries; additional information on legal residence status for foreigners from Non-EU countries; and even more detailed information (e.g. on educational background and language skills) on foreigners seeking asylum in Germany. Most importantly for classifying refugees and people in refugee like situations, the CRF centrally collects all information concerning foreigners' legal residence status (e.g. current residence permit) that is locally produced by immigration offices. As an important limitation, the register only covers foreigners. Once a person receives German citizenship, s/he is immediately deleted from the register. Hence, second generation and former refugees who acquired German citizenship cannot be identified with the data at hand.

376. In many countries, including Germany, residence and asylum law is excessively complex. When classifying foreigners seeking protection, numerous types of residence permits, and a wide range of additional legal statuses must be considered. Accordingly, filtering and structuring the relevant legal information that reliably identify the target population is key in compiling statistics on refugees and people in refugee like situations from administrative data. The data usually does not come in a form that is ready for use for statistical purposes. Before foreigners seeking protection can be identified in the data, the various variables related to the legal residence status must be consolidated. In this process the relevant legal status for each person in the register is determined. For instance, consider a Croatian citizen who fled persecution during the war for independence from Yugoslavia in the early 1990s. Further, imagine this person receives a permanent humanitarian residence permit in Germany. Even three decades later, this residence permit is still valid and recorded in the register. However, it is no longer relevant for this person, since Croatia became a member of the European Union in 2013 and meanwhile Croatian citizens enjoy the right of free movement in Germany.

Figure 5.1: Statistics on foreigners seeking protection



377. Building upon the consolidated legal residence status, the target population can be classified. People seeking humanitarian protection constitute a sub-group of the foreign population who specifically refer to humanitarian reasons for their current stay in Germany. This population in turn consists of three distinct sub-groups that are distinguished by whether their request for humanitarian protection is yet unsettled, granted or denied.

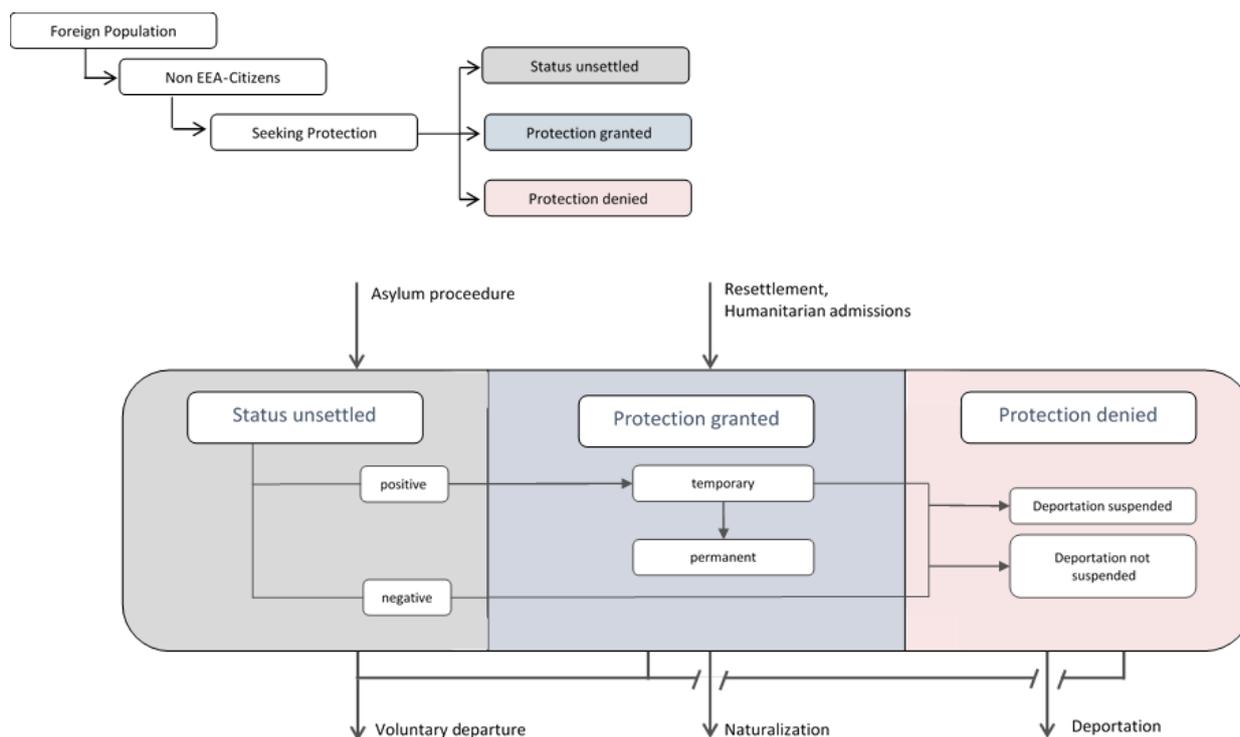
378. A foreigner's purpose of residence is inferred from his/her legal residence status. One necessary precondition for identifying foreigners with a granted protection status is that residence permits reveal the underlying purpose of residence. In Germany, information on the purpose of residence can be deduced from a foreigner's current residence permit since in 2005 the German Residence Act became effective. Since then residence permits distinguish between

residence either for educational purposes, for the purpose of economic activity, for family reasons or based on humanitarian grounds. Residence permits based on humanitarian grounds can either be temporary or permanent.

379. Besides foreigners already holding a humanitarian residence permit, asylum-seekers residing in Germany who intend acquiring a humanitarian residence permit also belong to the stock of foreigners seeking protection. Foreigners with an unsettled protection status have either expressed a request for asylum towards German authorities (e.g. police or immigration office) or they already filed an asylum application at the Federal Office for Migration and Refugees.

380. Foreigners whose asylum applications were denied but who still reside in Germany constitute the third final sub-group of foreigners whose request for humanitarian protection has been denied. The sub-group of foreigners with a denied protection status also includes foreigners who previously held a temporary humanitarian residence permit that was either not prolonged, renounced or withdrawn. All foreigners belonging to this category are legally obliged to leave the country. However, in many cases this obligation cannot be enforced due to practical (e.g. missing passports) or legal (e.g. family ties) reasons leading to prolonged durations of stay in this status.

Figure 5.2: Statistics on foreigners seeking protection



381. By including granted and unsettled cases, the outlined concept follows the IRRS. However, the definition of foreigners seeking protection deviates from the IRRS in one aspect: The group of foreigners with a denied protection status does not include persons who entered on tourist, student or work visa but who have nonetheless fled persecution, war or violence in their home countries (IRRS 2018: pp. 26-29). This group cannot be identified in the administrative data at hand.

382. Also see, **UNECE-Eurostat Work Session on Migration Statistics October 2018:** https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.10/2018/mtg1/DEU_Asym ENG.pdf

5.3 Review and validate

383. This would include reviewing the data and assessing its quality, including checking for consistency, outliers and missing values (see also Phase 6.2 paragraph 463).

1) EXAMPLE: HOLOCLEAN PROJECT

384. “HoloClean is a statistical inference engine to impute, clean, and enrich data. As a weakly supervised machine learning system, HoloClean leverages available quality rules, value correlations, reference data, and multiple other signals to build a probabilistic model that accurately captures the data generation process and uses the model in a variety of data curation tasks. HoloClean allows data practitioners and scientists to save the enormous time they spend in building piecemeal cleaning solutions, and instead, effectively communicate their domain knowledge in a declarative way to enable accurate analytics, predictions, and insights from noisy, incomplete, and erroneous data.” Further information is to be found at this [Link](#).

2) EXAMPLE: JIPS GUIDANCE ON DATA PROCESSING:

385. This guidance includes the following topics. 1) Why processing the raw data is so important, 2) Checklist for processing the data, 3) Before you start, 4) Preparing the dataset, 5) Cleaning common errors, 6) Checking for unexpected values, and 6) Preparing variables for analysis. The link to this guidance is : <https://jet.jips.org/wp-content/uploads/Guidance-Data-Processing-Phase5-JET.pdf>

3) EXAMPLE: DATA QUALITY PROBLEMS IN ADMINISTRATIVE REGISTERS AFTER A MASS INFLUX OF ASYLUM SEEKERS - FEDERAL STATISTICAL OFFICE OF GERMANY

386. 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;

387. **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 lack.

388. **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.

389. **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.

5.5 Derive new variables and units

390. This sub-process derives data for variables and units that are not explicitly provided in the collection but are needed to be estimated or derived to deliver the required outputs specified in Phase 2 "Build".

- Computing variables to measure the characteristics of IDPs and refugees and refugee related populations. Some examples of this, depending on user demands are
 - Female headed household
 - Family type/household composition
 - Employment or labour market activity rate
 - Total consumption
 - Regional classifications
 - Unaccompanied minors
 - Returned to place of habitual residence
 - Having overcome key displacement-related vulnerabilities (IDPs)
 - Measures of integration (refugees)
 - Durable solutions progress measure
 - Refugee/IDP status
 - Total household consumption/income

391. Computing a variable for measuring progress towards a durable solution and

identifying those who have overcome key displacement-related vulnerabilities is discussed in the IRIS, 2020. A set of indicators and related questions will be developed based on the 10 identified sub-criteria for overcoming key displacement related vulnerabilities (IRIS, Table 4.3) as well as sub-criteria related to the three remaining criteria for measuring progress towards durable solutions (see IRIS Chapter 4).

392. The methodology for developing these indicators and a progress variable will be added to this Manual in future editions, once more data becomes available and the methodology agreed.

393. Computing variables for measuring integration and the satisfaction of immediate and on-going needs of refugees and refugee related populations are set out at length in IRRS, 2018 Chapter 5. Further discussion of the indicators used to measure European refugee integration can be found at link https://ec.europa.eu/home-affairs/sites/homeaffairs/files/e-library/documents/policies/legal-migration/general/docs/final_report_on_using_eu_indicators_of_immigrant_integration_june_2013_en.pdf.

394. Both the IRRS in Chapter 5, and IRIS in Chapter 4 Section B, discuss the use of the SDGs indicators for measuring the satisfaction of basic and immediate needs of displaced people, and measuring their progress towards integration or durable solutions. In constructing these indicators the following link to UNSD's SDG metadata repository provides details of constructing the indicators concerned, [SDG Metadata Repository](#).

5.6. Calculate weights for data derived from probability samples

395. This step involves Calculating weights for displaced populations obtained from household surveys based on probability samples to derive an estimate for the population (including hidden households, frame problems etc.).

396. Utilise contents from tool sponsored by UNHCR and developed by Stats4SD <https://stats4sd.org/resources/resource-bank/>

Phase references and links

- HoloClean: A Machine Learning System for Data Repair and Predictions on Structured Data. It is a statistical inference engine to impute, clean, and enrich data. As a weakly supervised machine learning system, HoloClean leverages available quality rules, value correlations, reference data, and multiple other signals to build a probabilistic model that accurately captures the data generation process, and uses the model in a variety of data curation tasks. HoloClean allows data practitioners and scientists to save the enormous time they spend in building piecemeal cleaning solutions, and instead, effectively communicate their domain knowledge in a declarative way to enable accurate analytics, predictions, and insights from noisy, incomplete, and erroneous data. Link <http://www.holoclean.io>
- The link to JIPS Guidance on Data Processing is : <https://jet.jips.org/wp-content/uploads/Guidance-Data-Processing-Phase5-JET.pdf>

Phase 6. ANALYSE DATA

“In this phase, statistical outputs are produced, examined in detail and made ready for dissemination. It includes preparing statistical content (including commentary, technical notes, etc.), and ensuring outputs are "fit for purpose" prior to dissemination to customers. This phase also includes the sub-processes and activities that enable statistical analysts to understand the statistics produced. For statistical outputs produced regularly, this phase occurs in every iteration. The "Analyse" phase and sub-processes are generic for all statistical outputs, regardless of how the data were sourced³⁹”.

(UNECE, 2013)

397. This section of the Compilers' Manual provides detailed guidance on the analysis required to produce the outputs specified in Phase 2.1 Design Outputs (see paragraph 611). The outputs for displacement statistics can be divided into four categories;

- i. Basic tabulations - estimates of counts and proportions of stocks and flows
- ii. Basic indicators (statistics, proportions and estimates)
- iii. SDG indicators recommended for refugees and IDPs
- iv. More complex analysis and commentary which might include an analysis of the integration of displaced people into host communities, the progress made towards overcoming key displacement-related vulnerabilities, and other key policy issues.

398. Taking point iv, the more complex analysis, the different indicators specified to measure aspects of displacement and integration are based on the indicator libraries of the *Integrated Household Survey Framework and the Durable Solution project*, as well as the recommended indicators included in the IRRS and IRIS. They include linkages to SDGs and form part of the reporting of the progress towards meeting the SDGs. In addition, analysis will measure progress towards durable solutions for IDPs, and of the integration for refugees into their host communities. It will also include the assessment of the end of IDP status for some members of the IDP stocks.

399. The analysis process includes two steps:

1. Utilizes the analytical framework specified in Phase 2.1 “Design” to produce the statistical outputs required,
2. Interpretation, commentary and checking the outputs.

400. In essence, it covers the steps required to convert data into a variety of useable products for the users and potential users of statistics. Producing these products to the quality standards designated for official statistics is the ultimate goal of the statistical production process.

401. The text below reminds readers of the recommendations on outputs discussed in earlier phases of this Manual, and links these recommendations to some recent examples, some of which are cited in Part II - Case Studies and Examples section of this manual.

6.1. Prepare draft outputs

402. Much of the analysis phase will be pre-defined by selections made and priorities chosen at the Design phase discussed in Phase 2. This section sets out the basic outputs recommended

³⁹ Ibid para 84

in the earlier phases.

a. Basic tabulations - estimates of counts and proportions of stocks and flows

403. Phase 2 Design defines stocks and flows. A stock corresponds to an already-defined category of individuals and refers to a set of characteristics. 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.

404. For example, IDPs have a specific set of characteristics, which make them fit into this category. This is mainly the forced nature of their displacement, its origin being the habitual place of residence of the concerned persons or groups, the movement remaining within internationally recognised borders and the individuals having not yet found a durable solution to their displacement.

405. In contrast a refugee has been forcibly displaced by reason of a well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group, or political opinion and is *“outside the country of his or her nationality, has crossed an internationally recognised border, and is unable to, owing to such a fear, is unwilling to avail himself of the protection of that country”*⁴⁰ in need of international protection. The statistical categories recommended go beyond these definitions, in order to classify those with a refugee or IDP background and second-generation populations for statistical analysis (see relevant Chapters 3 in both IRRS, 2018 and IRIS, 2020).

406. Flows affect this stock over time, either by making it increase (“in-flows”) or decrease (“out-flows”). This in-flow corresponds to the number of individuals who acquire the characteristics of those in scope of the stock during a specified time period. For out-flows, it is the number of people who cease to be refugees or IDPs, e.g. by changes in status, death, birth or out-migration.

407. There can also be in-flows and out-flows between the sub-stocks, 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). Therefore, there can be several stocks and flows, corresponding to the various classifications of populations being analysed. If the stocks are interconnected, i.e. if a person or group move from one status to another, the flows can also be interconnected; the out-flow from one stock can be the in-flow into another stock.

408. **The basic tabulations of refugee and refugee related stocks and flows** are set out in Phase 2 a. *Refugee statistics outputs* in Table 2.1 Basic Refugee Stock Tabulations and Table 2.2 Basic Refugee Flow Tabulations.

409. **The basic tabulations of IDPs and IDP related stocks and flows** are set out in Phase 2 *IDP statistics outputs* in Table 2.5 - Basic IDP Flow Tabulations, which covers both outflows and inflows to the IDP and IDP-related stock, and Table 2.6 IDP Stock Tabulations.

410. These stock tabulations include both i) IDPs, ii) IDP-Related and iii). Those Who Have Overcome Key Displacement-Related Vulnerabilities. Some analysis will be required to identify the stock of the third category, and to estimate the flows of those who have been

⁴⁰ 1951 Convention(Article 1). See paragraph 28 of IRRS

removed from the IDP stock because they *have overcome key displacement-related vulnerabilities*. Analysis will be needed 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. This is identification process described below in *section d*, but 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.

1) REFUGEE EXAMPLE OF RECLASSIFYING DATA TO MEET IRRS RECOMMENDATIONS – GERMAN FEDERAL STATISTICS OFFICE

411. 'In light of the mass influx between 2014 and 2016, the Federal Statistical Office of Germany explored the feasibility of complementing existing statistics on the inflow of refugees derived from asylum applications with information on the stock of foreigners living in Germany for humanitarian reasons. The concept was designed to take a comprehensive approach by not only including foreigners with a granted protection status but also foreigners with an asylum application still pending as well as foreigners who still remain in Germany after having been refused protection. Thereby, the approach follows the International Recommendations for Refugee Statistics that were recently adopted by the United Nations Statistical Commission. The concept relies on residence permit data obtained from the Central Register of Foreigners. This data allows identifying foreigners currently applying as well as foreigners having been granted or denied humanitarian protection. At the end of 2016, about 1.6 million foreigners were identified as seeking humanitarian protection in Germany. That represents about 2 percent of the resident population and about 16 % of all resident foreigners. During to the recent surge in humanitarian immigration to Germany between 2014 and 2016, the stock has more than doubled'⁴¹.

412. Examples of the tabulations produced can be found in the link to the Asylum in Europe website⁴² <https://www.asylumineurope.org/reports/country/germany/statistics>

2) REFUGEE EXAMPLE OF BASIC STOCK TABULATION – STATISTICS NORWAY

413. An example of a basic stock tabulation is an example from Statistics Norway published in 2019. It provides a stock estimate together with a percentage change estimate for the net flow. An example of output is provided below, the first is taken from Statistics Norway (2019). Norway presents stock figures of '*persons with a refugee background*' with their reason for

⁴¹ **United Nations Economic Commission for Europe**

Conference of European Statisticians, Work Session on Migration Statistics, Geneva, Switzerland, 24-26 October 2018. Item 6 of the provisional agenda Asylum seekers and refugees **Foreigners seeking humanitarian protection** Measuring the stock of foreigners seeking humanitarian protection in Germany Note by German Federal Statistical Office. Link accessed 9 December 2019:

https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.10/2018/mtg1/DEU_Asylum_ENG.pdf

⁴² The Asylum Information Database (AIDA) is a database managed by the European Council on Refugees and Exiles (ECRE), containing information on asylum procedures, reception conditions, detention and content of international protection across 23 countries. This includes 20 European Union (EU) Member States (Austria, Belgium, Bulgaria, Cyprus, Germany, Spain, France, Greece, Croatia, Hungary, Ireland, Italy, Malta, Netherlands, Poland, Portugal, Romania, Sweden, Slovenia, United Kingdom) and 3 non-EU countries (Switzerland, Serbia, Turkey).

The overall goal of the database is to contribute to the improvement of asylum policies and practices in Europe and the situation of asylum seekers by providing all relevant actors with appropriate tools and information to support their advocacy and litigation efforts, both at the national and European level. These objectives are carried out by AIDA through the following activities:

immigration, at the time they immigrated, and this departs from the IRRS as it does not present the stock by current refugee status.

Box 6.1: Norway refugee variables

Statistics Norway Variable Definitions:

- *Persons with refugee background:* refers to people with refugee as reason for immigration, as well as immigrants with family as reason for immigration who are reunited with a person with reason refugee.
- *Type of family unification:* The variable specifies all family immigrations, distinguishing between reunification, accompanying person and formation/extension data. The classification is mainly based on assessments of dates of immigration and marriage (when relevant) of both the immigrant and the reference person, and on registrations of that variable in the UDB data.
- *Asylum seekers* refers to asylum cases or residence on humanitarian grounds.
- *Resettlement refugees* refers to refugees who are permitted to come to Norway following an organised selection, normally in conjunction with the UN High Commissioner for Refugees (UNHCR). In accordance with a proposal from the government, the parliament sets an annual quota for the number of resettlement refugees to be received by Norway.
- *"Other refugees"* refers to refugees with families from mainly Bosnia and Herzegovina who have been granted a collective assessment.
- *Simplified reason for immigration:* Variable based on data from the Norwegian Directorate of Immigration, in which escape has been given as a reason for the application for asylum.

Table 6.1 Statistics Norway Refugee Stock Tabulation

	2019	Change in per cent. 2018-2019	2019	
			Persons with refugee background in per cent of all immigrants	Persons with refugee background in per cent of total population
Total	233 794	2.5	30.6	4.4
Principal applicants	170 087	2.1	22.2	3.2
Asylum seekers	114 028	0.9	14.9	2.1
Resettlement refugees	38 147	7.4	5.0	0.7
Other refugees	9 696	-1.0	1.3	0.2
Unspecified	8 216	-1.0	1.1	0.2
By family connection to refugee	63 707	3.5	8.3	1.2
Family enlargement	17 294	2.1	2.3	0.3
Family reunification	46 366	4.0	6.1	0.9
Family unspecified	47	0.0	0.0	0.0
Source: Statistics Norway				

414. There are several other tables published by Statistics Norway for refugees,

- Persons with refugee background, by years of residence and country

background

- b. Persons with refugee background. Norwegian and foreign citizens, by sex (C)
- c. Persons with refugee background, by age and groups of country background
- d. Persons with refugee background, by duration of residence in Norway and groups of country background
- e. Persons with refugee background, by citizenship, sex and groups of country background

<https://www.ssb.no/en/befolkning/statistikker/flyktninger#relatert-tabell-1>

3) REFUGEE EXAMPLE UNHCR KENYA

415. This is an example of UNHCR refugee supported statistics in a Statistical Summary as of 31-Aug-19 of Refugees and Asylum Seekers in Kenya. It is an example of the production of rapid statistical outputs from administrative data. <https://www.unhcr.org/ke/wp-content/uploads/sites/2/2019/09/Kenya-Statistics-Package-31-August-2019.pdf>

416. The [2019 Economic Survey of Kenya](#) published by the Kenya Bureau of Statistics provides official refugee official statistics on behalf of the Government.

Table 17.22. Refugees and Asylum Seekers in Kenya by Age and Sex, 2014-2018

Category	Number				
	2014	2015	2016	2017	2018*
Children (< 18 yrs)					
Male	173,573	174,344	149,972	134,109	132,999
Female.....	155,350	157,276	134,173	138,994	122,178
Sub Total-children	328,923	331,620	284,145	273,103	255,177
Adults (18+ yrs)					
Male	123,991	129,616	100,781	103,299	105,318
Female.....	132,449	132,645	109,937	112,013	111,229
Sub Total-adults	256,440	262,261	210,718	215,312	216,547
Total (all ages)					
Male	297,564	303,960	250,753	237,408	238,317
Female.....	287,799	289,921	244,110	251,007	233,407
Total	585,363	593,881	494,863	488,415	471,724

Source: Refugee Affairs Secretariat

* Provisional

Note: The figures in this table include both refugees and asylum seekers

4) IDP EXAMPLE GEORGIA

417. Part II of this Manual provides an example of IDP tabulations derived from the General Population Census of 2014 in Georgia. The data can be accessed in the Census-related publications: <http://census.ge/en/publication>

418. The tables produced estimate;

1. IDP's by age and sex (by Age, Urban/Rural and Sex); and
2. IDP's by place of usual residence and place of residence before acquiring the IDP status (by Place of usual residence, Sex and place of residence before acquiring the IDP status).

5) IDP EXAMPLE COLOMBIA

419. Official statistics on displacement due to internal armed conflict in Colombia are published in Spanish in the website of the National Information Network of the Victims' Unit (<https://www.unidadvictimas.gov.co/es/registro-unico-de-victimas-ruv/37394>). The website is open to the public and does not require any credentials.

420. The statistics related to the Effective Enjoyment of Rights are shared currently to the relevant government agencies in charge of social assistance in order to target access to aid and social programs. In this case, the dataset of the vulnerability assessment is shared via web through a secure site.

421. The following reports, regarding displacement are available at the website:

1. Number of IDPs disaggregated per time of expulsion, time of reception, time of declaration, with information presented yearly and accumulated for the period before 1985.
2. Number of IDPs disaggregated per differential approach into the variables of ethnicity, sex, age and disability.
3. Open data with numbers of IDPs mixing the above variables.

b. Basic indicators (statistics, proportions and estimates)

422. Once the various stocks have been defined in the Design phase and the data collected, cleaned and input into databases, the variables and categories can be developed in the database ready for analysis. This requires a clear definition of the indicators and metrics to be used in the analysis

423. **The relevant Refugee Indicator** tables in Phase 2 are Table 2.3 Refugee Stock Indicators, Table 2.4 Refugee Flow Indicators both found at paragraph 633 of this Manual.

424. **The relevant IDP Indicator** tables in Phase 2 is *Table 2.8: Key stock indicators of IDP, IDP-related, and those who have overcome key displacement related vulnerabilities*. These indicators include 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.

c. SDG indicators recommended for refugees and IDPs

425. In the years leading up to 2030, the SDGs will play an important role in policy discussions. During the 2020 Comprehensive Review of the SDG indicator framework, a

specific indicator on refugees was included in the framework, but there is no specification for other forcibly displaced populations, including IDPs. However, the issue cuts across many SDGs, and multiple indicators can be disaggregated by migratory status and, more specifically, by displacement status. When SDG indicators are collected as part of statistics on forced displacement it is recommended that countries use the SDG framework and metadata in developing the statistics.

Table 6.2 Recommended SDG Indicators to be disaggregated by forced displacement by priority policy areas (IRIS Table 4.1)

Policy Area 1: Basic needs and living conditions	
2.2.1	Prevalence of stunting (height for age < - 2 standard deviation from the median of the World Health Organization (WHO) Child Growth Standards) among children under 5 years of age
3.1.2	Proportion of births attended by skilled health personnel
6.1.1	Proportion of population using safely managed drinking water services
11.1.1	Proportion of urban population living in slums, informal settlements or inadequate housing
Policy Area 2: Livelihoods and economic self-reliance	
1.2.1	Proportion of population living below the national poverty line, by sex and age
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
7.1.1	Proportion of population with access to electricity
8.3.1	Proportion of informal employment in non-agriculture employment, by sex
8.5.2	Unemployment rate, by sex, age and persons with disabilities
Policy Area 3: Civil, political and legal rights	
1.4.2	Proportion of total adult population with secure tenure rights to land, (a) with legally recognized documentation, and (b) who perceive their rights to land as secure, by sex and type of tenure
16.1.4	Proportion of population that feel safe walking alone around the area they live
16.9.1	Proportion of children under 5 years of age whose births have been registered with a civil authority, by age

426. The Inter-Agency and Expert Group on SDG Indicators accepted the proposal of EGRIS members to recommend disaggregation by forced displacement categories for 12 priority SDG indicators, as well as further broken down by age and sex whenever possible.⁴³ In order to ensure availability of minimum quality statistics on refugees and IDPs, it is recommended that national statistics providers include statistics on these priority indicators in the reporting plans for the 2030 Agenda. However, governments should not feel limited to this priority list, and data collection on additional indicators is encouraged.

⁴³ <https://unstats.un.org/sdgs/iaeg-sdgs/disaggregation/>

427. A second initiative, the Expert Group meeting on Improving Migration Data in the Context of the 2030 Agenda,⁴⁴ has identified another 12 indicators which they recommend be disaggregated by migratory status. Both selected lists of SDG indicators are a useful resource for identifying indicators to measure the characteristics of IDPs within the criteria set forth by the IASC Framework.

428. Matching the collected data with the defined metrics and indicators, stocks and flows, can be done in various ways, depending on the data collection method used. Stock data can be populated based on questions asked within surveys and census, that aim to directly determine if and how many people match the selected characteristics for each type of stock (e.g. a survey with a question such as “have you been forced to leave your place of residence in the last x months/years?”), or using data obtained from registers of IDPs or refugees (see Phase 2).

d. More complex analysis and commentary

429. This might include an analysis of the integration of displaced people into host communities, the progress made towards overcoming key displacement-related vulnerabilities, and other key policy issues. The body of literature available is extensive, and guidance can be obtained in the IRRS, 2018 and the IRIS, 2020.

IRRS RECOMMENDATIONS

430. From a legal perspective, durable solutions for refugees occur when refugees can once again enjoy the protection of a state, either through voluntary repatriation to their country of origin, integration in their host country, or resettlement to a third country. However, given the protracted nature of many conflicts, repatriation is not an immediate prospect for most refugees, and resettlement and naturalization are solutions for only a small minority. In reality, the majority of refugees aspire to, or enjoy, de facto integration in their host countries. Rather than a one-off event, the integration of refugees involves an incremental process. The IRRS focuses on the satisfaction of refugees' immediate and ongoing needs, their access to rights and on their integration. Living conditions can serve as a proxy for the satisfaction of these needs, and as data about peoples' living conditions have been routinely collected across the world for over 25 years, the results for the general population are already widely available. The satisfaction of basic needs, and the acquisition of rights is the first step in the integration of refugees.

431. Refugees have particular challenges compared to general migrants. Due to the forced nature of their migration and the traumatic experiences frequently associated with it, they often suffer from psychological distress and disabilities. Moreover, unlike labour migrants who tend to settle where their skills can be used more productively than in their home country, refugees may settle in areas with limited or no employment or other livelihood opportunities.

432. The IRRS presents recommendations for indicators that can be used to measure satisfaction of needs, access to rights and the integration of refugees with their host country or original country communities. The recommendations are grouped into three levels of indicators: priority, second level, and third level indicators. Table 5.1 of the IRRS presents the proposed indicators organized by the dimension of need and integration that it measures, specifically the following: (a) legal; (b) civil-political; (c) demographic and migration; (d)

⁴⁴ For more information, see <https://unstats.un.org/unsd/demographic-social/meetings/2017/new-york-egm-migration-data>

education; (e) economic; (f) social inclusion; and (g) health. In addition, several core classificatory variables are also recommended for the production of comparative analyses.

433. These classificatory variables are cross-referenced to those tabulations recommended earlier in Phase 2.1. These include country of birth, age, sex, reason for migrating, legal status, citizenship and length of time in the host country. Length of residence in the host country is especially important for statistics on integration because integration is a temporal process that begins when the refugee arrives in the host country. The pace of integration is usually different for refugee and refugee related groups compared with general migrants.

434. The characteristics of age and sex are used to disaggregate indicators of refugee living conditions and integration. For instance, using the sex variable to look at the difference in employment status between male and female refugees can highlight gender differences for that indicator. Classificatory variables are described in detail in *Table 2.9: Variables for refugee statistics*. As far as possible the proposed indicators correspond with those of the SDGs. Where they correspond, the equivalent SDG indicators have been noted in IRRS Table 5.1, the correspondence is often not direct. Guidance is given on the composition of the proposed indicators. Additionally, metadata for the SDG indicators are published by the United Nations Statistical Division on their website.¹ The last column of Table 5.1 gives the closest SDG indicator equivalence, it is not proposed that all these indicators are used.

435. The Indicators of integration and the satisfaction of immediate and ongoing needs of refugees and refugee related populations are;

1. Legal indicators
2. Civil-political indicators
3. Demographic and migration indicators
4. Education indicators
5. Economic indicators
6. Social inclusion indicators
7. Health indicators

IRIS RECOMMENDATIONS

436. In 2015, an interagency process was established to operationalise the Inter-Agency Standing Committee (IASC) Framework. Under the leadership of the Mandate of the UN Special Rapporteur on the Human Rights of IDPs, a group of development, humanitarian and peacebuilding actors started work on developing and testing indicators and guidance for comprehensive durable solutions analysis in internal displacement situations and to measure progress over time. The work, coordinated by the Joint IDP Profiling Service, resulted in a library of standardised indicators and guidance for the operationalization of the eight IASC⁴⁵

⁴⁵ Inter-Agency Standing Committee "Framework on Durable Solutions for internally displaced persons", April 2010. Available at: <https://goo.gl/BydBcF>

criteria.

437. For a statistical approach for measuring progress towards durable solutions and whether key-displacement related vulnerabilities have been overcome. Two measures are proposed:

1. The durable solutions progress measure designed 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. Therefore, this measure will be able to show whether progress is being made in overcoming the different vulnerabilities, and cumulatively, the progress made towards the achievement of durable solutions.
2. The composite measure designed to specify whether all key displacement-related vulnerabilities have been overcome and thus, whether or not persons with displacement-related protection needs and vulnerabilities can be taken out of the total IDP stock.

438. For the purpose of statistically measuring durable solutions progress, displacement-related vulnerabilities, sub-criteria were specified for each of the eight IASC criteria as is shown in Table 6.3 Criteria and sub-criteria included in the composite measure (IRIS, Table 4.3) below. The first four criteria of the IASC Framework are relevant in all contexts and should therefore always be included. The last four should be considered and can be included if deemed relevant for the specific displacement context. Therefore, all eight criteria are included in recommendations to measure progress towards durable solutions (see IRIS, paragraph 146). Regarding the development of the composite measure for overcoming key displacement-related vulnerabilities, a smaller set of criteria are chosen, including the four deemed to always be relevant in displacement contexts. Given the prevalence of challenges linked to lack of documentation in many displacement settings, it is also recommended to include the fifth criteria on access to personal and other documentation in any statistical measure that can result in taking IDPs out of the stock. This measure therefore prioritises five of the eight criteria (see IRIS, paragraph 164).

439. For each sub-criterion, there are many different indicators which can be chosen, although a final list has not yet been agreed. 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 Durable Solutions Library⁴⁶). Where relevant, they should also mirror the SDG indicators recommended to be disaggregated by different categories of displacement. 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.

440. A preliminary proposal of indicators for the sub-criteria has been developed by the EGRIS' IDP subgroup (see IRIS for discussion of indicator selection), taking into consideration the above listed aspects. The aim of this work was to provide a short, manageable list of one or two indicators per sub-criteria. However, given the limited availability of data, particularly for the sub-criteria which experts determined to be displacement-specific, it is too preliminary to recommend a final list of indicators for each sub-criterion at this stage. Further testing is

⁴⁶ <https://inform-durablesolutions-idp.org>

required to detail out this proposal. As this work progresses, information will be added to future editions of the Compilers' Manual, including implications for target setting based on different types of indicators (see IRIS, paragraph 166). Once finalized an updated version of these Recommendations could be considered. In the meantime, countries and international organisations should use the durable solutions indicator library as a reference and starting point for indicator selection and should as far as possible follow principles described (see IRIS, paragraphs 135-137).

Table 6.3 Criteria and sub-criteria included in the composite measure (IRIS, Table 4.3)

Criteria		Sub-criteria
1. Safety and security	1	1.1 Victims of violence
	2	1.2 Freedom of movement
2. Adequate standard of living	3	2.1 Food security
	4	2.2 Shelter and housing
	5	2.3 Medical services
	6	2.4 Education
3. Access to livelihoods	7	3.1 Employment and livelihoods
	8	3.2 Economic security
4. Restoration of housing, land and property	9	4.1 Property restitution and compensation
5. Access to documentation	10	5.1 Documentation

441. The targets for each sub-criterion should be defined. A comparative approach against the general/national population is recommended. If determined relevant by the national context a sub-set thereof can be used for this purpose.

442. The two measures on progress and the composite measure, should be used for the production of official statistics on internal displacement.

1. The durable solutions progress measure designed 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
2. The composite measure designed to specify whether all key displacement-related vulnerabilities have been overcome and thus, whether or not persons with displacement-related protection needs and vulnerabilities can be taken out of the total IDP stock.

443. 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.

1) EXAMPLE COLOMBIA

444. The example in Part II paragraph 574 of this manual onwards describes the *Use of administrative records to assess displacement related vulnerabilities in Colombia*. The Socio Economic Re-establishment Index is used to determine the extent of their vulnerability, and are assessed as no longer having vulnerabilities once the SRI is equal to one.

Box 6.2: Colombia Social Re-establishment Index

Example 6.1: Overcoming the situation of vulnerability linked to displacement in Colombia

In 2008 under the mandate of the Constitutional Supreme Court, the Government of Colombia developed the Effective Enjoyment of Rights indicators. The main purpose of this set of indicators is to evaluate progress towards durable solutions and the extent to which an IDPS has overcome displacement related vulnerabilities.

The minimum criteria to measure the overcoming of the situation of vulnerability is published in Presidential Decree 2569 of 2014. The variables included in the assessment are; identification according to the age, access to food, access to health services, access to education for the underage, income generation for the household, quality of housing and family reunification.

The set of criteria used to measure the overcoming of the situation of vulnerability represents the extent to which an IDP has achieved socio-economic stabilization and the thresholds are compiled in the Socio Economic Restoration Global Index that is calculated for every IDP since 2016. Once a victim overcomes displacement related vulnerabilities, they are no longer targeted for assistance related to displacement and removed from the stock estimates.

2) EXAMPLE SUDAN DARFUR: MEASURING PROGRESS TOWARDS DURABLE SOLUTIONS

445. The Government of Sudan and the international community are working together to jointly support durable solutions for Darfur's internally displaced people. The commitment is rooted in the Doha Document for Peace in Darfur (DDPD) from 2011, signed by the Government of Sudan, other major parties to the conflict and the international community, and which sets out the framework for peace in Darfur.

446. A durable solutions process was piloted in El Fasher in North Darfur. To reach durable solutions and end displacement, long-term planning based on agreed and jointly-owned comprehensive data is needed. To establish an agreed evidence base, a collaborative profiling approach was adopted that brought actors together to ensure trust and ownership of the results of the profiling exercise. An important aspect of this durable solutions profiling is that it places IDPs centre-stage alongside the two other major stakeholders—the Government of Sudan and the international community—permitting the profiling results and recommendations to be owned and signed off by all parties.

447. The durable solutions profiling exercise in El Fasher specifically aims to:

1. Provide a comprehensive profile of IDPs residing in Abu Shouk and El Salaam IDP camps.
2. Develop a better understanding of IDPs' vulnerabilities, coping mechanisms, capacities and provide insight into IDPs' perceptions and their future settlement intentions.

3. Provide a jointly agreed upon data set to help inform durable solutions programming responses.
4. Pilot a profiling exercise of displacement and joint durable solutions planning that could be replicated in other Sudan contexts with displaced populations.

448. The profiling exercise adopted the Inter-Agency Standing Committee (IASC) Framework on Durable Solutions for Internally Displaced Persons as an analytical framework. The profiling was designed to take into account the displaced persons' perspectives on durable solutions including preferred settlement options and places a strong emphasis on understanding the rationale behind these preferences. It is also concerned with understanding the progress IDP communities have made across the IASC durable solutions criteria and applied an area-based comparative approach.

449. This comparative analysis was used to understand whether hardships are a result of IDPs' displacement or development challenges shared with non-displaced communities in El Fasher. Finally, it included a macrolevel urban analysis. The profiling exercise used mixed methods: mapping and enumeration, a household survey, and key informant interviews. The profiling applied a gender lens and questions for further analysis, which were included in the report narrative when significant gender related differences were identified. The report gives details of the indicators used as part of the profiling exercise. Link <https://www.jips.org/jips-news/selecting-core-durable-solutions-indicators-sudan/>

3) EXAMPLE NORWAY REFUGEE LABOUR MARKET ANALYSIS

450. Statistics Norway publishes a report every year describing the labour market situation for the refugee workforce population (15-66 years old) settled in Norway under its Refugee and the Labour Market technical series. The report is based on the register-based employment statistics of Norway.

451. It is important to understand the necessity of producing such statistical analysis from Statistics Norway's perspective. The participation of the refugees in the labour market is an imperative to understand the underlying driving forces for the refugee integration in the host countries' labour market. The result of the analysis is pivotal in designing effective policy responses to integrate the refugees, asylum seekers in the host communities in Norway. The outcome of the analysis will aid the Norwegian policy makers as well as the programme planners with a solid baseline / starting point to assess the impact of an intended policy or programme and at the same time will pave the way for further improvement in the policy.

Box 6.3: Analysis of refugees in the labour market, Statistics Norway

Analysis method used: The report presents the analysis outputs, consisting of statistics of various vital labour market indicators, e.g.:

- The number of employed refugees that were registered officially in Norway,
- Their employment rate,
- Rate of increase in employment among refugees.

The analysis is conducted in comparison with the total population and additionally in comparison with different variations of the overall refugee population (country of origin, age group, gender, time of arrival in Norway and so on) using cross tabulation methods and statistical inference.

Furthermore, the analysis shows the impact on integration in the local labour markets of different control variables such as, education level, stock of the existing refugees of the same nationalities and so on, with different varying factors across the refugee and the nonrefugee population using a tabulation method.

4) NEPAL EXAMPLE – MOBILE PHONE TRACKING OF POPULATION DISPLACEMENT

452. Using alternative data sources from mobile phone records Flowminder and Ncell collaborated with UN OCHA and other disaster relief agencies to estimate the number of people displaced. See Part II of this Manual *Nepal*.

5) SOUTH SUDAN EXAMPLE: POVERTY ASSESSMENT

453. South Sudan's National Bureau of Statistics and the World Bank conducted the fourth wave of the High Frequency Survey and the Crisis Recovery Survey to monitor welfare and perceptions of citizens in accessible urban areas and IDP camps across South Sudan. The HFS and CRS data contains information on security, economic conditions, education, employment, access to services, and perceptions. The data combines detailed household questionnaire information with displacement-specific information including drivers of displacement, access to resettlement mechanisms, and return intentions. It also includes comprehensive information on assets and consumption, to allow estimation of poverty based on the Rapid Consumption methodology as detailed in Pape and Mistiaen (2015).

454. More information can be found in Part II of this Manual *South Sudan Example: Poverty Assessment* and the 2017 South Sudan Poverty Assessment Report can be found at <https://microdata.worldbank.org/index.php/catalog/3392/related-materials>.

6) YEMEN DISPLACEMENT TRACER MATRIX EXAMPLE

455. The statistical series was started as a result of the dramatic increase of displacement in Yemen due to the outbreak of conflict in March 2015. In order to better inform the humanitarian community about the location and needs of the displaced populations. The data was collected using the IOM's Displacement Tracer Matrix (<https://displacement.iom.int/content/methodological-framework-used-displacement-tracking-matrix-operations-quantifying>).

456. The data collection will seek to identify the local demographic profile, displacement dynamics, and to answer the following questions;

1. Who and how many people face moderate and severe humanitarian needs?
2. Where are these people located?
3. What do their survival and livelihood problems consist of?
4. How are the needs expected to evolve in the future, based on ongoing/planned responses and other likely events?

457. For reports see [Yemen Area Assessment Round 37 March 2019](#) and <https://displacement.iom.int/Yemen>.

458. For more information see Part II of this Manual *Yemen IDP Example*.

References and Further Examples

459. For more information on the Durable Solutions library, please see: <https://www.jips.org/tools-and-guidance/durable-solutions-indicators-guide/>. For results of the work see: <http://inform-durablesolutions-idp.org/>

460. European Commission, '**Using EU Indicators of Immigrant Integration, Final Report for Directorate-General for Home Affairs**'. This report on the further development and use of EU immigrant integration indicators in policy debates is prepared at the request of the European Commission by the European Services Network (ESN) and the Migration Policy Group (MPG). Link https://ec.europa.eu/home-affairs/sites/homeaffairs/files/e-library/documents/policies/legal-migration/general/docs/final_report_on_using_eu_indicators_of_immigrant_integration_june_2013_en.pdf.

6.2. Validate outputs

461. This sub-process is where statisticians validate the quality of the outputs produced, in accordance with a general quality framework (developed in Phase 2) and with expectations based on local knowledge. This sub-process also includes activities involved with the gathering of intelligence, with the cumulative effect of building up a body of knowledge about a specific statistical domain. This knowledge is then applied to the current collection, in the current environment, to identify any divergence from expectations and to allow informed analyses. Validation activities can include checking the population coverage in the sample group and looking for content errors.

462. An example is Population Census coverage that can involve two or three major problems such as under representation, over representation, and selection bias in the study design. The most common problem is under-representation of the population in the sample group or due to selection bias in the research design. Generally, an aggregate undercount of 3 to 6% is considered acceptable, but some population groups tend to be disproportionately affected, e.g. very young children, young mobile men and undocumented migrants. Over-coverage also happens, especially if local authorities, with an interest in exaggerating their

populations, are able to influence the data collection.

463. Content errors might involve non-response bias and response bias for some of the key variables in the data collection phase. Individual data are often provided by an informant who is either the head of household or another responsible adult household member. These proxy respondents may not know all the details regarding each member of the households, or sometimes may not want to share the facts. The household members themselves may be unsure or misunderstand the question. They may even hide or distort information on purpose (see also under-reporting of consumption⁴⁷). As a result, data quality may be poor, despite excellent coverage. Therefore, it is necessary to detect those errors in the analysis phase and select the right analysis techniques to eradicate or minimise the content error.

464. There are several methods for quality assessment. Content errors can often be detected through internal consistency analysis by crosschecking the data across different datasets. Furthermore, the analyst can resort to the following mechanisms to validate the findings:

- Comparing the statistics with previous cycles (if applicable and available),
- Checking that the associated metadata and paradata (process metadata) are present and in line with the expectations,
- Checking the statistics for consistency against other relevant data (both internal and external), comparing with other national sources, operational and global data,
- Investigating inconsistencies in the statistics,
- Agreeing with the accepted threshold level in applying the quality standards and validation processes
- Performing macro editing, and
- Validating the statistics against expectations and domain intelligence.

465. The main criteria are: is the data understandable, usable for analysis and fit for purpose, particularly the purpose that was articulated in the Design phase. Does it respond to the questions that were set initially?

466. Comparators are important, either from another dataset relating to a similar group or population, or to the same group/population across different time periods. Particularly important is coherence when comparing datasets are those obtained via different methods (e.g. surveys versus registration). Although the figures obtained via different methods will not always match, they need to show clear relative trends, to allow comparability. The availability of good metadata for both datasets will enable judgments to be made about the reasons for differences and for a judgement to be made about the quality of the data.

467. The *analytical framework* should include *commentary and notes*, to allow proper

⁴⁷ https://link.springer.com/chapter/10.1007/978-3-030-25120-8_11#Fn3. This chapter is a summary of Kaplan, Pape, and Walsh (2018, forthcoming), "Eliciting Accurate Responses to Consumption Questions Among IDPs in South Sudan Using "Honesty Primes", Policy Research Working Paper Series. See also open source ebook The World Bank and Johannes Hoogeveen, Utz Pape Editors. *Data Collection in Fragile States Innovations from Africa and Beyond* <https://link.springer.com/content/pdf/10.1007%2F978-3-030-25120-8.pdf> and Johannes Hoogeveen, Emil Tesliuc, Renos Vakis with Stefan Dercon, *A Guide to the Analysis of Risk, Vulnerability and Vulnerable Groups*; <http://siteresources.worldbank.org/INTSRM/Publications/20316319/RVA.pdf>

understanding of how the data was obtained and treated, in order to support the accessibility and credibility of the results. This commentary should be accompanied by clear explanatory metadata pertaining to how the data were collected, processed and analysed (see Phase 7 on metadata)

468. Complement the estimates with *contextual updates*, to inform the reader and to allow for better understanding of trends. If numbers go up, does this correspond to greater coverage, or to an increase in violence, attacks or incidence of disasters. Try to explain what the likely correlates are.

469. *Triangulation*: use operational data for comparison, particularly for flow data, as operational data is more likely to capture micro-movements, spontaneous movements, short-term displacements. It is necessary to make sure the trends match between the various datasets without any major deviation. If the variation between two stocks is not significant, but operational 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 allow official statistics to verify/correct their flow data.

470. Linking to cross border data, *interoperability*: 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. Refugee data could be compared to third country nationals, refugees, migrants in other countries, to validate estimates and trends. Example: the stocks of IDPs in South Sudan only increase marginally between two data collection periods; however, the number of South Sudanese nationals in the Central African Republic increase exponentially, then the likelihood is that people were displaced over the border. Operational flow data from displacement event should reflect that significant variation, even if the stocks do not.

471. 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 same 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.

Examples and References

1) EXAMPLE OF USING SEVERAL DATA SOURCES TO VALIDATE DATA: THE WELFARE OF SYRIAN REFUGEES:

472. This joint publication of UNHCR and World Bank on the welfare of Syrian Refugees (2016) aimed to increase the understanding and ultimately improve the well-being of Syrian refugees. Due to the overwhelming humanitarian response, there was an abundance of data on the vulnerability, well-being and access to basic services aspects of the Syrian refugees. These data had not yet been fully explored and utilised for analytical, policy, and planning purposes at that time. The report used different sources of data to understand the vulnerability and poverty conditions among the Syrian refugees.

473. <https://www.worldbank.org/en/topic/poverty/publication/the-welfare-of-syrian-refugees-evidence-from-jordan-and-lebanon>

Box 6.4: Welfare of Syrian refugees using multiple datasets to resolve inconsistencies in consumption data

Example 6.2: Joint publication of UNHCR and World Bank on the welfare of Syrian Refugees

Validation of the data: The report relies on the data extracted from the global registry system of refugees maintained by the UNHCR (ProGres) and on four other data sets collected by the UNHCR and the World Food Programme (WFP) in Jordan and Lebanon.

The Jordan data includes the second and third rounds of the home visits data (JD-HV2 and JD-HV3), which are home visits undertaken by the UNHCR for the purpose of targeting.

The Lebanon data includes the second round of the vulnerability assessment (LB-VASyR) and the verification survey (LB-Verif), the latter is also an instrument designed for targeting purposes.

The welfare and poverty analysis of the report was replicated for three other data sets:

1. The third round of the Home Visits data in Jordan (JDHV3);
2. The second round of the Lebanon Vulnerability Assessment of Syrian Refugees (LB-VASyR);
3. The Lebanon Verification data set (LB-Verif), a data set compiled using information collected during the appeal process for the World Food Programme (WFP) food voucher distribution system.

Since multiple datasets from different sources were utilized for the analysis, the researchers addressed the first problem involving inconsistencies in the datasets, given that the available data was the construction of the main welfare aggregate to be used for the welfare model.

Questions on income and expenditure referred to a recall period of one month prior to the interview. These questions were not ideal to construct the welfare model, and under-reporting was expected. Under-reporting of expenditure is problematic challenging for statistics that use the population mean, e.g. mean expenditure or the poverty rate. To ameliorate this content error, the researchers performed statistical testing under mean changes by using, for example, different poverty lines or stochastic dominance analysis. By performing these sensitivity tests, they overcome the problem of exact estimates of the welfare aggregate and also the normative problem of selecting a particular poverty line for the analysis.

6.3 Interpret and explain outputs

474. This sub-process calls upon the in-depth expertise of statisticians to understand, interpret and explain the statistics produced by assessing how well the findings reflect their initial expectations. This involves viewing the statistics from all perspectives using different tools and media, and also carrying out in-depth statistical analyses.

475. The analyst should check to see if the data respond to the question that were initially set and do the outputs match what was determined in the Design phase? A basic analysis might

ask the following questions:

1. What are the variations from one reporting period to another?
2. Are those variations attributed to contextual or methodological considerations or both? (i.e. more to do with displacement trends or to do with the data collection and treatment methods)
3. Are the variations consistent across the entire context in scope of the analysis? Are there different trends within that scope? How can these be explained?
4. How are stocks and flows affected by the changes? Has one indicator changed more than others, i.e. have flows increased/decreased significantly whereas stocks have changed more marginally?
5. Which flows have changed the most, and what does this tell us? E.g. has the number of IDPs increased/decreased much more in comparison with cross-border flight and refugee numbers abroad?
6. Have methodologies changed? If so, why? How does this impact the accuracy and comparability of the analysis?
7. Do the variations in numbers correspond to contextual changes that can explain them?
8. Which groups, as highlighted by the sex and age disaggregated data (SADD) and other forms of disaggregation, are the most impacted? Why? Has that changed since the last reporting period?
9. Do the official statistics vary significantly in relation to other datasets produced by other sources, particularly operational actors? If so, why?
10. Do the variations correlate with specific actions taken to prevent or solve situations of displacement/provision of assistance/interventions/any form of state operation?

2) EXAMPLE OF INTERPRETATION OF OUTPUTS USING MULTIPLE DATASETS:
HUMANITARIAN SHELTER

476. “The State of Humanitarian Shelter and Settlements” report by Global Shelter Cluster (GSC) report provided an in-depth review of the evidence base of the shelter and settlements sector, and the sector’s data collection practices. One of the main difficulties encountered when attempting to measure shelter adequacy is obtaining accurate data across various contexts. This is due partly to the multi-sectoral nature of the criteria, and the wide range of environments in which humanitarian agencies work, as well as to variations between shelter designs, materials and costs. In addition, different adequacy indicators may sometimes conflict with each other. Most shelter actors therefore agree that a one-size-fits-all definition of adequacy is almost impossible.

477. To compare and analyse the data, GSC created a sub-set of information in order to gather data for disasters, conflicts and crises between 2013 and 2018 for which a Shelter Cluster (SC) response or SC-like response had been activated. The methodology is explained in detail

in Part 3 of this report https://www.sheltercluster.org/sites/default/files/The_State_of_Humanitarian_Shelter_and_Settlements_2018.pdf

Box 6.5: Example of interpreting outputs from multiple data sources in 'The State of the Humanitarian Shelters Settlements 2018' report

Methods of Interpreting the Outputs used in the "The State of the Humanitarian Shelters Settlements 2018"

Context: "The State of Humanitarian Shelter and Settlements" report by Global Shelter Cluster (GSC) aims to deliver a better understanding of the humanitarian shelter and settlements sector by providing a wide-ranging picture of humanitarian shelter needs and offering an analysis of current trends with the supporting statistical evidence base.

Interpretation structure used by Global Shelter Cluster: The third part of the report provided an in-depth review of the evidence base of the shelter and settlements sector, and the sector's data collection practices. To compare and analyse the data, GSC created a sub-set of information in order to gather data for disasters, conflicts and crises between 2013 and 2018 for which a Shelter Cluster (SC) response or SC-like response had been activated.

The report mainly provided an understanding of the evidence currently supporting shelter analysis across 153 attributes such as funding levels and the counts of damaged households, which they sourced, mapped out and combined the datasets from SC operations, UNHCR Displacement Data, ReliefWeb, Financial Tracking Service, DesInventer etc.

The analysis and interpretation of the findings part is principally structured in four main domains: disaster impact and shelter needs, shelter response, shelter gap and shelter impact, which they effectively analysed using different datasets.

Across different domains, the report provided thorough comparative analysis of the findings based on the activation of Shelter Cluster. Few parts of the analysis is done in terms of correlations comparing individual indicators between population groups and different geographical regions with a special focus on crises-prone regions.

6.4. Apply disclosure control

478. At the Design phase, the outputs should be designed with a clear understanding of the required information needed from those outputs to respond to the agreed questions. This should dictate how much information is necessary to collect, and what falls outside of that scope.

479. Due consideration should be given to privacy in accordance with national and international law, and to any type of information that could lead to the respondents being identified or suffering from discrimination as a result of their participation or responses in a survey/census should not be disclosed.

480. 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. This can enable possibilities for combining datasets to improve coverage and quality (see section on *Data Source D: Integrated Data*). This can also help prevent duplication of data. A data protection policy should be in place from the start, and Phase 2D discusses the a. Legal

environment and data protection legislation and the other capacity requirements for this.

481. Data integration may be possible where the data being collected by various government databases relates to individuals, groups or populations who are registered in one or more government databases. Inter-agency data sharing protocols can enable this and also help to prevent duplication. As the various service needs of the concerned populations are likely to be addressed by different entities within the government (or even by entities outside of it), data sharing between those entities will ensure greater capacity for analysis.

482. There are technical computer languages that allow data sharing more effectively, such as tagging, e.g. using HXL, allowing the data outputs to be identifiable by various data sharing platforms, and for the various elements of the datasets to be identified and shared individually as well.

483. For an example of how anonymisation is done by different governments and the decision rules that are applied to enforce it. See Canadian Example of a secure data enclaves and data use offering higher levels of security, IRRS, Box 6.4.

6.5. Finalise outputs

484. Final outputs and datasets must come packaged with all the fields or variables fully explained, with good metadata. The metadata should explain anything that might have affected the accuracy or quality of the data. Data treatments, weighting and any other transformations such as derived variables should be fully documented and made available to users.

485. It is crucial to account for any caveats/limitations that may affect the data and therefore prevent analysis, as well as comparability. An example of a caveat would be to specify the geographical coverage of the datasets. This should include coverage problems which have resulted from difficulty of access in particular locations. All caveats must be detailed: specific to each collection method, highlighting any potential difference in the collection exercises at different times, obstacles met preventing the data to be collected according to the parameters set in the Design phase, and include calculations of the sampling errors which apply to key estimates.

486. The outputs should also specify what quality assurance and quality control mechanisms were applied, and the findings of the quality assessments should be made available to users.

ⁱ See: <https://unstats.un.org/sdgs/iaeg-sdgs/metadata-compilation/>

Phase 7. DISSEMINATION

“This phase manages the release of the statistical products to customers. It includes all activities associated with assembling and releasing a range of static and dynamic products via a range of channels. These activities support customers to access and use the outputs released by the statistical organisation⁴⁸.” (UNECE, 2013)

487. Dissemination is a phase in the statistical process, in which the data collected, and outputs compiled by national statistical offices (NSOs) and other relevant national authorities are released to the public. It features as a step in the GSBPM

488. The responsibility for disseminating refugee and IDPs statistics varies from country to country and the dissemination practices differ significantly in their methods and effectiveness. In many countries, immigration authorities or other government ministries, departments or agencies (e.g. Ministry of Interiors) provide the NSO and UNHCR with the relevant data for compilation, validation and analysis. After validation, the NSO and/or relevant authorities disseminate the statistics, sometimes in parallel. In other countries, refugee data are collected directly by UNHCR, who also disseminate them.⁴⁹ A more harmonized approach is needed to the dissemination of refugee and IDP statistics.

489. The phase is composed of five sub-processes, that are briefly described below. Then on the subsequent pages, each section is discussed at length. The sub-process sections include general advice and specific examples of refugee and IDPs applications. The 5 sub-processes are:

7.1 Update of output systems⁵⁰

490. This sub-process manages the update of an existing system where data and metadata are stored ready for dissemination purposes, including: a) formatting data and metadata ready to be put into output systems; b) loading data and metadata into output systems; c) ensuring data are linked to the relevant metadata.

7.2 Production of dissemination products

491. This sub-process describes the various forms available for dissemination products, including: a) online data (e.g. databases with predefined datasets presenting aggregated data), b) analytical publications/reports, c) user-friendly leaflets, booklets, flyers, posters; d) digital publications; and e) public-use micro-data sets and downloadable files. This section of the Manual also provides examples of publication guidelines to facilitate the production process.

7.3 Management of the release of products

492. As highlighted in the IRRS and IRIS documents, dissemination "should be guided by a

⁴⁸ Ibid para 94

⁴⁹ Source IRRS: <https://ec.europa.eu/eurostat/documents/3859598/9315869/KS-GQ-18-004-EN-N.pdf/d331c9cc-1091-43c2-b589-2c250bccc281> - page 125.

⁵⁰ Output system is considered as a dissemination system (data portal, platform, database etc.) which is offered to external/end users of the statistical information/products.

number of principles including those established in the *Fundamental Principles of Official Statistics*⁵¹, such as objectivity, impartiality, confidentiality, privacy, timeliness and accessibility".

493. This section discusses the characteristics of an effective data dissemination system. It aims to ensure that all the elements for the statistical release are in place, including managing the timing of the release, briefings for specific groups such as the press, as well as the arrangements for any pre-release embargoes. It also covers the provision of products to subscribers and managing access to confidential data by authorised user groups, such as researchers and the error correction and notification policy.

7.4 Promotion of dissemination products

494. This sub-process includes the activities that are naturally associated with the data dissemination such as data publishing, and also activities related to the promotion of the dissemination products. This active promotion of the statistical products helps to reach the widest possible audience. It includes the use of different communication tools to facilitate the sharing of statistical information to users.

7.5 Management of user support

495. This sub-process explains how the dissemination of statistical products requires proper management of the user support. User queries and requests for other services such as microdata access should be recorded, and responses to these requests should be provided within agreed deadlines. These queries and requests should be regularly reviewed to provide an input to the overarching quality management process, as they can indicate new or changing user needs.

7.1. Update output systems

496. As formatting, loading and linking of metadata should take place in earlier phases, this sub-process focuses on final checks that all of the necessary elements are in place ready for dissemination. Therefore, this upload of the data and metadata should pay attention to the following principles:

1. **Manual or automated update process:** in order to streamline the update, it is advisable to automate this process. The update process should be defined in the design phase and executed as well as governed in this phase.
2. **Checking Technical requirements of the update:** Technical requirements for data exchange between systems like application programming interface (API), data exchange format (like SDMX), authentication/access rights and control, etc. should be defined in the design phase, and implemented and tested in the Build phase. Checking takes place in this sub-process to avoid potential errors, and prevent delays in the update of the output system.
3. **Monitoring of the upload process and verification of its correctness:** If the update is done automatically, it is important to control and verify the results of this

⁵¹ <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>

process (e.g. check for errors in the data, metadata, proper layout, descriptions). Therefore, it is necessary to assign this responsibility to a person who works with the selected output system.

4. **Detection and diagnostics of potential loading issues (consistency of data, no loading errors etc.):** As the electronic transfer might face interruptions, it is important to control this process and also to perform manual checks. As above, this checking should be assigned to a responsible person.
5. **Ensure Security:** In order to ensure the secure transfer of data between different systems, security rules should be established. The application of the rules can come from the output system 'owner' or sender to the recipient output system/database. In all cases, the minimum security settings should be defined, e.g. transfer of the data from the source to the destination system, output database should be encrypted and protected, and the security "health checks" should be part of the monitoring process.

497. As highlighted in the IRRS document, it is important to use a single portal/platform and single access point for disseminating all national refugee and IDP statistics. This approach can minimise confusion for users who consult various non-integrated portals often with conflicting results.

498. The decision on whether to use an existing dissemination hub or to develop a new one should be taken by national/regional or international statistical agencies while developing dissemination tools. However, it is recommended that statistical agencies take expert advice and access available information before choosing the system that best meets their dissemination needs.

7.1.1 Existing output systems (portals and platforms)

1) EXAMPLE 7.1 EUROSTAT CENSUS HUB:

499. The Census Hub is an easy-to-use tool that allows users to access census data. It can quickly produce customised tabulations; users specify the tables that they need, and the tool extracts the necessary data from the databases held by the NSOs. Data from the national databases are then compiled by the Census Hub, with output either displayed on screen or in spreadsheet-readable files.

Link: <https://ec.europa.eu/eurostat/web/population-and-housing-census-data/2011-census>

2) EXAMPLE 7.2 CROS PORTAL (PORTAL ON COLLABORATION IN RESEARCH AND METHODOLOGY FOR OFFICIAL STATISTICS):

500. The CROS Portal is dedicated to the collaboration between researchers and Official Statisticians in Europe and beyond. It provides a working space and tools for dissemination and information exchange for statistical projects and methodological topics. The services provided include hosting of statistical communities, repositories of useful documents, research results, project deliverables, and discussion fora.

Link: <https://ec.europa.eu/eurostat/cros/>

3) EXAMPLE 7.3 AFRICAN DEVELOPMENT BANK'S OPEN DATA PLATFORM:

501. As part of its Africa Information Highway (AIH), the African Development Bank (AfDB) provides an Open Data Platform (ODP) to all African countries. The key role of these portals is to facilitate statistical data access across Africa.

502. The ODP integrates many functionalities that allows countries to:

- Upload and update databases on the system,
- Create dynamic dashboards based on uploaded data,
- Share data with partners in different format (CSV, SDMX, ODATA, etc.).

503. The portal also gives countries the opportunity to integrate data from NSOs and line ministries within the NSS. In the context of refugees and IDPs, the ODP could play a key role in facilitating access to this data in African countries already using the platform for their dissemination. This is already done for SDG data and macro-economic data, the refugees and IDP data could be easily integrated to the African countries' ODP.

504. Recently, African countries have populated SDG hubs, having installed them as part of the ODP and also developed 'National Summary Data Pages'. In this respect, a module could easily be added to the platform for disseminating refugees and IDP data.

505. The links (below) show examples for Mauritius and Kenya, respectively.

<http://dataportal.opendataforafrica.org/>

<http://mauritius.opendataforafrica.org/konedhd>

<http://nso.kenya.opendataforafrica.org/dssevf/national-summary-data-page-nsdp>

4) OTHER EXAMPLES OF PORTALS/PLATFORMS FOR THE DISSEMINATION OF STATISTICS:

- i. **EU Open Data Portal** gives access to open data published by EU institutions and bodies (<https://data.europa.eu/euodp/en/home>),
- ii. **Knowledge Centre on Migration and Demography**, provides a single point of entry for information relevant to EU policies on migration and related fields (<https://bluehub.jrc.ec.europa.eu/portal/>)
- iii. **World Bank – DataBank** – an analysis and visualisation tool that contains collections of time series data on a variety of topics. Users can create their own queries; generate tables, charts, and maps; and easily save, embed, and share them (<https://databank.worldbank.org/data/home.aspx>)
- iv. **OECD data portal** gives access to charts, maps, tables and related publications (<https://data.oecd.org/>)
- v. **International Organization for Migration – Migration Data Portal**, aims to serve as an access point to information about migration data (<https://gmdac.iom.int/data-analysis-migration-data-portal>)

- vi. **Devinfo** tracks and disseminates key socio-economic development indicators (www.devinfo.org),
- vii. **NADA** makes it possible to archive technical documents, metadata and microdata from surveys and censuses (<http://www.ihsn.org/software/nada>)
- viii. **REDATAM**– facilitates the processing, analysis and dissemination of Statistics and Census (<https://redatam.org/redatam/en/index.html>)

5) EXAMPLES OF NATIONAL DATA PORTALS – IRELAND AND CANADA:

- i. Ireland: <https://data.gov.ie/>
- ii. Canada: <https://open.canada.ca/en/open-data>

7.1.2 *New Output system*

506. An alternative to using one of these existing systems is the creation of a new portal exclusively dedicated to 'Refugee and IDP statistics – Open Data Portal'. In which case a choice must be made as to which organisation will create and maintain such a portal and what content will be offered.

507. The following options could be considered:

1. Copy and centralise all content from various sources and present a full spectrum of 'Refugee and IDP statistics'. This option would require deep integration and data exchange between various stakeholders and therefore could be considered only as a long-term solution.
2. Create a centralised “dictionary” of all the 'Refugee and IDP statistics'. This solution could be considered as a “quick-win” and would aim to be a centralised place with links to all relevant materials and dissemination products (portals, various products like databases, publications, leaflets, social media, etc...) from all engaged national agencies, and international bodies. Such a portal can act as a single point of access for collected data. However, the portal's structure can be decentralised, which means the data remains on the data provider's website and the provider is responsible for its quality and availability. All the respective metadata are also stored on the source portals and are managed by data providers.

508. In both scenarios it is important to consider the importance of international comparisons and standards.

7.1.3 *Data Portals – guidelines*⁵²

509. There is a growing use of data portals among the NSOs (and regional and international agencies) for data dissemination. This is a positive development which allows for greater data availability and accessibility. It is recommended that statistical agencies take expert advice and access available information to support the selection of the system that meets their dissemination needs.

⁵² Source: PARIS21 NSDS guidelines on data dissemination: <https://nsdsguidelines.paris21.org/node/796>

510. Links to examples of relevant guidelines include:

- [The Handbook on Major Statistical Data Management Platforms by UNECA](https://nsdsguidelines.paris21.org/printpdf/node/798) is an important resource for NSOs and can provide guidance during the decision-making process to help NSOs select the appropriate platform for managing and disseminating statistical data to their users. <https://nsdsguidelines.paris21.org/printpdf/node/798>
- [World Bank's technical assessment of open data platforms for NSOs](http://documents.worldbank.org/curated/en/744241468334210686/pdf/928100WP0Tchn0Box385378B000PUBLIC0.pdf). This research report is intended to provide a better understanding and assessment of the technical issues related to data dissemination tools that NSOs use (or could use) to distribute data to the public under an open data initiative. The report defines a list of key criteria and evaluates relevant technology products according to those criteria. <http://documents.worldbank.org/curated/en/744241468334210686/pdf/928100WP0Tchn0Box385378B000PUBLIC0.pdf>
- [Making Data Portals work for SDGs: A view on deployment, design and technology](https://paris21.org/sites/default/files/Paper%20on%20Data%20Portals%20wcover%20WEB.pdf). This paper offers lessons on how the deployment, design and technology considerations can be improved as NSOs enter the implementation and monitoring phase of the SDGs – which will lead to a further push for data portals. The paper finds that while there have been many good intentions to make these portals available to countries, when it comes to actual implementation the outcomes are rather mixed. One particular problem is the set-up of multiple data portals with overlapping functionalities and their lack of integration – particularly in the most aid-dependent countries. This results in (i) a duplication of workload for already resource-constrained NSOs who have to maintain several portals and update information manually, (ii) confusion for users who consult the various portals with often conflicting results, and (iii) overall high costs for demonstrably low usage of these portals. [https://paris21.org/sites/default/files/Paper on Data Portals%20wcover WEB.pdf](https://paris21.org/sites/default/files/Paper%20on%20Data%20Portals%20wcover%20WEB.pdf)

7.2. Produce dissemination products

511. This sub-process focuses on the production of the dissemination products, which should be designed in earlier sub-processes (i.e. Phase 2.1. *Design Outputs*), however, it is presented here as a useful set of examples and best practices.

512. This sub-process focuses on three main aspects:

- i. **“what”** will be produced: selecting from the options for dissemination products, e.g. downloadable files, analytical reports, interactive graphics, maps, public-use microdata sets;
- ii. **“for whom”**, matching products to the types of users and their needs;
- iii. **“how”** it will be produced, following standards and guidelines in order to produce quality assured products

7.2.1. Dissemination products (“what”)

513. The following Dissemination products are among the most widely used:

- **Datasets** - multi-dimensional tables used to store the base data and more appropriate

for use by statistical and other experts via special applications. For dissemination purposes the data presented in the datasets should be aggregated to a level that is safe to share and accompanied by relevant metadata.

- **Predefined tables** - used to provide easy access to the main statistical indicators. They are generally based on datasets and are derived from them. They are predefined, non-modifiable and present a limited number of dimensions.
- **Statistics explained articles** - contain statistical data and explanations, but also direct links to all kinds of relevant background information which may be useful to understand the statistics.
- User-friendly **leaflets, booklets, flyers, posters**.
- **Statistical reports** - presenting a selection of data on specific topics, including descriptive statistics that provide a general picture of the findings.
- **Digital publications** – composed of short texts, dynamic infographics, maps, videos, graphs, photos, etc.
- **Visualisations**, including **maps and infographs** - visualisation tools, which are easily understandable, interactive and storytelling.
- **Public-use micro-data sets and downloadable files** – that consist of sets of records containing information on individuals, households or business entities. The files are created to allow the general public to become familiar with statistical microdata. The files are prepared in such a way that individual entities cannot be identified, but with a consequent loss in information value.

1) EXAMPLES OF DATABASES AND PREDEFINED TABLES:

- i. UNHCR: http://popstats.unhcr.org/en/time_series
- ii. Eurostat database – Asylum and managed migration: <https://ec.europa.eu/eurostat/web/asylum-and-managed-migration/data/database>
- iii. <https://ec.europa.eu/eurostat/web/asylum-and-managed-migration/data/main-tables>
- iv. Eurostat Metadata – Asylum and managed migration: https://ec.europa.eu/eurostat/cache/metadata/en/migr_asyapp_esms.htm

2) EXAMPLES OF STATISTICS EXPLAINED ARTICLES:

514. Statistics explained articles (Eurostat): https://ec.europa.eu/eurostat/statistics-explained/index.php/Main_Page

3) EXAMPLES OF STATISTICAL REPORTS:

515. The following reports are produced by the IoM using operational data.

- i. Overview reports that cover all the data collected during the specified timeframe.

These include descriptive statistics providing a general picture of the findings. There is no specific geographic subset of focus or analysis theme. The example is for **Turkey - Flow Monitoring Surveys Analysis (April 2019)**
<https://migration.iom.int/reports/turkey-%E2%80%94-flow-monitoring-surveys-analysis-april-2019?close=true>

- ii. **Thematic reports:** focus on a specific research question, analysis theme (e.g. health, education or livelihoods, etc.) or population subset. This example is for **West and Central Africa - Youth and Mobility: Children and Youth on the Move Across West and Central Africa in 2018 (April 2019)**.
<https://migration.iom.int/reports/west-and-central-africa-%E2%80%94-youth-and-mobility-children-and-youth-move-across-west-and-central?close=true>
- iii. **Sub regional reports:** similar to overview reports but focusing on a specific geographic area. This example is **South Sudan - Jonglei I Catalogue -Mobility Tracking Round 4**.
<https://displacement.iom.int/reports/south-sudan-%E2%80%94-jonglei-i-catalogue-%E2%80%94-mobility-tracking-round-4?close=true>
- iv. **Brief report:** shorter version of a narrative report summarizing the key findings. This report is **Europe — Migrant Vulnerability to Human Trafficking and Exploitation (November 2017)**
<https://migration.iom.int/reports/europe-%E2%80%94-migrant-vulnerability-human-trafficking-and-exploitation-november-2017?close=true>

516. A model Report will depend on the availability and comprehensiveness of the data collected, however a structure suitable for refugee and IDP statistics will likely include:

- **A Foreword:** usually written by an authority figure, presenting the report and highlighting its relevance. This can be procured in advance from the invited authority.
- **Introduction/Executive Summary:** including a background context, justification for the exercise, description of the methodology, data collection period, locations covered, reference population, if applicable a description of the sampling and construction of weights, who was collecting data, limitations and constrains. Finally, the sections of the report should be presented and the main findings or most relevant figures/charts.
- **Glossary:** if needed, it is recommended to include a glossary defining key terms in the beginning of the document.
- **List of figures.**
- **Suggested content sections:**

Demographics chapter: providing the snapshot information of observed population groups based on sex-age and geographic disaggregation, household size; if available and applicable comparing changes with previous periods, also in terms of demographic balance (births and deaths). If available highlight vulnerable population (children and youth, women in vulnerability, disability, dismembered households).

Sectoral /living conditions chapters (e.g. Housing, Education, Health, Employment): detailing the current living conditions of the observed population.

Mobility chapter: focusing on the subset of the population that has moved in the last 12 months. It should outline places of origin and destination, habitual and current residence, as well as current displacement status (IDP, refugee or neither), and previous status if available.

- **Conclusion**
- **Annexes**

4) EXAMPLES OF EUROSTAT DIGITAL PUBLICATIONS:

- Digital publications (Eurostat) on e.g. 'The life of women and men in Europe – a statistical portrait': <https://ec.europa.eu/eurostat/cache/infographs/womenmen/>. More Eurostat hosted digital publications can be found [here](#) and details of all Eurostat's publications can be accessed here; <https://ec.europa.eu/eurostat/publications/all-publications>

5) EXAMPLES OF CREATING/USING VISUALISATION TOOLS:

- Visualisation tools (UNHCR):
<https://data2.unhcr.org/en/countries/>
- Visualisation tools (World Bank):
<https://data.worldbank.org/indicator/sm.pop.refg>
- Visualisation tools (European Commission):
<https://migrationdataportal.org/>
- Visualisation tools (Eurostat):
<https://ec.europa.eu/eurostat/cache/RCI/myregion/-?reg=RO31>
<https://ec.europa.eu/eurostat/cache/infographs/economy/desktop/index.html>
https://ec.europa.eu/eurostat/cache/infographs/youth/index_en.html
https://ec.europa.eu/eurostat/cache/infographs/youineu/index_en.html

6) EXAMPLES OF CREATING/USING VISUALISATION TOOLS: NON-INTERACTIVE:

<https://displacement.iom.int/reports/burundi-%E2%80%94-baseline-dashboard-march-2019?close=true>

7) EXAMPLES OF CREATING/USING VISUALISATION TOOLS: INTERACTIVE:

<https://app.powerbi.com/view?r=eyJrijoimjdhnm1mgityzdkzs00ntawltkwodetyjm2n2e2otu5zdy5iwideci6ije1odgynjklitzmitndninc1izdzllwjztq5yzhlnje4niisimmiojh9>

1. **Overview maps:** maps covering data that would be aligned with an overview report. Could be disseminated as an annex of the narrative report, as a stand-alone product or compiled with other maps in an atlas;

<https://displacement.iom.int/reports/tripoli-displacement-map-7-07-may-2019?close=true>

2. **Thematic maps:** similar to the thematic report, and focused on a specific analysis theme or population subset; <https://displacement.iom.int/reports/mozambique-%E2%80%94-tropical-cyclone-idai-map-temporary-relocation-sites-layout-1-april-2019?close=true>

8) EXAMPLES OF PUBLIC-USE MICRO-DATA SETS:

517. Micro-data sets (World Bank):

<http://microdata.worldbank.org/index.php/home>

Micro-data sets (Eurostat):

<https://ec.europa.eu/eurostat/web/microdata/public-microdata>

7.2.2. *Users / recipients of the dissemination products (“for whom”)*

518. The production of dissemination products must be adapted to the different users and their needs, particularly in the volume and detail of information required. Thus, there might be a need to have different user segments (as proposed below by the Task Force on Value of Official Statistics).⁵³ Such segmentation can serve as a basis for constructing well-based production and communication of statistics. These user groups include:

- A. Users with a general interest
 - Citizens
 - Media and journalists
 - Students and teachers
- B. Users with a pre-defined/structured interest
 - International policies and monitoring frameworks
 - International organizations
- C. Users with a specific subject/domain interest
 - Decision makers
 - Policy makers and analysts
 - Marketing analysts
 - Experts in a specific field
 - Private businesses
 - NGOs and associations

⁵³ <http://www.unecce.org/fileadmin/DAM/stats/publications/2018/ECECESSTAT20182.pdf> - page 8.

- D. Users with a reuse and reproduction interest
 - Other producers of official statistics
 - Private or government organizations providing information services/products
 - Other providers of information services (e.g. App builders)

- E. Users with a research interest
 - Scientific community – academics and researchers
 - Consultants and researchers in the government or private sector

519. This categorisation and the dissemination targeting that it makes possible was described as the “cornerstone” of NSDS (Guidelines)⁵⁴, and illustrated below (UNECA) to conceptually define the various user communities and their demand for data.

Figure 7.1: UNECA Information and Audience Model for ECA Data and Information and Knowledge products

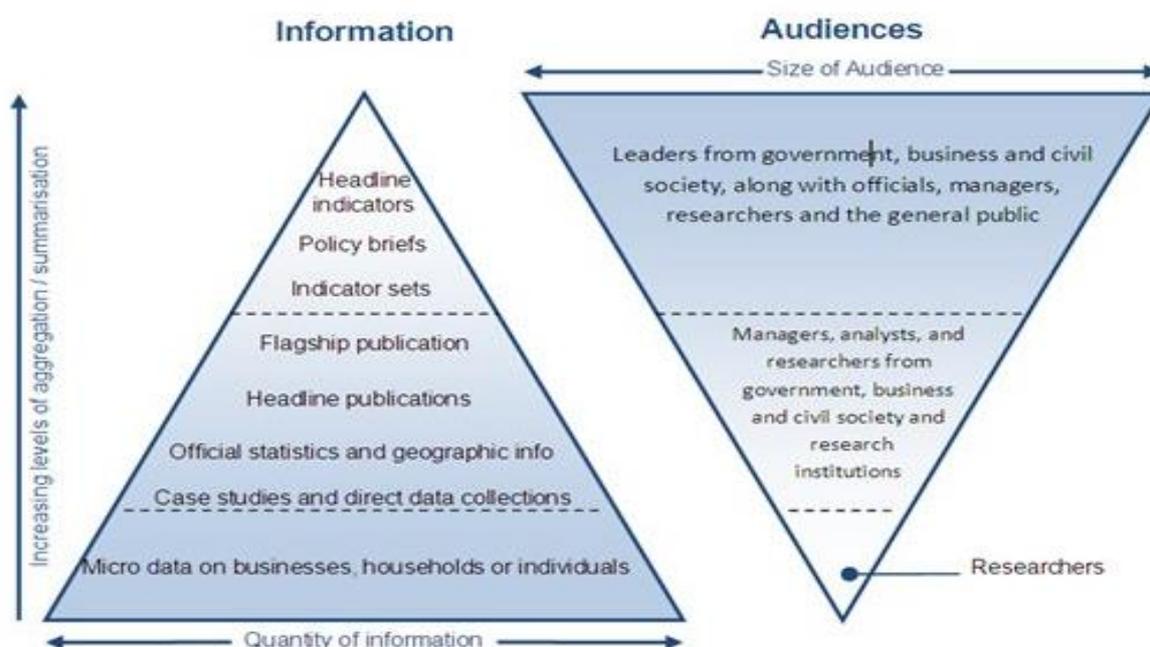


Figure 1: Information and audience model for the ECA data and related information and knowledge products.

Source: UNECA

7.2.3. Process ensuring proper creation of the dissemination products and best practices, guidelines (“how”)

520. When releasing the dissemination products, the style and formatting should be consistent and the quality of the product should be assured. At a simple level, the creation of

⁵⁴ <https://nsdsguidelines.paris21.org/node/796>

dissemination products is composed of the following steps:

1. Preparing the product components (explanatory texts, tables, charts, maps, quality statements etc.);
2. Assembling the components into products;
3. Editing the products and checking that they meet publication standards.

521. However, to ensure that quality standards are consistently met, statistical agencies should develop general guidelines for their publication products. Such publication guidelines are important for the following reasons⁵⁵:

1. they improve the quality of the publications by making them more user-friendly,
2. they set clear rules (that must be followed) and guidelines (that serve to advise) to facilitate the internal production process (and process of approval), by making it clear what the publication or website should contain and how this content should be structured and formatted.
3. Examples of publication guides and guidance on disseminating results.

522. Manuals produced by UNECE: "Making Data Meaningful" guide series- Parts 1, 2, 3 and 4: <https://www.unecce.org/stats/documents/writing/>

523. The **Making Data Meaningful** guides are intended as a practical tool to help managers, statisticians and media relations officers in statistical organizations to use text, tables, charts, maps and other devices to bring statistics to life for non-statisticians.

Part 1: A guide to writing stories about numbers - providing guidelines and examples on the use of effective writing techniques to make data meaningful.

Part 2: A guide to presenting statistics - providing guidelines and examples on preparing effective tables, charts and maps, and using other forms of visualizations to make data meaningful. It also offers advice on how to avoid bad or misleading visual presentations.

Part 3: A guide to communicating with the media - helping producers of statistics find the best way to get their message across and to communicate effectively with the media. It contains suggestions, guidelines and examples.

Part 4: A guide to improving statistical literacy - providing an overview of current initiatives and defining strategies for improving the statistical literacy of different user groups.

1) EXAMPLES OF GUIDELINES FOR DESIGN AND LAYOUT PUBLICATIONS:

⁵⁵ Source: https://paris21.org/sites/default/files/datapresentation_norway.pdf - page 61.

- Graphical Style guide (Eurostat)
https://ec.europa.eu/eurostat/documents/4187653/7192088/STYLE_GUIDE_2016.pdf/dbdf7b92-5ac5-4313-ae80-296f52692e5c
- Inter-institutional Style Guide:
<http://publications.europa.eu/code/en/en-000100.htm>

2) EXAMPLES OF SPECIFIC GUIDANCE ON METADATA, QUALITY ASSURANCE:

- [ESS Reference Metadata Reporting Standards](https://ec.europa.eu/eurostat/data/metadata/metadata-structure) (Eurostat)
<https://ec.europa.eu/eurostat/data/metadata/metadata-structure>
- [Quality Assurance Framework of the European Statistical System](https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646) (Eurostat)
<https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646>
- [ESS handbook for quality reports](https://ec.europa.eu/eurostat/documents/3859598/6651706/KS-GQ-15-003-EN-N.pdf) (Eurostat)
<https://ec.europa.eu/eurostat/documents/3859598/6651706/KS-GQ-15-003-EN-N.pdf>
- [Revised International Statistical Processes Assessment Checklist](https://unstats.un.org/unsd/dnss/QAF_comments/ISPAC%20v3%20Nov%202009%20formatted.PDF) (UNSD)
https://unstats.un.org/unsd/dnss/QAF_comments/ISPAC%20v3%20Nov%202009%20formatted.PDF
- [Administrative Data Quality Assurance Toolkit](https://www.statisticsauthority.gov.uk/archive/assessment/monitoring/administrative-data-and-official-statistics/quality-assurance-toolkit.pdf) (UK Statistics Authority)
<https://www.statisticsauthority.gov.uk/archive/assessment/monitoring/administrative-data-and-official-statistics/quality-assurance-toolkit.pdf>

7.3. Manage release of dissemination products

524. This sub-process should ensure that all elements for the release are in place. It includes managing the timing of the release, briefings for specific groups such as the press, as well as the arrangements for any pre-release embargoes. It also includes the provision of products to subscribers, and managing access to confidential data by authorised user groups, such as researchers and the error correction and notification policy.

525. Dissemination of statistical data to the wider public follows several dissemination principles set by dissemination policies. The management of the release of dissemination products requires careful planning and execution of the following aspects:

1. **Data release** – the statistical organization must ensure that all users have equitable and timely access to data that are disseminated to the public; there must be a clear guide for the users where they can find the data (on the respective portal/website and via different mobile apps),
2. **Time of release and data extraction** – the database should always contain the latest version of data; there should be regular intervals for the data' updates; the statistical organization should minimize the interval between the period to which the data refer and the date when the product is released to the public; in case of

revisions to the data after an initial release, the users must be notified about these changes in an equitable and timely manner; whenever preliminary data are released, they must be identified as preliminary; the ways that users can extract the data should be clearly explained (e.g. directly from the database or via 'bulk download' facility),

For example: Eurostat dissemination policy for Asylum statistics states that updates are done twice a week on Tuesday and Thursday, at 11:00 am

3. **Release calendar** should be available in advance of publication, e.g. prior to the beginning of the calendar year the statistical organization shall provide the users with a schedule of when statistical products are expected to be released during the upcoming year; such a calendar should be available on the portal/website and revised in a timely manner.

EXAMPLES

Eurostat release calendars: <https://ec.europa.eu/eurostat/news/release-calendar>

Statistics Norway: <https://www.ssb.no/en/omssb/lover-og-prinsipper/formidlingspolitikk/ssbs-statistikkalender>

4. **Relations with media** - when communicating with the media, two basic rules should be followed⁵⁶:
 - a. Equal treatment (all media should have access to releases of the new statistics at the same time),
 - b. Release calendar (all media should be informed about coming releases in advance in order to have some time for planning/research)
5. **Pre-release access/release of the statistical results under embargo** – this is the privileged pre-release access that should be limited, controlled and publicised; such access should be well described and published by the statistical organisation; the respective document should set out how the organisation fulfils its responsibilities in this respect and gives details of the conditions of pre-release access.

EXAMPLES

The '[*Protocol on impartial access to Eurostat data for users*](#)'⁵⁷ sets out how Eurostat fulfils its responsibilities in this respect and gives details on the conditions of pre-release access.

Eurostat signs specific Memorandum of Understanding (MoU) with the policy Directorates-General of the European Commission to allow the pre-release access to the statistical data.

⁵⁶ Source: <https://www.ssb.no/en/omssb/samarbeid/internasjonalt-utviklingssamarbeid/attachment/117795?ts=13f15ca60c0> page 50.

⁵⁷https://ec.europa.eu/eurostat/documents/10186/9268011/IMPARTIAL_ACCESS_2014_JAN-EN.PDF.pdf

6. **Error correction and notification** – a process of correcting errors should be established; the respective instructions should note errors both in online and printed statistical data and include the central principles and the situations in which a correction is required; errors when detected are recognised and documented, and the respective corrections/revisions are duly, rapidly and clearly disseminated to users.

Luxembourg example - Error correction and data revision at STATEC:

<https://statistiques.public.lu/en/actors/statec/policies/errors-revision/index.html>

7. **Access to microdata** - access to microdata should be restricted; e.g. for scientific purposes only; details and conditions on how to apply for microdata should be clearly explained;

EXAMPLES

- i. Eurostat examples: Eurostat pages dedicated to microdata access: <https://ec.europa.eu/eurostat/web/microdata/overview>
- ii. World Bank page: <http://microdata.worldbank.org/index.php/terms-of-use>

Figure 7.2: Microdata principles - UNFPOS etc.

Managing access to confidential data by authorised user groups⁵⁸

The sixth United Nations Fundamental Principle of Official Statistics is very clear on statistical confidentiality. Any principles for microdata access must be consistent with this Fundamental Principle.

“Individual data collected by statistical agencies for statistical compilation, whether or not they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.”

The following principles should be used for managing the confidentiality of microdata:

Principle 1: It is appropriate for microdata collected for official statistical purposes to be used for statistical analysis to support research as long as confidentiality is protected.

Principle 2: Microdata should only be made available for statistical purposes.

Principle 3: Provision of microdata should be consistent with legal and other necessary arrangements that ensure that confidentiality of the released microdata is protected.

Principle 4: The procedures for researcher access to microdata, as well as the uses and users of microdata, should be transparent and publicly available.

Links to guidelines and examples:

- UN Guidelines and good practices for managing statistical confidentiality and microdata: https://www.unece.org/fileadmin/DAM/stats/publications/Managing_statistical_confidentiality_and_microdata_access.pdf
- UN Principles and Guidelines for Managing Statistical Confidentiality and Microdata

⁵⁸ Source: https://www.unece.org/fileadmin/DAM/stats/publications/Managing_statistical_confidentiality_and_microdata_access.pdf - page 6.

Access: <https://unstats.un.org/unsd/statcom/doc07/BG-Microdata-E.pdf>

- Eurostat, Statistical confidentiality and personal data protection: <https://ec.europa.eu/eurostat/web/microdata/statistical-confidentiality-and-personal-data-protection>
- Eurostat, Statistical confidentiality: <https://ec.europa.eu/eurostat/web/research-methodology/statistical-confidentiality>

8. **Data re-use and copyright** – the statistical organization should have a policy of encouraging free re-use of its data, both for non-commercial and commercial purposes, provided that the respective data source is clearly indicated,

Example: Eurostat page dedicated to copyright notice and free re-use of data: <https://ec.europa.eu/eurostat/about/policies/copyright>

3) LINKS TO EXAMPLES OF DISSEMINATION POLICES, GUIDELINES, BEST PRACTICES:

- Principles for the dissemination policy of official statistics (UNSD): <https://unstats.un.org/unsd/dnss/docViewer.aspx?docID=2384>
- Dissemination policy (Eurostat): <https://ec.europa.eu/eurostat/about/policies/dissemination>
- Principles of communication and dissemination (Statistics Norway): <https://www.ssb.no/en/omssb/lover-og-prinsipper/formidlingspolitikk>
- JIPS guidance on data processing: <https://jet.jips.org/wp-content/uploads/Guidance-Data-Processing-Phase5-JET.pdf>

7.4. Promote dissemination products

526. Dissemination consists not only of the activities that are assumed with the data dissemination such as data publishing, but also activities related to the promotion of the dissemination products. Thus, this sub-process focusses on active promotion of the statistical products to help them reach the widest possible audience. It includes the use of different communication tools to reach the biggest audience and best facilitate the process of sharing statistical information with users:

7.4.1. Conferences, public debates, fairs

527. Conferences, public debates and fairs are still very useful and powerful channels of communication as the target audience (i.e. users interested in a specific topic) is present in one place. Through panels with speakers, meetings with subject matter experts, distribution of promotion materials, the users have a unique opportunity to acquire up-to-date information.

528. The information on forthcoming conferences, public debates, and fairs can be found on:

- UNHCR: <https://www.unhcr.org/conferences-and-meetings.html>
- European Commission: <https://ec.europa.eu/migrant-integration/upcoming-events>
- Research and Consulting on Migration and Asylum Policy Worldwide: <http://www.eurasylum.org/conferences/>

7.4.2. Social media

529. Social media can be a very useful tool in the dissemination and discussion of official statistics, as it offers tailor-made solutions to reaching individuals interested in official data. They are mainly used to distribute information, to promote the use of statistics and to increase statistical literacy. They can also be used to reach new audiences and to improve user support, as well as to receive user feedback and for marketing purposes.

530. Categories of social media to be taken into consideration⁵⁹:

1. **Collaborative projects, e.g. Wikipedia**, serve as an online database, allowing users to pool knowledge and information on particular topics to be viewed by other interested parties.
2. **Blogs and microblogs, e.g. WordPress or Twitter**, are a type of website or part of a website maintained by an individual with regular entries of commentary, graphics or video. Many blogs provide commentary or news on a particular subject; others function as more personal online diaries.
3. **Content communities, e.g. YouTube or Flickr, Delicious**, where users can upload, share, and view multimedia content such as videos, pictures, music or presentations.
4. **Social networking sites, e.g. Facebook or LinkedIn**, are online services, platforms or sites that focus on building and reflecting social networks or social relations among people who share interests or activities.

1) LINKS TO SOCIAL MEDIA GUIDELINES AND EXAMPLES:

- Social Media Guidelines: <https://ec.europa.eu/eurostat/documents/42577/43978/Social-Media-Guidelines-09-2011>
- Recommendations for Promoting, Measuring and Communicating the Value of Official Statistics: <http://www.unece.org/index.php?id=51139>
- UNHCR Twitter: <https://twitter.com/Refugees>
- Eurostat Twitter: https://twitter.com/EU_Eurostat

⁵⁹ Source: <https://ec.europa.eu/eurostat/documents/42577/43978/Social-Media-Guidelines-09-2011> - page 2.

- UNHCR Facebook: <https://www.facebook.com/UNHCR/>
- Eurostat Facebook: <https://www.facebook.com/EurostatStatistics/>
- UNHCR YouTube channel: <https://www.youtube.com/channel/UC2ZDAoIMGP3m9tAgnniNEYw>

7.4.3. Other forms of dissemination - visualisation tools, mobile apps, GIS portals

531. **Data visualisation tools** present data in an attractive and easy-understandable way for everyone to explore. **Mobile apps** make it possible to access the data on mobile devices. **GIS Portals** are specialised platforms to display geodata on maps and provide tools to interact with the data.

Links to examples of other dissemination tools:

- Eurostat data visualisation tools and mobile apps: <https://ec.europa.eu/eurostat/help/first-visit/tools>
- Eurostat – the Statistical Atlas: <https://ec.europa.eu/eurostat/web/gisco/gisco-activities/statistical-atlas>
- Central Statistical Bureau of Latvia thematic maps: [https://www.csb.gov.lv/en/statistics/search?product_type\[map_spatial\]=map_spatial](https://www.csb.gov.lv/en/statistics/search?product_type[map_spatial]=map_spatial)
- Tool for visualisation and dissemination of geospatial statistics in Slovenia: <https://www.efgs.info/information-base/case-study/dissemination/tool-for-visualisation-and-dissemination-of-geospatial-statistics-in-slovenia/>
- e-Learning Platform of the United Nations Statistics Division: <https://elearning-cms.unstats.un.org/course>

Links to the examples of communications strategies:

- UNHCR's communications strategy: <https://www.unhcr.org/excom/standcom/559644479/unhcrs-communications-strategy.html>
- NSDS Guidelines: [An example of advocacy and communication strategy \(Rwanda\)](#)
- Denmark: [Communication and dissemination strategy 2016-2020](#)

7.5. Manage user support

532. This sub-process ensures that user queries and requests for services such as microdata access are recorded, and that responses are provided within agreed deadlines. These queries and requests should be regularly reviewed to provide an input to the overarching quality management process, as they can indicate new or changing user needs.

533. Replies to user requests can also be used to populate a knowledge database or a "Frequently Asked Questions" page that is made publicly available, thus reducing the burden of replying to repeated and/or similar requests from external users.

534. User support provides assistance to all types of users: newcomers, sporadic visitors, more experienced users, journalists. For those who are visiting the portal/website for the first time a short 'first-time help' video can be offered, while for all users there are replies to frequently asked questions. Additionally, the statistical agency can propose more personal user support for those who would like to submit questions in their national language.

535. If data are not available from the portal/website at the level of detail or in the format or layout the user requires, the respective user support service may be able to provide customised data to meet the user's specific requirements, e.g. tabulations not readily available from Eurostat's online database can be requested via their user support service.

536. In some NSOs, statistics on refugees and IDPs may be a new product for those responsible for user support so staff may need additional training on dealing with the topic, particularly as it may be perceived as controversial.

537. Elements to be considered while establishing the user support service:

- i. Providing assistance to all types of users: newcomers, sporadic visitors or more experienced users. Example: Eurostat providing training to different policy Directorates-General of the European Commission on Asylum and managed migration statistics
- ii. Creation of the first-visit tutorial – for those who are visiting the website for the first time, Example: <https://ec.europa.eu/eurostat/help/first-visit>
- iii. Creation of the section dedicated to replies to frequently asked questions, Example: <https://ec.europa.eu/eurostat/help/faq>
- iv. Creation of the press centre for the use of journalists and fact-checkers, Example: <https://ec.europa.eu/eurostat/help/press-media-support>
- v. Creation of the multilingual user support giving the possibility to submit the questions in several languages (optionally, where relevant), Example: <https://ec.europa.eu/eurostat/help/support>;
- vi. Investigating new innovative solutions for user support – e.g. developing a Live Chat (real time conversations) Virtual Assistant tool with users, Example: <https://www.stat.ee/news-release-2019-042>

PART II – EXAMPLES AND CASE STUDIES

COUNTRY CASE STUDIES

a. Bosnia and Herzegovina

538. Statistics on IDPs and refugees are produced by the National Statistical System, namely the Ministry of Human Rights and Refugees. This is a state level organisation (Entity of Federation of B&H) responsible for unifying the data from other entities and districts, and for producing the report. The country's municipalities and cantons are responsible for collecting the data and for delivering it to the Ministry, together with The Ministry of Refugees and DPs of the Republika Srpska Entity; and the Government of Brčko District B&H, Department of Displaced Persons, Refugees, and Housing.

539. The reason for production of statistics about refugees and IDPs was the war in Bosnia and Herzegovina in the 1990s. Producing statistics on refugees and IDPs was initiated by the government at the beginning of the war. The local authorities during the war and after it were mostly engaged in this process due to the siege of the capital. When war ended the first comprehensive official registration of persons in BiH was carried out in late 2000 ministries and local authorities.

540. The revision of the number and the status of displaced persons started after compiling applications for the status revision throughout B&H, which was completed on 31/03/2005. Support from the international community was needed due to the fact that only international bodies had access to all parts of B&H during the war, and after the war they (UNHCR for example) participated in production of statistics related to recorded returns of displaced persons within B&H. The financial support given by international bodies participating, facilitated this collaborative production of refugee and IDPs statistics.

541. The main audience for the results are all the national, international, and NGO bodies involved or contributing to the B&H Strategy for the Implementation of Annex VII of the Dayton Peace Agreement, as well all other concerned parties.

542. The displacement statistics are now funded mainly by the government, but during and after the war the international community played an important role. The Swiss Agency for Development and Cooperation (SDC) ensured financial support covering the costs for the work of the consultative working groups while producing the Revised Strategy of B&H for the implementation of Annex VII in 2010.

2) LEGAL MANDATE AND OBSTACLES

543. The Legal and policy frameworks for the production of displacement statistics are:

- Annex VII of the Dayton Peace Agreement;
- Law on Refugees from Bosnia and Herzegovina and Displaced Persons in B&H (state level);
- Law on Displaced Persons and Returnees in Entity of Federation of B&H and Refugees from B&H;
- Law on Displaced Persons, Returnees, and Refugees in Entity of Republika Srpska.

544. There are no significant gaps in the legislation. The problem is within administrative inconsistencies which provides unequal and different structures of the entities and the district involved. There is no unique methodology at the state level, and statistical measurement is conducted by entities and the district, and there is little communication between them.

1) METHODOLOGY AND DATA SOURCES

545. During the war and after it, as well today, official status of the displaced person is determined through the registration in the Database of Displaced Persons (DDPR). The data were collected continuously during the war (by registration), and the first comprehensive official registration of persons in B&H was carried out in late 2000 by ministries and local authorities. The revision of the number and the status of displaced persons started after compiling applications for the status revision throughout B&H, which was completed on 31/03/2005 by establishing of DDPR. After that it was updated monthly updated by Municipality – Canton – Entity. Today there are no significant updates planned, due to a debate about durable solutions and the ending of displacement.

546. Administrative records are the main source, namely Database of Displaced Persons (DDPR) established by registering of DP in 2005. This key data source establishes someone's status as IDP or refugee. The secondary data source was the census.

547. **Census:** The census includes specific questions related to the place of residence and returns. – The personal questionnaire from the census in 2013 which 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.

Extract from the individual form of BiH 2013 Census

13. PLACE WHERE THE PERSON RESIDED DURING THE 1991 CENSUS (ONLY PERSONS BORN BEFORE 1 APRIL 1991)

In the settlement the person resides now 1) → Question 14.

In other place in BiH 2) {

SETTLEMENT

MUNICIPALITY

NAME OF COUNTRY ABROAD

Abroad 3)

14. WAS THE PERSON REFUGEE FROM BiH (AFTER 30 APRIL 1991)

Yes 1) No 2) → Question 16.

15. HAS THE PERSON RETURNED FROM REFUGEE

Yes, to the place the person emigrated from 1) Yes, to other place in BiH 2) No, the person has not returned to BiH 3)

16. WAS THE PERSON DISPLACED IN BiH (AFTER 30 APRIL 1991)

Yes 1) {

SETTLEMENT THE PERSON WAS DISPLACED FROM

MUNICIPALITY

No 2) → Question 19.

17. HAS THE PERSON RETURNED TO THE SETTLEMENT HE/SHE WAS DISPLACED FROM

Yes 1) → Question 19. No 2) → Question 19.

18. IS THE PERSON STILL FORMALLY-LEGALLY CONSIDERED DISPLACED PERSON IN BiH

Yes 1) No 2)

19. CITIZENSHIP OF

BH 1) BH and other country 2) Other country 3) Without citizenship 4)

WRITE DOWN THE CITIZENSHIP OF OTHER COUNTRY

19.1. ENTITY CITIZENSHIP

FBH 1) RS 2) Does not declare 3)

548. **The Database of Displaced Persons:** is in electronic format and contains extensive set of variables as follows;

1. Municipality of first registration, Date of first registration.
2. Personal data: (Name, Ethnicity, Sex, Birthdate, Place of birth, Municipality, State, Nationality, Registrar Number, Number of Household members in 1992 and today, Status in Household in 1992 – Head of household etc., Have you ever been a refugee abroad after 1992?)
3. The address before war (Entity, RS, DB, Municipality, zip code, Village or Settlement, Address, Prewar Status: 0-not marked, 1-private property, 2-state's property, 3-company's property, 4-renting, 5-with parents, 6-with family or friends... ; Current Status: 0-not marked, 1- sold, 2- replaced, 3-rented to other person, 4-renting contract expired, 5-family members from 1992 (1990) still living there, 6-cousins or friend living there with permit..)
4. Current Address (Entity, RS, DB, Municipality, zip code, Village or Settlement, Address, Prewar Status: 0-not marked, 1-private property, 2-state's property, 3-company's property, 4-renting, 5-with parents, 6-with family or friends... ; 0-not marked, 1- with permit from local authorities, 2-without permit from local

- authorities, 3-with owner's permit, 4-with permit of occupancy rights owner)
5. Other personal data (Have you requested return of the property; Have you requested voluntary readmission; Do you want to go back = answers: yes, no, not sure. If answer is "no" or "not sure", please quote your reason – you can chose 3 answers: 1-safety; 2-I'll settle in place of my current residence; 3-The house is already occupied; 4-I have found a job; 5-Neighbours haven't returned; 6-House is destroyed/not suitable for living; 7-There is no job; 8-Education. Number of supported household members by a person; Number of children in the family <18 years; Number of employed family members. Has any of your household members from 1992 returned to the municipality he lived before war? Yes, no, not marked) Statistics (Special needs: 0-No; 1-physically challenged; 2-mentally challenged; 3-chronically ill; 4-blind; 5-Missing needed aid; 6-Single parent family; 7-Elder without means; 8-Other. Education degree: 0-No education, or unknown; 1-Low qualified; 2-Qualified; 3-Highly qualified; 4-Secondary school; 5-High school; 6-BD; 7-MD; 8-PhD; 9-Other. Way of family support: 0-No support at all; 1-Employed; 2-Private busyness; 3-Pension; 4-Disablity pension; 5-Social help; 6-Family of KIA; 7-Humanitarian help; 8-Other. Employment status of registered person; Employment status of spouse; Profession of registered person; Profession of spouse.)
 6. Members of household (Ordinal number; Surname; Father's name; Name; Kinship/Relationship; Birthdate; Sex; Registrar number)
 7. Reports 1-Request of Status Revision; 2-Status Revision Stage; 3-Independently Created Report; 4-Report according to nationality; 5-Report according to prewar status of housing unit; 6-Report according to current status of housing unit; 7-Report according to statement of return (a); 8-Report according to statement of return (b); 9- Report considering type of employment; 10-Report considering special needs; 11-Report considering type of accommodation and occupancy; 12-Report according to profession; 13 Report according to education level; 14-Report according to family support by registered person; 15-Double entries; 16-Divided households; 17 Reports per status; 18 Reports per addresses, number of households' members, and status; 19-Report per addresses, number of households' members, and entry date (registration)

549. Data collection starts with the municipalities in all three administrative units of B&H – Federation of B&H, Republika Srpska, and Brčko District. In Republika Srpska and Brčko District, data is delivered to the Ministry of Refugees and DPs and Department of Displaced Persons, Refugees, and Housing respectively. In Federation of B&H municipalities deliver the data to the cantons and from cantons it goes to the entity level to the Ministry of Human Rights and Refugees (Entity of Federation of B&H is "federation of cantons"). After the process is finished all is delivered to the Ministry of Human Rights and Refugees of Bosnia and Herzegovina at the national level. This instance is responsible for final data integration and occasional reports publication.

2) ENDING OF DISPLACEMENT STATUS

550. Displacement ending criteria and durable solutions are currently under debate. The Revised Strategy of Bosnia and Herzegovina for the Implementation of Annex VII of the

Dayton Peace Agreement, defines it as: completion of the return process of B&H refugees and internally displaced persons; implementation of repossession of property and reinstatement of occupancy

551. rights; completion of reconstruction process of housing units for the return needs; ensuring conditions for sustainable return and reintegration process in B&H. There was dedicated information about returns during the period from 1995 – 2010 but in local language. Link:

<http://www.mhrr.gov.ba/PDF/Izbjeglice/INFORMACIJA%20O%20POVRATKU%20DO%202010.pdf>

3) DISSEMINATION

552. The reports available from the DDR are listed above. Data is occasionally exchanged or sent by the municipalities. At the central level (Ministry of Human Rights and Refugees) is responsible of integrating the data received, preparing the reports on it, controlling multiple entries, and other activities.

553. There was web-based dissemination of statistics in form of mentioned Revised Strategy of Bosnia and Herzegovina for the Implementation of Annex VII of the Dayton Peace Agreement. This report includes extensive statistical information

Link: <http://www.mhrr.gov.ba/PDF/Izbjeglice/Revidirano%20strategija%20Engleski.pdf>

b. Cameroon

554. Due to the situation in the country, there is a focus on migration in the methodological design for household data collection through i) population censuses and ii) sample surveys. The following definitions to identify refugees and IDPs are established through respective legal conventions in Cameroon:

1) DEFINITIONS

555. Refugee Definition: 'Qui, craignant avec raison d'être persécutée du fait de sa race, de sa religion, de sa nationalité, de son appartenance à un certain groupe social ou de ses opinions politiques, se trouve hors du pays dont elle a la nationalité et qui ne peut ou, du fait de cette crainte, ne veut se réclamer de la protection de ce pays; ou qui, si elle n'a pas de nationalité et se trouve hors du pays dans lequel elle avait sa résidence habituelle à la suite de tels événements, ne peut ou, en raison de ladite crainte, ne veut y retourner' (CERD 1998 ; also see 1951 Convention).

556. Internally Displaced Persons Definition: persons or groups 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 or armed conflict, situations of generalized violence, violations of human rights or natural or human-made disasters, and who have not crossed an internationally recognized State border (Kampala Convention 2009).

2) GENERAL CENSUS OF POPULATION AND HOUSING

557. As part of the fourth General Census of Population and Housing, information on refugees and IDPs is collected. Several variables (around eight) capture information for IDPs, such as previous place of residence, current place of residence, duration in the residence at the current location, and reason for change of place of residence. For refugees, the information is captured at two different levels: *collective household information* provides the numbers and characteristics, whereas the *ordinary household questionnaire* allows to collect data on the same variables as for IDPs. The crucial difference between both statistical domains are the variables "country of origin" and "nationality" that will differentiate IDPs from refugees.

3) SAMPLE SURVEYS

558. Three main steps are needed to screen refugees or IDPs in Cameroon through sample surveys: mapping of sampled enumeration areas, completion of pilot survey (testing of tools and methodology), and data collection. The collective household questions asked are adjusted to the country need to compile refugees and IDP statistics, for example:

- Does the household include refugees?
- Number of refugees living in the household
- Which country do the refugees come from?
- Does the household include IDPs due to insecurity in parts of Cameroon?
- Number of people in the household displaced by insecurity in parts of Cameroon

- Which region do the IDPs come from?

559. Additionally, the data collection through the household questionnaire captures valuable, but sensitive, individual information, for example:

- Why (Name) is he / she coming to settle in this locality?

01 = Work 02 = Job search (job search) 03 = Join the family 04 = Health problem
05 = Education, training 06 = Natural disasters 07 = Expropriation, eviction from
housing 08 = Repatriation 09 = Acquiring one's own home 10 = Crisis / security
situation 11 = Political exile 12 = Agro-pastoral conflicts 13 = Other conflicts /
crises (to be specified) 96 = Other (to be specified)

- (Name) does he / she intend to remain in this locality, or settle elsewhere or return to the place of origin in case of resolution of the problem causing his / her displacement?

0 = Not concerned 1 = Remain in this locality 2 = Return to the place of origin 3 =
Moving to another locality in the same region 4 = Moving to another region 5 =
Moved abroad 6 = Other (to be specified) 8 = DK

560. The case of Cameroon illustrates a methodological approach for refugee and IDP data through established data sources, including census and surveys. While generalising this example to other countries may not be possible, it provides an approach to be considered by countries by tapping on existent sources to derive displacement data on their population.

c. Colombia

1) USE OF ADMINISTRATIVE RECORDS TO ASSESS DISPLACEMENT RELATED
VULNERABILITIES IN COLOMBIA

561. In Colombia, 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 is 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. Additionally, IDPs update their location directly to the Victims' Unit through services points located throughout the country.

562. The National Planning Department and the National Department for Social Prosperity collect and produce official statistics on the country's most vulnerable population. These datasets account for most of the administrative records used by the Victims' Unit to update socio economic variables of IDPs at an individual level. Other relevant datasets are the official education enrolment system, the social security system, the registers' office, the housing and agricultural subsidies registers, among others.

563. 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.

564. As a rule, data from administrative records prevail over the data collected by phone surveys. This is because of the official nature of the administrative record produced by the relevant government agency. Nevertheless, there are cases in which chronologically, data from a survey prevails over the administrative record. For instance, in order to assess the criteria for decent housing, a survey collected after the event of displacement takes precedence over an official administrative record collected before the event of displacement.

565. Data from the relevant administrative records are programmed to calculate: possession of identification according to age, access to food, access to health services, access to education for children, access to income generation, participation in programs of training, quality of housing, access to land subsidies, and participation in family reunification programs.

2) COLOMBIAN SINGLE VICTIM'S REGISTRY

566. 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 is a demand-driven system counting IDPs since 1985. Up to October 2019, 7.5 million IDPs are registered.

567. One of the main purposes to keep a registry is to take account of, and to monitor the evolution and impact of the conflict. Besides being an instrument of memory of the impact of the armed conflict, the Single Victims' Registry is also the baseline to measure progress towards durable solutions to displacement.

568. IDP households who wish to be included in the Single Victims' Registry should present

a declaration before the Public Ministry (Ombudsman's Regional Office, Procurator's Municipal Office). According to the national legislation an IDP has to declare within two years after the event of displacement took place.

569. The declaration is sent by the Public Ministry to the Victims' Unit where it is assessed under the principles of dignity, good faith, legitimate trust and prevalence of the fundamental right. The assessment consists of a cross tabulation of variables regarding the circumstances of time, place and mode in which the displacement occurred.

570. Based on the information provided in the declaration form, as well as the information consulted during the verification process, the Victims' Unit makes a decision, either granting the status of IDP or not. IDPs are registered individually according to the household members who were forcibly displaced and thus included in the declaration form.

571. Once an IDP is granted the status of victim and they are included in the Single Victims' Registry. Once registered they are entitled to access relevant assistance and reparation. The registration in the Single Victims' Registry should be sufficient for the relevant government agencies to provide support, assistance and reparation in compliance with the law.

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

573. For statistical purposes, an IDP is taken out of the stock when they overcome displacement related vulnerabilities.

3) ASSESSING THE VULNERABILITY OF IDPs IN COLOMBIA

574. The Government of Colombia assesses the vulnerability of IDPs both to target social assistance and to report the progress in finding solutions to displacement. While this is not precisely the same methodology as that proposed in the IRIS (2020), the Colombian framework is similar as it uses the same assessment criteria as those proposed by IRIS to determine the end of displacement. These criteria come from the IASC framework on durable solutions and the indicators have been adapted to reflect the data that is available in the country. Overcoming vulnerability does not affect the stock of IDPs. A separate stock of IDPs, who have overcome vulnerability, is calculated periodically.

575. Overcoming vulnerability is linked to the process of achieving socioeconomic stabilization. For this purpose, the status of identification, health, education, food security, housing, income generation, psychosocial assistance and family reunification is assessed. The Socioeconomic Re-establishment Index (SRI) is used as a tool to calculate the extent of overcoming vulnerability. The SRI is calculated as follows:

$$\text{SRI} = (\text{ID} + \text{HE} + \text{ED} + \text{HO} + \text{FO} + \text{IG} + \text{PA} + \text{FR}) / 8$$

where,

ID = 1, if IDP possesses a national ID document according to age,

HE = 1, if IDP is covered under the national health system,

ED = 1, if IDP currently attends primary or lower secondary school and has finalised the previous grade

HO = 1, if IDP lives in a household with adequate housing conditions and benefits from

housing subsidy

FO = 1, if IDP exhibits acceptable frequency and diversity in the consumption of food,

IG = 1, if IDP lives above poverty line,

PA = 1, if IDP requested and accessed psychosocial assistance when needed,

and FR = 1, if IDP requested and achieved family unification when needed.

576. IDPs are considered to have overcome vulnerability, if SRI equals one. IDPs who have successfully overcome vulnerability are no longer referred to IDP support programs. They are instead supported through the social protection system. Regardless the vulnerability status, IDPs are entitled to accessing all reparation components.

577. As the circumstances of IDPs may change overtime, the purpose of the vulnerability assessment is to mark the point at IDPs no longer have any vulnerabilities related to displacement. Once an IDP successfully overcomes vulnerability, the person will not be assessed again unless a new event of displacement occurs.

4) IDP STATUS IN COLOMBIA WHO HAVE MOVED ABROAD

578. Even though Colombia complies with international standards on migration, the Colombian Victims' Law does not change a person's IDP status with any change in their current place of residence, including moves abroad, as long as the displacement occurred within Colombia's international borders. This situation may result in victims enrolling the Single Victim's Registry from another country through a declaration presented to the Colombian consulate. When an IDP wishes to be enrolled in the Single Victim's Registry from abroad, an internal displacement must have happened before they crossed the international border, for the enrolment to be successful when presenting the declaration at the Colombian consulate. Currently the government of Colombia does not count cross-border displacement.

579. Whether an IDP is entitled to receive assistance from the Colombian government while living abroad is subject to negotiation. The IDP status in Colombia will not normally be taken into account when applying for asylum abroad, it will depend on the laws and procedures of the country where asylum is sought. IDPs who later take up residence abroad are usually considered migrants, unless they seek international protection with a third country under the 1951 Convention relating to the Status of Refugees. A "refugee" is defined as a person who has crossed an international border "owing to well-founded fear of being persecuted for reasons of race, religion, nationality, membership of a particular social group or political opinion" In some contexts, the definition extends to persons fleeing "events seriously disturbing public order" (1969 OAU Convention; 1984 Cartagena Declaration).

d. Georgia

1) CENSUS COLLECTION OF IDP/REFUGEE STATISTICS

580. Geostat does not publish annual official statistics on refugees or IDPs. Data on IDPs were collected during the 2014 General Population Census. The person was asked the following questions:

10.1. Are you IDP or refugee?

10.2 Where are you IDP/refugee from? (Municipality, self-governing city/Country)

Also, the person was asked about the

Previous place of residence within the country in case of internal migration, and

Previous country of residence in case of immigration (for other details, please, see the Sections 8 and 9 of the questionnaire). Questionnaire from Census is available with this

<p>8. INTERNAL MIGRATION</p> <p>8.1 Have you ever lived in another settlement within the country for more than 12 months?</p> <p><input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>8.2 Previous place of residence within the country</p> <p><input type="checkbox"/> Urban settlement (City, borough)</p> <p><input type="checkbox"/> Rural settlement</p> <p>8.3 From which year have you been usually residing in the current settlement?</p> <p><input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/></p> <p>(Year)</p>	<p>9. IMMIGRATION</p> <p>9.1 Have you ever lived outside the country for more than 12 months?</p> <p><input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>9.2 Previous country of residence</p> <p><input type="text"/></p> <p>(Country)</p> <p>9.3 In which year did you move for permanent residence in Georgia from the country indicated in question 9.2?</p> <p><input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/></p> <p>(Year)</p>
<p>10. IDP OR REFUGEE STATUS</p> <p>10.1 Are you IDP or refugee?</p> <p><input type="checkbox"/> Yes <input type="checkbox"/> No</p> <p>10.2 Where are you IDP or refugee from?</p> <p><input type="text"/> <input type="text"/></p> <p>(Municipality, self-governing city) (Country)</p>	

581. Another source for the IDP statistics is administrative data from the Ministry of Internally Displaced Persons from the Occupied Territories, Labour, Health and Social Affairs of Georgia. The data is not in accordance with the definition on "usual place of residence" and it reflects the total number of registered IDPs for the time being rather than population with the IDP status permanently residing in Georgia.

582. According to the Law of Georgia on Internally Displaced Persons from the Occupied Territories of Georgia, "... A citizen of Georgia or a stateless minor in Georgia is eligible for IDP status if one or both parents have been granted and/or had IDP status."

583. The following tables are available online on the 2014 General Population Census [website](#) and [Geostat Statistics Database](#) in PC-Axis format tables:

1. IDP's by age and sex (by Age, Urban/Rural and Sex)⁶⁰;
2. IDP's by place of usual residence and place of residence before acquiring the IDP status (by Place of usual residence, Sex and place of residence before acquiring the IDP status).

⁶⁰ Law of Georgia "On Internally Displaced Persons from the Occupied Territories of Georgia": <https://matsne.gov.ge/en/document/view/2244506?publication=1>

Compilers' Manual Part II: Examples and Case Studies

IDP's by place of usual residence and place of residence before acquiring the IDP status					(person)
Place of usual residence	Total	place of residence before acquiring the IDP status:			
		Abkhazia A.R	Mtskheta-Mtianeti (Akhalgori municipality)	Tskhinvali region	
Both sexes					
GEORGIA	189,639	159,754	7,044	22,841	
C. Tbilisi	81,488	75,874	896	4,718	
Autonomous Republic of Adjara	3,414	3,345	...	60	
Guria	744	719	...	20	
Imereti	18,687	18,476	12	199	
Kakheti	1,586	1,250	11	325	
Mtskheta-Mtianeti	9,564	984	5,634	2,946	
Racha-Lechkhumi and Kvemo Svaneti	734	716	...	17	
Samegrelo-Zemo Svaneti	47,298	47,251	...	37	
Samtskhe-Javakheti	1,267	1,138	...	120	
Kvemo Kartli	11,314	8,236	168	2,910	
Shida Kartli	13,543	1,765	289	11,489	
Male					
GEORGIA	86,774	72,489	3,376	10,909	
C. Tbilisi	36,784	34,223	406	2,155	
Autonomous Republic of Adjara	1,524	1,500	...	21	
Guria	322	313	
Imereti	8,435	8,360	...	71	
Kakheti	661	510	...	150	
Mtskheta-Mtianeti	4,687	484	2,735	1,468	
Racha-Lechkhumi and Kvemo Svaneti	341	337	0	...	
Samegrelo-Zemo Svaneti	21,679	21,662	...	14	
Samtskhe-Javakheti	573	513	...	59	
Kvemo Kartli	5,312	3,818	84	1,410	
Shida Kartli	6,456	769	137	5,550	
Female					
GEORGIA	102,865	87,265	3,668	11,932	
C. Tbilisi	44,704	41,651	490	2,563	
Autonomous Republic of Adjara	1,890	1,845	...	39	
Guria	422	406	...	13	
Imereti	10,252	10,116	...	128	
Kakheti	925	740	...	175	
Mtskheta-Mtianeti	4,877	500	2,899	1,478	
Racha-Lechkhumi and Kvemo Svaneti	393	379	...	13	
Samegrelo-Zemo Svaneti	25,619	25,589	...	23	
Samtskhe-Javakheti	694	625	...	61	
Kvemo Kartli	6,002	4,418	84	1,500	
Shida Kartli	7,087	996	152	5,939	

IDP's by age and sex										(person)
Age	Total			Urban			Rural			
	Both sexes	Male	Female	Both sexes	Male	Female	Both sexes	Male	Female	
Total	189,639	86,774	102,865	140,070	63,835	76,235	49,569	22,939	26,630	
0-4	13,340	6,854	6,486	10,151	5,173	4,978	3,189	1,681	1,508	
5-9	14,159	7,410	6,749	10,650	5,532	5,118	3,509	1,878	1,631	
10-14	12,971	6,847	6,124	9,346	4,881	4,465	3,625	1,966	1,659	
15-19	13,033	6,670	6,363	9,033	4,569	4,464	4,000	2,101	1,899	
20-24	13,018	6,341	6,677	9,757	4,683	5,074	3,261	1,658	1,603	
25-29	13,979	6,617	7,362	10,602	5,055	5,547	3,377	1,562	1,815	
30-34	13,569	6,213	7,356	10,270	4,810	5,460	3,299	1,403	1,896	
35-39	13,156	5,888	7,268	9,884	4,600	5,284	3,272	1,288	1,984	
40-44	12,053	5,195	6,858	8,933	3,963	4,970	3,120	1,232	1,888	
45-49	11,672	4,904	6,768	8,592	3,588	5,004	3,080	1,316	1,764	
50-54	12,869	5,478	7,391	9,405	3,959	5,446	3,464	1,519	1,945	
55-59	12,207	5,192	7,015	9,055	3,768	5,287	3,152	1,424	1,728	
60-64	10,791	4,632	6,159	7,996	3,333	4,663	2,795	1,299	1,496	
65-69	7,066	3,044	4,022	5,144	2,174	2,970	1,922	870	1,052	
70-74	5,449	2,148	3,301	3,867	1,502	2,365	1,582	646	936	
75-79	5,473	1,818	3,655	3,903	1,235	2,668	1,570	583	987	
80-84	3,100	1,105	1,995	2,196	723	1,473	904	382	522	
85-89	1,347	352	995	1,006	239	767	341	113	228	
90-94	322	60	262	240	45	195	82	15	67	
95-99	57	...	51	35	...	32	22	...	19	
100+	...	0	0	0	...	

Source: Data are available in the Census-related publications: <http://census.ge/en/publication>

3. Also, Geostat disseminates the data to the users upon their requests.

584. The 2014 General Population Census results are available online on the 2014 General Population Census website and Geostat Statistics Database in PC-Axis format tables.

585. Access to microdata to the users is provided under the following circumstances: 10% of the 2014 General Population Census database can be shared upon the user's request and

following the principles of random sampling and Personal Data Protection.

e. Hungary

1) GENERAL INTRODUCTION TO NATIONAL SYSTEM FOR STATISTICS ON REFUGEES AND PERSONS TAKEN UNDER INTERNATIONAL PROTECTION

586. In Hungary both the National Directorate-General for Aliens Policing (NDGAP) and the Hungarian Central Statistical Office (HCSO) publish official statistics on persons involved in the forced migration.

587. The NDGAP is the authority which is responsible for the asylum procedure in Hungary. The IAO produces statistics on the asylum procedure, on numbers of persons taken under international protection status (refugees under the Geneva Convention and persons taken under subsidiary protection) and the figures of the persons who receive national humanitarian status (persons who are taken under the principle of non- refoulement).

588. The Hungarian Central Statistics Office (HCSO) receives the aggregated statistical data on asylum applicants from the NDGAP on a monthly basis and the aggregated data on positive decisions (refugees, subsidiary protected, and persons taken under subsidiary protection) on a quarter-year basis.

589. The NDGAP decides on the asylum applications⁶¹, recognising refugees under the Geneva Convention and persons who are under subsidiary protection. Data is forwarded to the Ministry of Interior which runs the Population and Address Register (PAR). In Hungary the persons taken under international protection are issued with identity card (and have similar rights to Hungarian citizens) and therefore their data is stored in the PAR.

590. In contrast to asylum seekers who are not considered to be part of the usually resident population, refugees and subsidiary protected persons are part of the usual residence population in Hungary and included in the PAR.

2) DISSEMINATION

591. Both of the HCSO and the NDGAP publishes only aggregated data on their web sites, the statistical publication of the NDGAP is accessible on the web site of the office: [NDGAP](#) . The NDGAP publishes its statistical data in an infrequent time period (mostly quarter yearly) the HCSO publishes the number of recognized refugees, subsidiary protected and the number of persons taken under the principle of non-refoulement once in a year while the data on asylum seekers registered in Hungary is published once in every three months. The date of the publication is indicated in the dissemination calendar of the HCSO which is transparently available on the following web site: [HCSO calendar](#).

⁶¹ Since the establishment of the so called 'transit-zones' the asylum procedure takes approximately 30 days but even before the existence of the 'transit-zones' the asylum procedure did not take longer than 60 days which indicates that this population subgroup is not part of the usual residence population of Hungary.

592. The statistical data are downloadable from the HCSO's Stadat – tables, the quarterly number of asylum applicants is available on the following web page: [HCSO applications](#). While the data on the positive decisions can be downloaded from the following Stadat – table: [HCSO recognition](#).

593. The HCSO also publishes statistical publication series 'Statistical Reflections' for the International World Refugee Day (20th of June) a short but comprehensive publication on the asylum related issue {([Statistical Reflections World Refugee Day](#)) unfortunately the last release was in 2016)}

594. The Ministry of Interior publishes the monthly flow data of persons who are registered in the PAR on a monthly basis (refugees and subsidiary protected), the data are available on the web site of the PAR: [Recognized refugees CPR](#). Unfortunately, these statistics are published in the Hungarian language only.

595. Asylum applicants are persons who applied for international protection at the authority responsible in Hungary (NDGAP) and the persons who are issued with a positive decision by the asylum authority can be recognized refugees or persons taken under subsidiary protection or taken under the principle of non-refoulement.

596. The NDGAP runs the Asylum-Register (Asylum Information System, abbreviation in Hungarian: MIR) where the relevant data in connection with the whole asylum procedure are stored. The Asylum-Register was established parallel with the foundation of the Immigration and Asylum Office in 2000.

597. Within the European Statistical System, the NDGAP is the in-line authority which is responsible for the data provision to the Eurostat under Article 4 of the Regulation 862/2007 EC and to other organizations. The HCSO does not provides statistics on asylum to the Eurostat and any other international organizations.

3) DESCRIPTION OF DATA SOURCES. USE OF ADMINISTRATIVE RECORDS FOR STATISTICS, ACCESS TO ADMINISTRATIVE DATA.

598. For asylum statistics, the NDGAP and the PAR are the main data sources of the HCSO. According to the Act No. CLV. of 2016 on Official Statistics ([Act NO CLV 2016](#)) the HCSO is entitled to have fully access to the administrative registers and can link record-level data with other administrative data. The law, Act No. LXXX of 2007 on Asylum ([Act LXXX 2007](#)) authorised the HCSO to have access to the record level data of the Asylum-Register prior to the adoption of the current Statistical Law. Nevertheless, this right has never been utilized by the HCSO, because the data identifying asylum seekers are not needed for the establishment of the usually resident population.

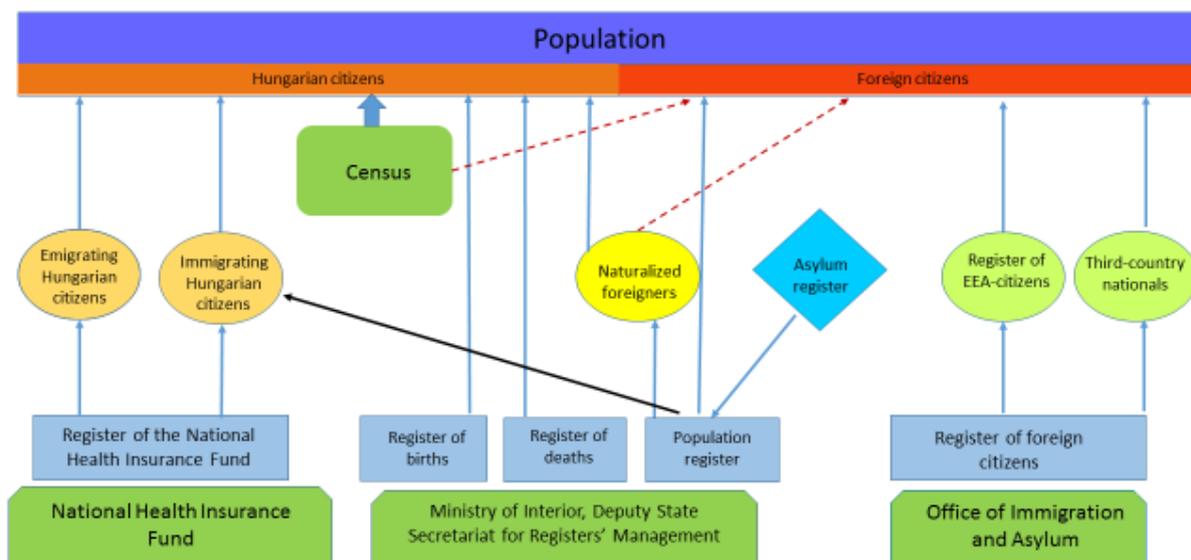
599. Before the new statistical law came into force in 2016, the HCSO had already entered into a memorandum of cooperation with NDGAP (since 2011) which regulates the formal background of the data exchange between the two authorities.

600. From the Ministry of the Interior the HCSO receives data on refugees and persons taken under subsidiary protection which enables their identification. After processing these data its number is added to the number of usually resident population in Hungary.

601. The statistical data exchange between HCSO, MOI and NDGAP is part of the national

statistical data collection programme (its abbreviation from the Hungarian is OSAP. Registration numbers are OSAP 2243 for asylum statistics from NDGAP , OSAP 2228 for persons taken under international protection from PAR). The channel of the data transmission is the secured KARÁT data exchange application of the HCSO ([KARÁT](#)).

Annual population calculation used by HCSO



4) POSSIBLE LEGAL/ADMINISTRATIVE CONSTRAINTS IN CREATION/USE OF ADMINISTRATIVE REGISTERS.

602. With the adaption of the new Statistical Act of 2016 all the existing constraints to the utilization of administrative data were removed. Both of the legal- and technical backgrounds are available for HCSO for the use of administrative registers.

5) EXISTING NATIONAL COORDINATION MECHANISMS PRODUCING REFUGEE STATISTICS; WHICH AGENCIES ARE INVOLVED; ROLE OF THE NATIONAL STATISTICAL OFFICE.

603. The main institution responsible for the storage of asylum and refugee statistics are the MOI and the NDGAP . Statistics on international protection are produced by both of these two institutions. According to the Regulation No. 223/2009 on European Statistics ([Regulation 223 2009 EC](#)) the NSIs in the Member States are the main coordination bodies on European Statistics, it means all types of statistics which are provided to the European Commission (Eurostat). Beyond this, the Act on Statistics regulates the statistical coordination role of the HCSO in Hungary.

604. Recently the Interreg Danube Transnational Programme⁶² which ended in 2019,

⁶² YOUMIG <http://www.interreg-danube.eu/approved-projects/youmig>; Improving institutional capacities and fostering cooperation to tackle the impacts of transnational youth migration
The two workshops were organized on 8th November 2018 and on 7th March 2019.

founded by the EU, called YOUMIG whose lead-partner is the HCSO. The partnership of the YOUMIG includes 16 financing project partners and 3 associated strategic partners from 8 countries (HU, AT, SI, RO, BG, SK, DE, RS). The project's duration is 30 months, from January 2017 to June 2019.

605. During the implementation of the project there were two workshops organized at the HCSO which aimed to improve of data exchange between HCSO and its main data suppliers. These workshops supported the coordination of access to administrative data sources. The YOUMIG workshop aims were to support statistical development in the area of international and domestic migration. Unfortunately, the two workshops were rather an exception in the general practice of the cooperation between the HCSO and its administrative data sources.

f. Nepal

1) USING MOBILE PHONE DATA FOR TRACKING POPULATION DISPLACEMENT IN NEPAL AFTER THE 2015 EARTHQUAKE

606. In April 2015, a devastating earthquake with a magnitude of 7.8 struck Nepal. After the initial earthquake, more than 300 aftershocks followed. In the earthquake aftermath, mass movements of the population occurred as people fled affected areas. Due to critical infrastructure failures and limited capacities, very limited information is available to understand where displaced people are located once these movements have begun, which is essential for the coordination of humanitarian response operations.

607. For this purpose, Flowminder and Ncell, the largest mobile operator in Nepal, agreed to collaborate in analysing mobile operator data to produce updated static population density maps, 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. Flowminder was given access to anonymised mobile phone data directly at the operator's premises for producing the maps after the earthquake. The WorldPop project mapping team estimated that around 1.8 million people above normal levels had left their home districts because of the 2015 earthquake.

608. Since 2015, Flowminder continues to analyse post-earthquake population movements and displacement in Nepal, and keeps working with key stakeholders to ensure that the capacity to respond with similar analyses in potential future earthquakes and disasters is maintained.

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

g. Norway

1) ADMINISTRATIVE DATA AND REFUGEE STATISTICS IN NORWAY

609. Refugee statistics in Norway are produced from administrative data. All asylum seekers must register with the Norwegian immigration authorities, and their case data and personal data are stored in a database maintained by the Norwegian Directorate of Immigration (UDI). The database contains data on UN convention refugees as well as refugees who enter Norway on their own and apply for asylum.

610. The UDI statistics division has access to copies of raw data from the database and has the ability to extract, aggregate and process the data in whatever way it chooses. At the same time, selected personal and case data are transferred directly from the case processing system to the Central Population Register (CPR), and from there transferred to Statistics Norway through daily batch jobs. UDI also provides supplemental, processed data to Statistics Norway in a data set that is compiled every year. The most important variable in this data set is each immigrant's reason for immigration.

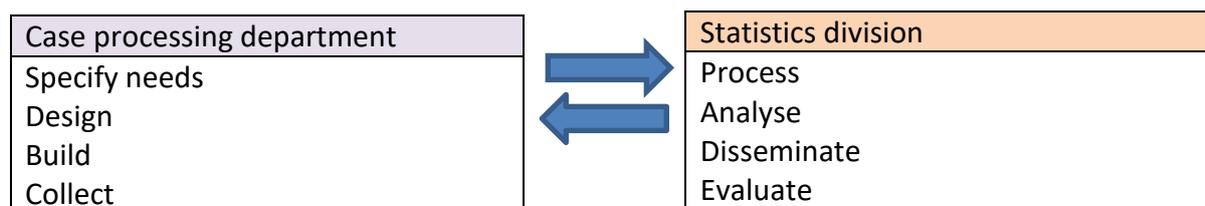
611. An obvious advantage of this system is that the producers of official statistics on refugees – Statistics Norway and UDI – have excellent access to data as well as control of how the data is processed, analysed and disseminated.

612. A major challenge is that data collection first and foremost is designed and built for case processing, not to produce statistics. Hence, the main purpose of the data collection is to facilitate the flow of applications through the administrative systems, while keeping track of the applicants' legal status and rights. Many people are involved in the handling of each application, and the system is built to enable each case handler to finish tasks and transfer the case to the next person, through all the steps from application to final decision. For this purpose, it may not even be necessary to access or collect structured data, as many case processing tasks can be resolved simply based on case documents.

The process above can be related to the main phases GSPBM model:

GSPBM: Main phases							
Specify needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate

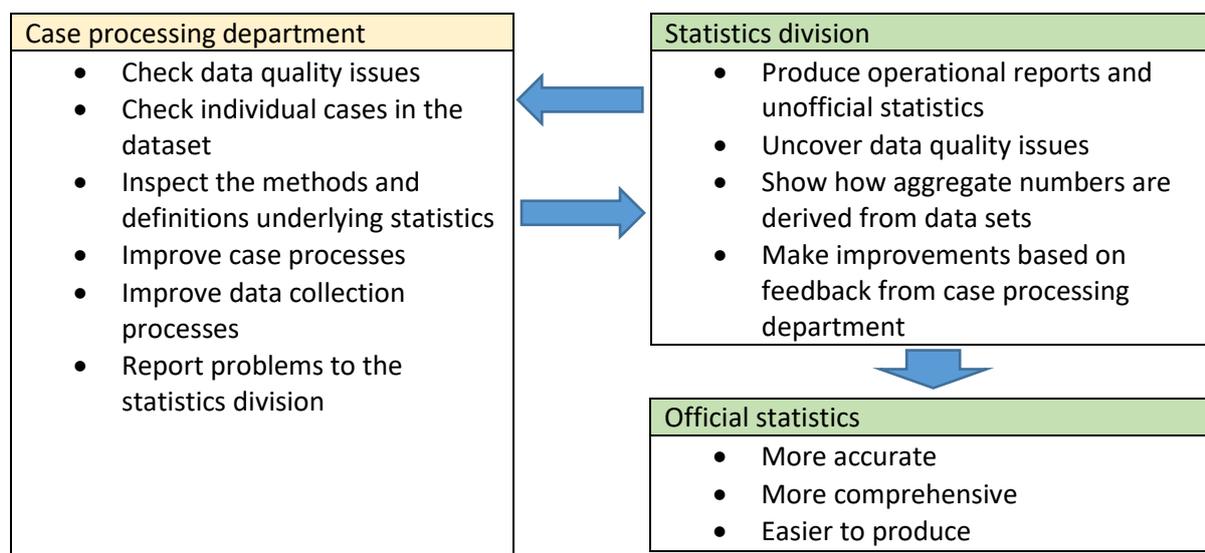
613. The first four phases, indicated in yellow, are designed mainly by the IT department in cooperation with the case processing department, while the last four phases, indicated in brown, are designed mainly by the statistics division. This can be illustrated as in the following diagram, showing the reciprocal relationship between these two groups of phases and the main responsibility for each group:



614. Ideally, during the first four phases – needs, design, build and collect – care should be taken to facilitate both case handling and statistics. The following four points are particularly important:

1. Data should be made accessible to the statistics division in raw form, not merely as aggregates from report modules.
2. Data should include personal ID and case ID, so that they can be linked to other data.
3. If possible, data should be structured, not in the form of free text.
4. Data that are logically linked, such as rejection and complaint, should also be physically linked with pointers in the database.

615. Norway experienced some problems pertaining to all four points above. The main issue was not a lack of access to data, but rather that the modelling of data is not always ideally suited for statistical purposes. While the UDI produces official statistics on refugees and other immigrant groups, it also produces unofficial and operational statistics, for internal use and ad hoc purposes. Such unofficial statistics are important both as operational tools and as means to improve the quality and completeness of official statistics. The mechanisms involved can be summarized as follows:



616. The UDI and the Central Population Register (CPR) are currently in the process of modernising the methods of data synchronisation, so that batch jobs and manual tasks will be replaced by continuous and automated data transfer. For asylum seekers this has already been implemented. When a person applies for asylum, a task event in UDI's case processing system triggers the requisition of a temporary personal ID number in the CPR. If the application is later approved, the decision triggers an update of the CPR, whereby the applicant is granted a permanent 11-digit personal ID-number. Statistics Norway receives data from the CPR through daily updates, and the 11-digit ID number ensures that data can be linked to data from other public registers.

h. South Sudan High Frequency Survey

1) METHODOLOGY

617. The High Frequency Survey: [Wave 4 and Crisis Recovery Survey 2017](#), South Sudan, 2017 Central, South Sudan National Bureau of Statistics was used to collect data about IDPs. Between May and August 2017 The World Bank in collaboration with South Sudan's National Bureau of Statistics, funded by DfID, conducted the fourth wave of the High Frequency Survey and the Crisis Recovery Survey to monitor welfare and perceptions of citizens in accessible urban areas and IDP camps across South Sudan. The HFS and CRS data contains information on security, economic conditions, education, employment, access to services, and perceptions. The data combines detailed household questionnaire information with displacement-specific information including drivers of displacement, access to resettlement mechanisms, and return intentions. It also includes comprehensive information on assets and consumption, to allow estimation of poverty based on the Rapid Consumption methodology as detailed in Pape and Mistiaen (2015).

618. **Coverage:** The High Frequency Survey covered urban areas of seven of South Sudan's ten former states: Western Equatoria, Central Equatoria, Eastern Equatoria, Northern Bahr-El-Ghazl, Western Bahr-El-Ghazal, Warrap and Lakes state. The CRS covered four of the largest Protection of Civilian (PoC) camps with defined boundaries: Bentiu, Bor, Juba, and Wau.

619. **Sampling:** Wave 4 of the High Frequency South Sudan Survey revisited urban households interviewed in Waves 1 and 2. Fifteen urban enumeration areas (EAs) visited in the first two waves were randomly selected from each state, and all of the households interviewed in the selected EAs were to be revisited. In Waves 1 and 2, the sampling strategy consisted of a stratified clustered design. Within each of the 7 strata (7 states, urban and rural) the primary sampling units are EAs that were drawn randomly proportional to size. Within EAs, a listing was conducted, and 12 households were drawn randomly as unit of observation.

620. 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 counted using satellite imagery of the EAs and strata. Each EA was divided into 12 blocks, and a micro listing was done in the blocks to randomly select households. One structure per block was selected, and one household per structure was interviewed.

621. EAs were replaced if security rendered field work unfeasible. Replacements were approved by the project manager. Households were not replaced and were dropped from the sample after a total of three unsuccessful visits.

622. **Weighting:** The selection probability for a household can be decomposed into the selection probability of the EA and the selection probability of the household within the EA. In the HFS, the selection probability of an EA is calculated as the number of households within the EA divided by the number of households within the stratum multiplied by the number of selected EAs in the stratum estimated using the 2008 census. The selection probability for a

household within an EA is constant across households and is calculated as the number of households selected in the EA over the number of listed households in the EA. Sampling weights were then scaled to equal the number of households per strata using the Census 2008 data.

623. In the CRS, the selection probability of an EA is calculated as the number of structures in the EA divided by the number of structures in the stratum multiplied by the number of EAs selected in the stratum. The number of structures was estimated using satellite imagery of the strata (camps). The selection probability of a household within the EA is decomposed into the selection probability of a block within the EA, the selection probability of a structure within a block, and the selection probability of a household within a structure. The sampling weights are then scaled to equal the number of structures per stratum as per the satellite imagery. In the Bor camp, the total number of households was similar to the target sample size, thus a census was conducted -- therefore, each household had a probability of selection of 1 and thus a sampling weight equal to 1.

2) QUESTIONNAIRE - MODULE I: CONFLICT AND DISPLACEMENT

Previous Section CONFLICT EXPOSURE (see xxxx)	
I. 16 What was your households' last place of permanent residence before displacement?	01 A different home in the same boma 02 A different location in the same payam 03 A different location in the same county 04 A different location in the same state 05 A different state in South Sudan 06 Outside South Sudan -98 Don't know -99 Refused to respond
I.17 In which state was your household's last place of permanent residence before displacement?	Code for State
I.18 Which country or region did you live in before displacement?	Code for Country
I.19 When were you displaced from your last place of permanent residence?	Date
I.20 Why did your household leave your last place of permanent residence before displacement?	01 Armed conflict 02 Increased crime, violence and insecurity but not armed conflict 03 Ethnic/political/religious discrimination and persecutions 04 Drought/famine/flood 05 Lack of access to home/land/livestock 06 Lack of access to education services and health services 07 Lack of employment opportunities 08 Death of husband/family reasons 09 IDP relocation program 10 Other reasons (specify) -98 Don't know -99 Refused to respond
I.19a Specify the main reason your household left your last place of permanent residence?	Text for main reason
I20 Please tell me more about the armed conflict. Was the conflict in your	01 Armed conflict in my village

Compilers' Manual Part II: Examples and Case Studies

village or nearby villages?	02 Armed conflict in nearby villages
I.21. How many times has your household changed residence after being displaced from the place of permanent residence?	Number of times
I.22 What was the main reason for you to leave the place you were staying in before coming to this place?	01 Armed conflict 02 Increased crime, violence and insecurity but not armed conflict 03 Ethnic/political/religious discrimination and persecutions 04 Drought/famine/flood 05 Lack of access to home/land/livestock 06 Lack of access to education services and health services 07 Lack of employment opportunities 08 Death of husband/family reasons 09 IDP relocation program 10 Lack of aid 11 Lack of proper management of the site/ site was crowded 12 Tensions with the host community 1000 Other reasons (specify) -0098 Don't know -0099 Refused to respond
I.23 When did the first members of your household arrive in the current location?	Date
I.24 Why did you choose to stay in this place?	01 Better security 02 Better access to home/land/livestock 03 Better access to education and health services 04 Better access to livelihood/employment opportunities 05 To join family / family reasons 06 I knew people who are settled in this place 07 Access to humanitarian aid (particularly food and water) 10 Other reasons (specify) -98 Don't know -99 Refused to respond
I.24 Please specify the main reason you chose to stay in this place.	
I.26 With whom did you arrive at this location?	01 Alone 02 With my family 03 With a larger group -98 Don't know -99 Refused to respond
I.29 Are you able to contact relatives or community members left behind at your original place of residence?	01 Yes 02 No
I.30 Have you or your household members ever gone back to your original place of residence after your displacement?	01 Yes 02 No
I.32 For what purpose did you go back to your original place of residence?	01 To visit household members/ relatives left behind 02 To check the status of property 03 For planting and harvesting 04 For income generation through business or trade 05 To gather information about the situation 06 To return and settle back 10 Other (Specify) -98 Don't know -99 Refused to respond
I.40 Where are most other people who were displaced from your place of origin now living?	01 In the same boma as this 02 In the same payam as this 03 In the same county as this 04 In the same state as this 05 A different state in South Sudan 06 Outside South Sudan -98 Don't know -99 Refused to respond

Compilers' Manual Part II: Examples and Case Studies

I.41 Did you or any of the household members lose any legal documents in the course of displacement?	01 Yes 02 No
I.42 Do you or the household members have access to mechanisms to obtain new documents or replace lost documents?	01 Yes 02 No
I.44 Do you feel free to move in and out of this settlement whenever you choose?	01 Yes 02 No
I.45 Did you have to give money or goods in order to secure a specific shelter here?	
I.46 Who did you give money or goods to? Select all that apply.	01 Government authorities 02 Military 03 Community organization or gang (gatekeepers) 04 Tribal authorities 05 NGO or UN personnel 10 Other
I.46 Please specify who did you give money or goods to?	
I.47 What did you give them? Select all that apply.	01 Food 02 Assets 03 Money 04 Labor 10 Other
I.47 Please specify what you give them.	
MOVEMENT AND RETURN	
I.48 Do you want to leave this location?	01 Yes 02 No
I.48 Do you plan to return to your original place of residence or go to a new location?	01 I plan to return to my original place of residence 02 I plan to go to a new area
I.49 If you decided to leave this location, when do you think you will start relocating?	01 In less than 6 months 02 In 6 to 12 months 03 In more than 12 months 04 Don't know yet
I.50 To which place would you like to move?	01 A new home in the same boma 02 A new location in the same payam 03 A new location in the same county 04 A new location in the same state 05 A different state in South Sudan 06 Khartoum or elsewhere in North Sudan 07 A new country other than South Sudan or Sudan
I.51 Thinking about the place where you live now, what are the three most important reasons you want to stay? Please start with the most important reason, then state the second most important reason and then the third most important reason.	01 Better security here 02 Better access to home/land/livestock 03 Better access to education and health services 04 Better access to livelihood/employment opportunities 05 To continue living with family or community members / family reasons 06 Access to humanitarian aid (particularly food and water) 10 Other reasons (specify)
I.51 Please Specify the main reason you want to stay.	
I.52 Thinking about the place where you would go if you needed to move, what are the three main reason you don't want to move? Please start with the most important reason, then state the second most important reason and then the third most important reason.	01 Armed conflict 02 Increased crime, violence and insecurity but not armed conflict in the area 03 Fear of ethnic/political/religious discrimination and persecutions 04 Drought/famine/flood 05 Lack of access to home/land/livestock 06 Lack of access to education services and health services 07 Lack of employment opportunities

Compilers' Manual Part II: Examples and Case Studies

	<p>08 No family there anymore / family reasons</p> <p>09 Will lose access to humanitarian aid</p> <p>10 Other reasons (specify)</p>
I.52 Please specify the main reason you don't want to move.	
I.53 Please tell me more about the armed conflict here in the area you live now. Is the conflict in that area or nearby areas?	<p>01 Armed conflict in the area</p> <p>02 Armed conflict in nearby areas</p>
I.54 Thinking about your place of origin, what are the three most important reasons you don't want to move back? Please start with the most important, then state the second most important and then the third most important.	<p>01 Armed conflict</p> <p>02 Increased crime, violence and insecurity but not armed conflict in the area</p> <p>03 Fear of ethnic/political/religious discrimination and persecutions</p> <p>04 Drought/famine/flood</p> <p>05 Lack of access to home/land/livestock</p> <p>06 Lack of access to education services and health services</p> <p>07 Lack of employment opportunities</p> <p>08 No family there anymore / family reasons</p> <p>09 Will lose access to humanitarian aid</p> <p>10 Other reasons (specify)</p>
I.54 Please specify what is the main reason you do not want to go back to your place of origin.	
I.55 Thinking about the place where you would go, what are the three most important reasons you want to move? Please start with the most important, then state the second most important and then the third most important reason.	<p>01 Better security here</p> <p>02 Better access to home/land/livestock</p> <p>03 Better access to education and health services</p> <p>04 Better access to livelihood/employment opportunities</p> <p>05 To live with family or community members / family reasons</p> <p>06 Access to humanitarian aid (particularly food and water)</p> <p>10 Other reasons (specify)</p>
I.55 Please specify the main reason you want to move.	
I.56 Please tell us the most important things you need to be able to settle in the location of your preference. Please start with the most important thing, then state the next most important thing and then third most important thing.	<p>01 Security</p> <p>02 Mine clearance</p> <p>03 Access to original house/land</p> <p>04 Provision of new shelter / housing</p> <p>05 Access to a farm plot and / or grazing land</p> <p>06 Agricultural tools & seeds / livestock</p> <p>07 Funds / productive assets for re-establishing business</p> <p>08 Skills training for job</p> <p>09 Connections for jobs</p> <p>10 Access to markets</p> <p>11 Access to educational services</p> <p>12 Access to health services</p> <p>13 Access to water and/ or sanitation</p> <p>14 Access to aid</p> <p>15 Transportation arrangement</p> <p>16 Regrouping of family</p>
I.56 Please specify the most important things you need to be able to settle in the location of your preference?	
Next section INFORMTION NEEDS (see	

i. Turkey

1) BACKGROUND TO INTERNATIONAL MIGRATION STATISTICS

624. In Turkey, main data sources for international migration statistics are population censuses and administrative registers. In 2007, following the last traditional census in 2000, Address Based Population Registration System (ABPRS) was established. Since then, annual population and migration statistics started to be obtained from ABPRS by the Turkish Statistical Institute (TurkStat). TurkStat is responsible for producing stock and flow statistics on international migration by demographic and other related indicators. These statistics are produced without the breakdowns of refugee or refugee-like populations.

625. TurkStat is responsible for producing stock and flow statistics on international migration by demographic and other related indicators of migrants. For foreigners residing in Turkey, citizenship and country of birth statistics are produced by ABPRS. TurkStat produced information on immigration and emigration, namely annual migration flow statistics, for the first time based on administrative registers in 2018 for the years 2016 and 2017. Statistics on foreign stock population, migration flow statistics covering both Turkish citizens and foreigners are produced in “country of birth” and “country of citizenship” breakdowns (with no disaggregation by refugee or refugee like populations).

2) THE POPULATION REGISTER USED FOR STOCKS AND FLOWS STATISTICS

626. After establishment of the ABPRS in 2007, the Population and Housing Survey (PHS) was conducted with combined method (by conducting ABPRS based large-scaled sampling survey) in 2011. 2011 PHS was implemented to obtain information which is not available in the ABPRS at provincial level. Within the scope of international migration statistics, information on the place of birth, ever resided abroad, and the country of residence one year ago was compiled from 2011 PHS. In the next census round (2021 PHC), Turkey plans to conduct a register based census.

627. According to the Population Registration Law No. 5490, all addresses within the boundaries of the country were registered in the National Address Database (NAD) in 2006. By linking addresses of Turkish citizens living within the boundaries of the country with the Turkish identification number, Address Based Population Registration System (ABPRS) was established in 2007. The system was subsequently transferred to the Ministry of Interior General Directorate of Civil Registration and Nationality (GDCRN). In the context of Official Statistics Programme prepared according to the Turkish Statistical Law No.5429, annual information on population size by administrative division (province, district, town, village and quarter), and its basic characteristics (age-sex structure, place of registration, nationality, literacy and educational attainment, legal marital status, place of birth, type of households, and internal migration in province level) is announced to the public in January of the following year based on the registers.

628. In order to obtain foreign population residing in Turkey, foreign registers held by the General Directorate of Civil Registration and Nationality (GDCRN) and the residence-work permit registers are mainly used. For stock migration, citizenship and country of birth statistics are produced from ABPRS.

629. Within the scope of studies on development of statistics obtained from administrative registers, analyses on existing records were carried out and thereafter Turkish Statistical Institute produced information on immigration and emigration, namely annual migration flow statistics, for the first time based on administrative registers in 2018 for the years 2016 and 2017.

630. Similar to stock migration statistics, migration flow statistics are produced by “country of birth” and “country of citizenship” breakdowns. International migration variables that are produced by TurkStat from 1927 to the present are given in the table below.

International Migration Statistics in Turkey

International Migration Variables	Traditional censuses (1927-2000)	2011 PHC	ABPRS (Since 2007)
Stock Population			
<i>Population</i>	De-facto	De-jure	De-jure
<i>Country of Birth</i>	✓	✓	✓ -Since 2014-
<i>Country of Citizenship</i>	✓	X	✓ -Since 2012-
<i>Ever resided abroad</i>	X	✓	X
<i>First Year of Residence</i>	X	X	✓
Migration Flows			
<i>Immigration (country of birth, country of citizenship, previous country of residence) - Five years-</i>	✓	X	X
<i>Immigration (country of birth and country of citizenship) -One year-</i>	X	✓ X	✓ -Since 2016-
<i>Immigration (Previous country of residence) -One year-</i>	X	✓	X
<i>Emigration (country of birth and country of citizenship) -One year-</i>	X	X	✓ -Since 2016-
<i>Emigration statistics (Next country of residence) -One year-</i>	X	X	X
<i>Reason for migration</i>	✓ (2000)	✓	X -Studies are ongoing-

631. Along with the Göç-Net (migration management system), registers of residence permits, international protection, illegal migration and other events related to foreigners are kept in connection to each other. DGMM use the Göç-Net system for collection, analysis and dissemination of the data on international protection, residence permits and irregular migration. The system contains personal information and provides detailed statistical reports.

632. The principal variables defined concerning applicants of international protection are age, gender, educational status, nationality, marital status etc. Statistical figures on international protection applications are shared on DGMM’s official website and published in Annual Migration Reports of DGMM on a yearly basis.

3) SYRIANS IN TURKEY

633. The conflict in Syria has caused the biggest displacement crisis that humanity has ever faced and forced millions of Syrians to take refuge in neighbouring countries. With the “Temporary Protection Regulation” issued on 22.10.2014, Syrians in Turkey were given temporary protection status. Syrians can get a foreign identification number to perform social

and legal procedures and benefit from health, education, labour market access and social support services on account of the Regulation.

634. DGMM is the responsible authority for the registration and status decisions within the scope of the "temporary protection" regime which grants beneficiaries right to legal stay as well as some level of access to basic rights and services. Additionally, the DGMM is also responsible for monitoring the impact of refugee crisis on our country together with the developing integration policies. In order to improve the information on Syrians, the DGMM is performing an intensive effort in collaboration with the related institutes.

635. Furthermore, with the issuance of "Regulation on Work Permits of Foreigners under Temporary Protection", published in 15 January 2016, participation of Syrians into the employment was legalized. Moreover, in cooperation with international organizations, a number of projects are carried out for integration of Syrians under temporary protection.

636. Syrians under temporary protection are not covered in Address Based Population Registration System (ABPRS), where annual population statistics are produced. Because, "Temporary Protection Identity Documents" or work permits issued for this group are not equivalent to residence permits or documents which substitute for residence permits. Thus, this population group is not officially accepted as "resident population". Only the Syrians who entered the country legally (with the required official documents such as a passport) are included in ABPRS population like other foreigners holding residence permit/work permit. Briefly, while approximately 3.5 million Syrians under temporary protection are not covered in ABPRS, the number of Syrians in ABPRS was just 87,955 as of December 31, 2018.

637. The number of persons under temporary protection in Turkey which was 14,000 in 2012 reached 225,000 thousand in 2013, 1.5 million in 2014, 2.5 million in 2015, 3.5 million in 2017 and almost 3.6 million as of June 2018.

4) LEGAL BASIS FOR THE USE OF REGISTERS AND STATISTICAL COORDINATION

638. Administrative registers are the data sources that institutions/organizations use to carry out their mandate of providing services within the framework of their governing legislation. Administrative registers play a major role in compiling timely information for the production of official statistics. Therefore, the transformation of administrative registers by integrating them with other administrative registers is very important in the development of official statistics.

639. In Turkey, TurkStat has the authority to access administrative registers for statistical purposes according to Article 9 of Statistics Law of Turkey numbered 5429. With the Law, TurkStat is also responsible for determining the basic principles and standards for national registration systems, and for supporting and monitoring the organisations involved in statistical processes and outputs.

640. Within the scope of this law, in order to improve annual population and international migration statistics, a protocol was signed between TurkStat and DGMM on the use of resident permit records of DGMM. Studies on similar protocols or agreements on the use of records of other population groups (i.e. persons under international protection) will continue.

641. The use of administrative data is a prerequisite for statistics that TurkStat is to produce or elaborate within the scope of both national priorities (Government Programme, Priority

Transformation Program, The Tenth Development Plan and linked strategy plans) and international requirements (closing criteria of the EU Statistics Chapter and data transfer to international organizations).

642. Statistics on foreign population residing in Turkey are obtained from ABPRS by using the foreign registers held by the General Directorate of Civil Registration and Nationality (GDCRN), residence permits held by DGMM, work permits held by Directorate General of International Labour Force and birth/death records. For foreign stock population and foreign migration flows, same data sources are used.

643. The Foreigners and International Protection Law (LFIP) No. 6458 was adopted on 4 April 2013 and entered into force on 11 April 2014. By this law, Directorate General of Migration Management (DGMM) was established. Since 11 April 2014, DGMM has been responsible for not only carrying out the procedures related to registering foreign citizens and implementing policies and strategies related to migration. The LFIP and Official Statistics Programme, nominates DGMM as the executive authority for the production of statistics on residence permits, international protection (applications, refugees, conditional refugees, subsidiary protection), temporary protection and illegal migration.

644. The DGMM was established under the Ministry of Interior with a view to implementing policies and strategies related to migration; ensuring coordination between the related agencies and organizations in these matters; and carrying out the tasks and procedures related to foreigners' entry into, stay in, exit and removal from Turkey, international protection, temporary protection and protection of victims of human trafficking (Article 103, LFIP).

645. Along with the Göç-Net (migration management system), registers of residence permits, international protection, illegal migration and other events related to foreigners are kept in connection to each other. DGMM use the Göç-Net system for collection, analysis and dissemination of the data on international protection, residence permits and irregular migration. All tasks and procedures that are under the responsibility of DGMM's central and provincial organization are carried out and kept in this system.

5) STATISTICAL DEFINITIONS

646. **Foreigners:** In the ABPRS, the foreign population is calculated according to international de-jure concept by considering data derived from administrative registers of;

1. Foreigners holding a valid residence/work permit at the reference day,
2. Foreigners holding an identity document equivalent to residence permit (international protection, etc..) with a valid address declaration at the reference day,
3. Individuals who have already renounced his/her T.R. citizenship and who are residing in the country with a valid address declaration at the reference day.

647. In addition to Syrians under temporary protection, foreigners holding visas or residence permits shorter than 3 months with the purpose of training, tourism, scientific research, etc. are not covered.

648. Foreigners who are covered in the official population statistics are given as follows:

Population group	Explanation	Covered in total population ?
Foreigners holding residence permit valid shorter than 3 months	Individuals present in Turkey for a short period of time with the intention of tourism, attending courses, scientific research, etc.	No
Individuals under temporary protection	Individuals with a temporary protection identification document, who are granted the right to stay in Turkey. However, the document issued to this group is <u>not equivalent to a residence permit or a document which substitute for residence permits</u> . Additionally, the document does not grant the right for transition to long term residence permit; its duration is not taken into consideration in calculating the total term of residence permit durations and in Turkish citizenship application.	No
Refugees	Refugees are issued identity documents which have <u>three years validity and substitute for residence permits</u> .	Yes
Conditional refugees and individuals under subsidiary protection	Conditional refugees and individuals under subsidiary protection are issued identity documents which have <u>one-year validity and substitute for residence permits</u> .	Yes

649. **International protection:** The status granted to refugees, conditional refugees, and those with subsidiary protection.

1. **Refugee:** A person who as a result of events occurring in European countries and owing to 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 citizenship and is unable or, owing to such fear, is unwilling to avail himself or herself of the protection of that country; or who, not having a nationality and being outside the country of his former residence as a result of such events, is unable or, owing to such fear, is unwilling to return to it, shall be granted refugee status upon completion of the refugee status determination process (Article 61, LFIP).
2. **Conditional Refugee:** A person who as a result of events occurring outside European countries and owing to 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 or herself of the protection of that country; or who, not having a nationality and being outside the country of former habitual residence as a result of such events, is unable or, owing to such fear, is unwilling to return to it, shall be granted conditional refugee status upon completion of the refugee status determination process. Conditional refugees shall be allowed to reside in Turkey temporarily until they are resettled to a third country (Article 62, LFIP).
3. **Subsidiary Protection:** A foreigner or a stateless person, who neither could be qualified as a refugee nor as a conditional refugee, shall nevertheless be granted

subsidiary protection upon the status determination (Article 63, LFIP).

4. **Temporary protection:** The status may be provided for foreigners who have been forced to leave their country, cannot return to the country that they have left, and have arrived at or crossed the borders of Turkey in a mass influx situation seeking immediate and temporary protection (Article 91, LFIP).

6) TURKEY'S STATISTICAL WORK PROGRAMME

Detailed description of each sub-groups in Official Statistics Programme 2017-2021

Code	Sub Subject	Responsible Institution/ Organization	Related Institutions/ Organizations	Data Collecting Frequency	Dissemination Frequency	Dissemination Period	Data Collecting Methodology	Data Availability Status	Classifications Used	Estimation Level	Release Version	Explanation
2.11.1	Foreign Population Statistics	TURKSTAT	<ul style="list-style-type: none"> • Ministry of Interior (General Directorate of Civil Registration and Nationality, General Directorate of Security, General Directorate of Migration Management) • Ministry of Labour and Social Security (DG for Labour) • Ministry of Economy Directorate General of Free Zones, Overseas Investment and Services) • Council of Higher Education 	Annually	Annually	+/- 1 months	Administrative Register	Available	-	NUTS Level 3	Electronic Documents (doc, xls, pdf etc.)	Each year, the foreign population statistics will continue to be released based on the ABPRS and will be developed on the basis of indicators. In the context of improving the registers, registers of international protection applications, refugees, conditional refugees, subsidiary protection, temporary protection will be linked with existing registers (ABPRS, residence permit and work permit registers, foreign registers kept by GDCRN) until 2018. In addition, registers will be linked with the registers on acquisition and loss of citizenship. Until 2018, border registers will be linked with existing registers in order to keep the population mobility under control.
2.11.2	Immigration and Emigration	TURKSTAT	<ul style="list-style-type: none"> • Ministry of Interior (General Directorate of Civil Registration and Nationality, General Directorate of Security, General Directorate of Migration Management) • Ministry of Labour and Social Security (DG for Labour) • Ministry of Economy (General Directorate of Foreign Investments and Services General Directorate of Free Zones) • Council of Higher Education 	Annually	Annually	+/- 11 months	Administrative Register	Not available	-	Turkey	-	Basic statistics on immigration and emigration (in citizenship, place of birth, country of previous and next residence details) will be produced annually until 2019.
2.11.3	Statistics on Turkish Citizens Living Abroad	TURKSTAT	<ul style="list-style-type: none"> • Ministry of Interior (General Directorate of Civil Registration and Nationality) • Ministry of Foreign Affairs (Directorate General for Consular Affairs) • Ministry of Labour and Social Security (DG for Labour) 	Annually	Annually	+/- 11 months	Administrative Register	Partly Available	-	Turkey	-	The studies on production of stock data on Turkish citizens living abroad will continue and will be developed by related institutions until 2018.
2.11.4	Citizenship Statistics (Statistics on the acquisition and loss of citizenship)	Ministry of Interior (DG of Civil Registration and Nationality)	Ministry of Interior (DG of Migration Management)	Annually	Annually	+/- 6 months	Administrative Register	Partly Available	-	-	Electronic Documents (doc, xls, pdf etc.)	As of 2017, statistics on acquisition and loss of citizenship occurred during the year will be produced in the country detail by gender and age groups.
2.11.5	Statistics on Residence Permits, Refugees, Conditional Refugees, Subsidiary Protection, Temporary Protection and Illegal Migration	Ministry of Interior (DG of Migration Management)	TURKSTAT	Quarterly	Quarterly	+/- 1 months	Administrative Register	Partly Available	-	NUTS Level 3	Electronic Documents (doc, xls, pdf etc.)	Studies on keeping the registers of international protection applications, refugee, conditional refugee, subsidiary protection, temporary protection and irregular migration in compatible with international standards and transforming into statistics will continue. These statistics will be produced beginning from 2018.
2.11.6	Work Permit Statistics	Ministry of Labour and Social Security (DG of Labour)	<ul style="list-style-type: none"> • Ministry of Interior (DG of Migration Management) • TURKSTAT 	Annually	Annually	+/- 11 months	Administrative Register	Available	ISCO 88, ISCED 97, NACE	NUTS Level 3	-	Studies on improvement and diversification of subjected statistics will be performed.

650. The Official Statistics Programme (OSP) is based on the Statistics Law of Turkey No 5429. It is prepared for a 5-year-period to determine the basic principles and standards dealing with the production and dissemination of official statistics, and to produce reliable, timely, transparent and impartial data required at national and international level.

651. The OSP includes International Migration Statistics. The responsible institutions are determined for each topic. In order to produce complete, reliable, regular, comparable and compatible international migration statistics, TurkStat works on transformation of registers into statistical sources by development of existing data sources and exploring the availability of new data sources in cooperation with Ministry of Interior (DG of Civil Registration and Nationality, DG of Migration Management), Ministry of Family, Labour and Social Services, Council of Higher Education and Ministry of Trade.

652. Turkish Statistical System started its planning cycle in 2007 with "OSP 2007-2011". The second OSP covered the years 2012-2016, and the third OSP covers the period 2017-2021. The main aim of the OSP is to create a disciplined statistical production process. The Turkish

Statistical Law covers and regulates not only the TurkStat, but all of the institutions in the Turkish Statistical System. Detailed information about OSP is given in the following link;http://www.resmiistatistik.gov.tr/misc/Official_Statistical_Programme_2017-2021.pdf

7) STATISTICAL OUTPUTS

653. Annual population statistics derived from ABPRS are disseminated through “Press Releases” on the TurkStat web page. Users can also use the static tables and the special data dissemination database on TurkStat web page. DGMM, on the other hand, disseminates results of the statistics on residence permits, international protection (applications, refugee, conditional refugee, subsidiary protection), temporary protection and illegal migration on their official web site. Figures are continuously updated. DGMM also publishes “Annual Migration Report”.

8) SUMMARY

654. The context for the creation/use of administrative data on international migration in Turkey are as follows:

1. There is a legal basis for harmonization and processing of administrative data by TurkStat.
2. Personal identifiers used nationwide for both Turkish citizens and foreigners make administrative data integration and linkage possible.
3. When the administrative sources reflect cases or applications instead of "individuals" (i.e. records of work permits issued not the records of work permit holders), statistical processes may be more difficult.
4. Since the registers are kept for administrative purposes, even though good progress has been achieved, there is still need for some quality improvements in process of transforming administrative registers to statistical data.
5. The institutional mandates and data confidentiality concerns of administrative register holders may outweigh the Statistical Law, and these are the main factors slowing down the process of accessing or transferring data for statistical purposes.

655. The Population Registers cover foreigners including special groups such as asylum seekers and refugees, but these groups cannot be disaggregated.

656. Registers of DGMM works in a similar way with the population registers; but only cover foreigners who have legal residence in the country.

j. Ukraine

1) BACKGROUND

657. Several studies have been made about IDPs in Ukraine. Most of them are done by humanitarian organisations providing assistance to IDPs. The studies do not generate official

statistics but give an account for needs and services rendered within a limited geographical area. The International Organization for Migration regularly produce statistics on IDPs in Ukraine through their National Monitoring System, based on their Displacement Tracking Matrix approach. It collects and analyses information on the socio-economic characteristics of IDPs and IDP households. One challenge the IOM faces, is to have an updated sample for their survey.

658. The Ministry of Social Policy in Ukraine has a register of IDPs in the country. They are continuously working to improve the quality of the register.

659. The State Statistics Service of Ukraine is planning a Census, also aiming to identify IDPs, in order to make statistics disaggregated by IDP status.

660. In January 2019, the Ministry of Social Policy, the State Statistics Service, the International Organization for Migration and the Ministry of Temporary Occupied Territory first met to discuss the potential for a cooperation on development of official IDP statistics, aiming to align their efforts to produce policy relevant statistics of high quality.

661. In Ukraine, the Ministry of Social Policy of Ukraine is responsible for ensuring the creation and maintenance of a Unified Information Database on Internally Displaced Persons (UIB). A Cabinet of Ministers approved the decree, which determines the conditions for the creation, maintenance and access to information of the UIB.

2) LEGAL CONTEXT

662. In order to determine the legal status of the IDP and the creation of a legal basis for the development of a mechanism for the protection of rights and freedoms of the Parliament of Ukraine adopted:

1. The Law of Ukraine "On ensuring the rights and freedoms of citizens and legal regime in the temporarily occupied territory of Ukraine" (from April 15, 2014 № 1207-VII);
2. Law of Ukraine "On ensuring the rights and freedoms of internally displaced persons" (dated October 20, 2014 No. 1706-VII);
3. Law of Ukraine "On amendments to some laws of Ukraine on strengthening the guarantee of observance of the rights and freedoms of internally displaced persons" (December 24, 2015, No. 921-VIII).

663. Registration of the place of residence of IDP is carried out in accordance with the laws of Ukraine:

1. "On freedom of movement and free choice of place of residence in Ukraine", "On ensuring the rights and freedoms of internally displaced persons",
2. the Resolution of the Cabinet of Ministers of Ukraine dated October 01, 2014 No. 509 "On the registration of persons who move from the temporarily occupied territory of Ukraine and areas of anti-terrorist operation ",
3. the Resolution of the Cabinet of Ministers of Ukraine dated March 02, 2016 No.

207" On Approval of the Rules of Registration of the Residence and the Procedure for the Transfer of Bodies registration of information to the Unified State Demographic Register".

3) DATA COLLECTION

664. IDPs have a certificate of registration which is valid indefinitely and only those with the relevant certificate can be considered IDPs. According to Article 5 of the "On Ensuring the Rights and Freedoms of Internally Displaced Persons", residence in the territory in which the conflict occurred is the basis of registering a displaced person. The certificate of IDP registration includes information on the address of the person's abandoned place of residence. In addition, it is possible for the applicant to indicate both his or her last registered place of residence / stay in the conflict territory and his / her actual place of residence in the territory. In addition, the applicant must provide his current address in the controlled area. Unlike other citizens, IDPs have a document confirming their current residence in a controlled area which is different from their registered place of residence in the conflict territory.

665. In 2016, changes to the system of residence registration transferred the responsibility for registering peoples' residence from the State Migration Service authorities to the local self-governments, In March 2016 the Cabinet of Ministers adopted Resolution No. 207, which regulates the procedures for registering residence. Residence registration and the withdrawal of registration of residence, is carried out by the executive body of the village, settlement or city council. Where an executive body of the village councils do not legally exist, the village head undertakes this task.

666. The relevant civil registration acts do not provide for recording that a person is an IDP. Therefore, death registrations do not identify persons who were IDPs, and IDPs or IDP-related persons are not identified in birth registrations. Due to the lack of interconnectivity between administrative registers (Register of Territorial Communities, EDDR, EIBD IDPs, State Register of Civil Status Acts), it is impossible to identify persons as belonging to the IDP category in other administrative sources.

667. The State Statistics Committee of Ukraine (SSSU) receives only very aggregated information from UIB and is unable to carry out a quality assessment of the data and is unable to use this data in its demographic calculations. Therefore, SSSU in cooperation with SSB Norway created the Technical Working Group (TWG) on IDP, the members of which were representatives of the SSSU, the International Organization for Migration (IOM), the Ministry of Social Policy (MSP) and the Ministry of Temporary Occupied Territories (MTOT). SSSU leads this working group and arranges its meeting if necessary. The group aims are to increase the volume and quality of IDP statistics, as well as agreed on definitions and other metadata.

668. According to the Unified Information Database on Internally Displaced Persons (UIB), as of January 8, 2020, 1,433,454 IDPs from the temporarily occupied territories of Donetsk and Luhansk oblasts and the Autonomous Republic of Crimea were registered.

k. Yemen IDP Example

1) BACKGROUND

669. The statistical series was started as a result of the dramatic increase of displacement in Yemen due to the outbreak of conflict in March 2015. In order to better inform the humanitarian community about the location and needs of the displaced populations.

670. The IOM's Displacement Tracer Matrix (<https://displacement.iom.int/content/methodological-framework-used-displacement-tracking-matrix-operations-quantifying>) was launched in April, 2015. Data collection was split between IOM and UNHCR. IOM covered 12 governorates (Abyan, Aden, Al Bayda, Al Dhale'e, Al Jawf, Al Maharah, Hadramaut, Ibb, Lahjj, Shabwah, Socotra, and Taizz), and UNHCR covered the remaining 10 governorates (Al Hudaydah, Al Mahwit, Amanat Al Asimah, Amran, Dhamar, Hajjah, Marib, Raymah, Sana'a and Sa'ada). In May 2017, data collection in the 10 governorates falling under the responsibility of UNHCR, was no longer possible.

671. In July 2018, the access to governorates had improved and IOM signed a memorandum of understanding (MoU) with the National Authority for the Management and Coordination of Humanitarian Affairs and Disaster Recovery (NAMCHA), placing IOM responsible for data collection in all the 22 governorates. Within IRG (Internationally Recognized Government) areas, IOM's DTM has been operating throughout, with recent engagement from MOPIC and Central Statistical Office.

672. Currently, there is a coverage on 22 Governorates, 333 districts, 2,287 sub-areas and 16,492 locations, accounting for a stock of 3.6 million of IDPs and 1.3 million returnees. The results of this exercise inform strategic decision-making among key humanitarian and development actors, including national, regional and local authorities.

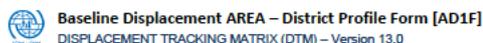
2) DESIGNING AND BUILDING THE STATISTICAL SYSTEM

673. There are five different exercise/modules implemented covering IDPs and the returnee population in Yemen. The data are collected once a year by the Task Force on Population Movement (TFPM). The task force is a Technical Working Group to the Inter-Cluster Coordination Mechanism. (ICCM) The TFPM implements an information management tool that gathers data and location of displaced persons across Yemen.

674. Two modules are related to the IDP and returnee stocks and their characteristics:

1. **Area Assessment:** The aim of the Area Assessment is to track and monitor IDP and returnee populations in Yemen's sub-districts. Information is collected on population size, area of origin, current location, duration of displacement, shelter types, priority needs and movement trends. The Area Assessment is based on the DTM Baseline Assessment tool.
2. **Multi-Cluster Location Assessment (MCLA):** A community-based assessment covering all 333 districts, and 7,715 locations in Yemen to understand the situation and needs of Host and Non-Host Communities, IDPs, Returnees, Refugees and Migrants.

not in the existing pre-existing OCHA location list. The estimated numbers of IDPs and IDP returnees are recorded on the form used.



D. 1-1 **RETURNEE:** List the village/neighborhood/sites, if there are more than 7 to be listed use the table on in section F and attach

For each RETURNEE location use one line to identify each location accommodating returnee populations in the district																									
For column 3&4. Period of arrival for majority; indicate the year (2015 '15' or 2016 '16') and month in 2015 or 2016 where the majority of IDP arrived where Jan =1 Feb = 2 Mar = 3 etc...																									
For columns 8. For type of residence/shelter enter the population of returnees who currently reside in each form of shelter within the location.																									
For column 9. Select the most predominant need among the returnee population in that location.																									
For column 9. Use one of the following Needs (Drinking Water=1 & Cooking/washing Water=2 & Food=3 & Health Sanitation/ Hygiene= 4 & Shelter/Housing=5 & Education for children=6 & Access to income=7 & Household items (NF)=8 & Legal help=9 & Civil Documents=10 & Psychological support=11 & Financial support=12 & Child protection services=13)																									
(1)	(2)	(3)	(4)	(5a)	(5b)	(5c)	(6)	(7a)	(7b)	(7c)	(8)							(9)							
Name of location - use the location list	Year of return for the majority	Month of return of majority in the identified year	Total number of RETURN HH	# of RETURN HH from Natural Disasters	Governorate of last displacement	Debt of last displacement	Sub-Area of last displacement	Returnee Population # per district	# of RETURN HH from Conflict per district	# of RETURN HH from Natural Disasters per district	In School Buildings	In Health Facilities	In Religious Buildings	In Other Private Building	In Other Public Building	In Settlements (Grouped of Families) Urban and	In isolated/dispersed settlements (detached)	In Rented Accommodation	With Host Families who are Relatives (no rent fee)	With Host Families who are not Relatives (no rent fee)	(Returnees) Original House of Habitual	In Second Home	Unknown - No data available/No access	Main priority need	
							(6) The sub-area of origin should be sorted by majority of the Returnee population, if more than 4 use the row after for the same location				[8] Shelter type accommodated. The total sum of Returnee accommodated in each shelter type should equal that of the total Returnee population in the location.														
1																									
2																									
3																									

680. Data are collected through an extensive Key Informant (KI) network within the operational area. The selection of the KI follows a standardized and structured approach, to ensure that the data collected in the Sub-Area Assessment is comprehensive and comparable across the different teams.

681. In the Yemen, the suggested practice for field teams and implementing partners is to select KI representatives of both the host and IDP communities while adhering to the humanitarian principles of humanity, neutrality, impartiality and operational independence. This ensures that the selected KIs are the most relevant and appropriate individuals to ensure the successful implementation of the exercise. The last report collected information from 9,917 Key Informants, of whom 5 per cent or 464 were female and 95 per cent or 9,453 were male.

682. The Sub-Area Assessment tool is used to verify and update the baseline information in quarterly cycles (Rounds). DTM staff, who are enumerators, are in regular communication with their KI network throughout each month and work continuously to maintain and expand this network to further triangulate the displacement statistics collected.

683. For the Multi-Cluster Location Assessment data collection is carried out by IOM, UNHCR, MOPIC, and CSO. Approx. 750 enumerators will visit the target locations and select Key Informants (KI) to complete the questionnaire based on the following criteria:

1. KIs should be selected based on their subject matter of expertise;
2. KIs should be either representatives from each population group or knowledgeable

about them;

3. KIs should be gender representative; and
4. KIs must adhere to the humanitarian principles and ethics in data collection.

684. The Area Assessment is a renaming of the DTM Baseline Assessment, and some specific modules had been developed for the MCLA methodology. As the MCLA questionnaire covers humanitarian needs in six different sectors (Shelter, Water and sanitation, Education, Health, Referrals, Access to income), the completion of each questionnaire may require more than one KI. KIs may be representatives from the following groups: NGO/humanitarian worker, social worker, IDP representative, community-based organization leader, religious leader, host community representative, local authorities, refugee representative, government official, education official, health worker, trader, engineer, and farmer.

685. This assessment follows a methodology centred on key informant (KI) interviews in geographic locations (villages and neighbourhoods) across all districts in Yemen.

686. Locations were randomly selected, whereas the Key Informants were selected by enumerators through purposive sampling, based on KI's level of knowledge on the themes covered in the MCLA questionnaire. The identification of target locations followed two different methods: one for IDPs, returnees, refugees, host communities, and non-host communities, and another for migrants.

687. A list of locations to be targeted in the MCLA was randomly drawn. The size measure of the locations was based on population statistics from secondary data sources. These were; the CSO 2004 census, 2018 projections (host communities and non-host communities), IOM DTM data (IDPs and returnees), and data collected by refugee leaders in 2018.). For each targeted population group district estimates were required, with 90% of confidence level and 20% of margin of error. Five locations were selected per district based on a random sampling approach: for IDPs, returnees, refugees, host communities, and non-host communities. The probability of each location being selected for this assessment was proportional to the population size of the population group in each location.

688. A final list of locations was prepared by merging the various sampling lists per population groups into one master list. Many locations repeated themselves, as they were selected more than once, that is, as part of the sampling lists of various population groups. In each location, one KII will be completed per target population group.

689. The number of forms and target locations per population group varies depending on the density and geographical distribution of the population groups in Yemen, not on the population group sizes. Therefore, a higher number of forms will be completed for population groups that are dispersed across the country (such as refugees), whereas a lower number of forms will be completed for groups that are more concentrated in fewer areas (such as returnees). Nonetheless, the MCLA coverage (the total size of each population group on which data will be collected through the MCLA) is proportional to the sizes of the population groups in Yemen.

690. The KoboToolBox data collection tools were used for data collection.

691. The data collection will seek to identify the local demographic profile, displacement dynamics, and to answer the following questions;

1. Who and how many people face moderate and severe humanitarian needs?
2. Where are these people located?
3. What do their survival and livelihood problems consist of?
4. How are the needs expected to evolve in the future, based on ongoing/planned responses and other likely events?

4) DEFINITIONS

692. Internally Displaced Persons (IDPs): persons or groups of persons who have been forced or obliged to flee 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 generalized violence, violations of human rights or natural or human-made disasters, and who have not crossed an internationally recognized State border (this includes individuals who moved within their locations, across locations, within their districts, across districts, within governorates, and across governorates).

693. Returnees: an IDP who has now returned to their place of habitual residence where they used to live prior to being displaced, irrespective of whether they have returned to their former residence or to another one.

694. Refugees: a person who, “owing to a well-founded fear of persecution for reasons of race, religion, nationality, membership of a particular social group or political opinions, 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” (Convention relating to the Status of Refugees, 1951). For the purpose of the MCLA, we will consider refugees to be: all individuals registered with UNHCR (asylum seekers and recognized refugees), all Somalis (as they receive prima facie refugee status in Yemen), and all foreigners who arrived in Yemen after August 2016 (and therefore have been unable to register with UNHCR) and who fled their countries of origin due to fear of persecution.

695. Migrants: any person who has crossed State borders on a voluntary basis for economic or other personal reasons.

696. Non-displaced persons: individuals who have never been displaced due to the crisis and who may or may not be hosting displaced populations within their households.

5) REPORTING

697. For reports see [Yemen Area Assessment Round 37 March 2019](#) and <https://displacement.iom.int/Yemen>.

Regional Capacity Building Examples

a. MED-HIMS

1) BACKGROUND

698. The Mediterranean Household International Migration Survey (MED-HIMS) is a joint initiative of the European Commission/Eurostat, The World Bank, ILO, IOM, UNFPA, UNHCR, and the League of Arab States, in collaboration with the National Statistical Offices of countries in the Southern and Eastern Mediterranean region. The programme is designed to overcome the lack of data on international migration for the region which can contribute to misperceptions about migration and can distort public opinion and debates on migration as well as policy planning. Link [MED-HIMS](#).

699. The main objectives of the MED-HIMS Programme are:

- i. to study the recent trends, causes, determinants, dynamics and consequences of international migration and mobility, and the inter-linkages between migration and development; and
- ii. to explore scenarios for a closer cooperation in the area of migration and development between the sending countries in the Southern and Eastern Mediterranean region and receiving countries, particularly the European Union.

2) METHODOLOGY

700. The MED-HIMS methodology is designed to deal with the multidimensional nature of international migration and mobility by carrying out specialized national household surveys in the sending countries that aim to capture current and recent developments in the Southern and Eastern Mediterranean region.

701. In order to maximize the quality, utility and comparability of the data collected through the MED-HIMS surveys, a set of model questionnaires (MQs) has been developed in collaboration with a panel of experts from the international organizations and the National Statistical Offices of countries participating in the programme. The MQs provide a holistic framework that deals with various dimensions of international migration and mobility through the collection of nationally representative data that are multi-topic, multi-level, retrospective and comparative.

702. The survey programme aims to cover the following five population groups: out migrants, return migrants, non-migrants (and among them prospective migrants), immigrants and refugees. Information on demographic and socio-economic characteristics of the household is also gathered. The MED_HIMS survey can be used for both immigrant refugee populations, and for those who have returned from seeking international protection abroad using the *return migrant* modules. The survey is also designed to identify those with a refugee background.

703. In addition to the Model Questionnaires, a series of manuals, guidelines and computer systems, covering the different phases of the survey from the initial organisation to training, data collection, tabulations, data analysis and reporting results, has been developed to provide countries with guidance on the design and implementation at the national level of the MED-HIMS survey or equivalent operations, whether under internationally coordinated programmes

or national stand-alone programmes, within or outside the Southern and Eastern Mediterranean region.

3) THE MODEL QUESTIONNAIRES

704. The first version of the Model Questionnaires was published in November 2013 and was used in the surveys in Egypt in 2013 and Jordan in 2014. In view of the rapid changes in the migration scene in the region, it has been decided to revise the MED-HIMS methodology including the set of MQs. The revised approach provides the following two survey types:

- i. MED-HIMS Standard Surveys (the long version) have large sample sizes and typically are conducted about every 5-7 years, to allow comparisons over time.
- ii. MED-HIMS Light Surveys (the short version) focus on the collection of information on key migration indicators and are conducted between rounds of MED-HIMS standard surveys. Although nationally representative, these surveys generally have smaller samples than the standard surveys.

705. This document presents the model questionnaires, version of 30 June 2019, for the light (short) MED-HIMS survey type. The two versions of the model questionnaires constitute the key model tools for harmonized and comparable data collection in the region.

706. For all participating countries, the MED-HIMS Light will make the following six manuals available:

- i. Manual 1: MED-HIMS Light Model Questionnaires
- ii. Manual 2: Design and Organization of MED-HIMS Light
- iii. Manual 3: Instructions to Supervisors
- iv. Manual 4: Instructions to Interviewers
- v. Manual 5: Sample Design
- vi. Manual 6: Guidelines for Country Report and Statistical Tabulations

707. The MED-HIMS Light will utilise the following seven questionnaires:

1. MQ-1 Light: Household Questionnaire
2. MQ-2 Light: Individual Questionnaire for Out Migrant
3. MQ-3 Light: Individual Questionnaire for Return Migrant
4. MQ-4 Light: Individual Questionnaire for Non Migrant
5. MQ-5 Light: Individual Questionnaire for Forced Migrant
6. MQ-6 Light: Individual Questionnaire for Immigrant
7. MQ-7 Light: Household Socio-economic Characteristics Questionnaire

4) MQ-5 LIGHT: INDIVIDUAL QUESTIONNAIRE FOR FORCED MIGRANTS

708. The questionnaire MQ-5 has a module for forced international migrants (refugees and refugee-related)

Box 5. MQ-5 Light: Individual Questionnaire for Forced Migrant. Details of Variables
Section 1: Migration Process
- Name of country of origin
- Date of birth, Age & sex
- Year of first move out of country of origin (CoO)
- Main reason for first move abroad
- Year of moving out of CoO as a refugee or for forced reasons
- Family members / relatives who left CoO with FM
- Other family members who joined FM in this country
- Number of countries stayed in since first move abroad before arriving in this country
- Name of last country stayed in before arriving in this country
- Main reason(s) for moving onwards from last country of asylum
- Ever applied for asylum in this country to the government or UNHCR
- Ever applied for asylum in any other country to the government or UNHCR
- Ever been recognized as a refugee
- Whether FM, when first moved abroad, knew of final destination & name of that country
- Factor(s) influencing decision to move to final destination before first move abroad
- Factor(s) determining chosen migratory route
- Financial source(s) for the migration at time of leaving CoO
- Whether FM had to pay money to anyone to get out of CoO or to move on to another country
- Difficulties FM encountered during outward journey
- Whether difficulties relayed to family still in CoO
- Means of transportation used on migratory route
- Conditions under which FM would have stayed in any of the countries travelled through
- Mother tongue / Native language
- Knowledge of other language(s)
Section 2: Situation in Host Country (THIS COUNTRY)
- Main reason(s) for choosing to move to this country
- Current status in host country
- Type(s) of documents FM has in country
- FM perceptions about local population's attitude to refugees
- Ever worked in CoO
- Status in employment in CoO
- Occupation in CoO
- Whether worked in six weeks before leaving CoO
- Whether currently working / Status in employment / Occupation
- Type of current job: permanent or temporary; full-time or part-time
- Days usually worked per week in current job
- Hours usually worked per day in current job
- Main production activity of current work place
- Whether currently looking for work

- Reason(s) for not looking for work
- Whether allowed to work in host country
- Access to and effective provision of public services (social protection)
- Assistance currently received from any person or organization in host country
- Whether receiving money from anyone living in another country
- Importance of money received for FM upkeep
- Whether sending money to anyone living in CoO or any other country
- Importance of money sent for their upkeep
Section 3: Prospects and Intentions
- FM future plans
- Reason(s) for wanting to stay in host country
Box 5. MQ-5 Light: Individual Questionnaire for Forced Migrant. Details of Variables, continued,
- Condition(s) required to return to home country / Meanwhile whether prefer to stay in this country or move to another country
- Reason(s) for wanting to move onwards to another country
- Name of country FM planning to move to
- Time-frame for planned move to another country
- Whether forced migrant felt discriminated against or harassed within the last 12 months on the basis of
- a ground of discrimination prohibited under international human rights law
- Awareness of risks of irregular migration
- Whether anyone residing with FM intending to move to another country/ Who? To what country? & When?
- Intention of members of FM family residing abroad to move to this country / Number intending to join FM
- FM advice to relatives/friends residing in CoO regarding moving abroad
Section 4: Marital Status and Reproduction
- Current marital status
- Age at and date of first marriage
- Whether married more than once
- Nationality of (first) spouse
- Country of current residence of (former) first spouse
- Number of times married
- Year of last marriage
- Nationality of (last/current) spouse at time of marriage
- Country of current residence of (last/current) spouse
- Number of own sons & own daughters living with FM
- Number of own sons & own daughters living elsewhere
- Among own sons living elsewhere, the number living in CoO & the number living elsewhere abroad
- Among own daughters living elsewhere, the number living in CoO & the number living elsewhere abroad
- Number of own children born in host country
- Number of own children under 15 years of age and their country of current residence
- Ever used (or spouse) family planning
- Number of sons and daughters when first used family planning

- Currently using any method of family planning
- Family planning method(s) currently used
- Intention to use family planning methods in future
- Reason(s) for not using family planning methods
- Ideal total number of children
Section 5: Health Status
- Self-rated health status
- Health insurance
- Whether FM or any member needed to have access to health care from any source in past 3 months
- Health care received in past 3 months
- Reason for seeking health care
- Type of health facility provided care
- Whether had to pay for health care received in past 3 months
- Total amount paid for all health care received in the past 3 months
- Degree of resilience using general self-efficacy scale

709. The Light Survey household questionnaire has a section designed to identify forced international migrants; this is reproduced below. These individuals identified as forced migrants are then selected for more detailed interview. The details include reasons for migration and information about the migrant process, demographic and fertility information and health information. The reasons for migration included in the individual questionnaire are longer and include;

Q. What was the main reason for moving from your country of origin for the first time?	a. Generalized insecurity / war
	b. Persecution related reasons
	c. Trafficking / Coercion
	d. Family reunification (within asylum procedure)
	e. To obtain asylum/refugee status
	f. Study abroad
	g. Find employment abroad
	h. Recruited to work abroad
	i. Business related reasons
	j. Tourism
	k. Health related reasons
	l. Visit holy places
	m. Family related reasons
	n. Other (specify): _

MED-HIMS Light Survey Household Questionnaire - Section 4: Identifying Forced Migrants and Immigrants (Non-Citizens)

Non-citizens residing in (SURVEY	Year of first arrival of non-citizens	Repeat migrant	IF 403=YES	Reason for leaving (COUNTR	Ability of non-citizen going back	Eligible forced migrants	Eligible immigrants
----------------------------------	---------------------------------------	----------------	------------	----------------------------	-----------------------------------	--------------------------	---------------------

Compilers' Manual Part II: Examples and Case Studies

COUNTRY)			Year of most recent arrival	Y OF ORIGIN)	to his/her country of origin		
401	402	403	404	405	406	407	408
INTERVIEWER: Check 108 (=2) & circle line number of every household member who is not a citizen of (SURVEY COUNTRY)	In what year did (NAME) come to reside for the first time in (SURVEY COUNTRY)?	Has (NAME) come to (SURVEY COUNTRY) more than once?	In what year did (NAME) most recently arrive in (SURVEY COUNTRY)?	What was (NAME)'s main reason for leaving his/her (COUNTRY OF ORIGIN)*?	Could (NAME) go back now to his/her (COUNTRY OF ORIGIN)?	Circle line number of every non-citizen who is currently aged 15+ years and whose reason for coming to (SURVEY COUNTRY)	Circle line number of every non-citizen who is currently aged 15+ years and whose reason for leaving (COUNTRY OF ORIGIN) in 405 = codes 2-11 and/or 406=1
	(YEAR) (DK=9998)	YES NO (GO TO 405)	(YEAR) (DK=9998)		YES NO (GO TO 408)	in 405 = codes 12-15 and/or 406=2	
01	_____	1 2	_____	_____	1 2	01	01
02	_____	1 2	_____	_____	1 2	02	02
03	_____	1 2	_____	_____	1 2	03	03
04	_____	1 2	_____	_____	1 2	04	04
.....	_____	1 2	_____	_____	1 2	05	05

*Codes for Q405: Reason for LEAVING COUNTRY OF ORIGIN:	
1. Transferred by employer (Diplomatic post)	9. Medical treatment
2. Transferred by employer (Non-Diplomatic post)	10. Family related reasons
3. Recruited to work abroad	11. This is (NAME)'s country of origin of parents
4. To look for employment	12. Insecurity/war in country of origin
5. No jobs available in country of origin	13. Persecution related reasons
6. Business / Investment related reasons	14. Trafficking / Coercion
7. Education / Study for self	15. To obtain asylum / refugee status
8. Education / Study for children	96. Other

710. The Standard Survey has a similar structure and methodology. The MED-HIMS Standard Survey model utilises eight model questionnaires (MQs) for the standard survey type:

MQ-1 (Standard): Household Questionnaire

MQ-2 (Standard): Individual Questionnaire for Out Migrant

MQ-3 (Standard): Individual Questionnaire for Return Migrant

MQ-4 (Standard): Individual Questionnaire for Non Migrant

MQ-5 (Standard): Individual Questionnaire for Forced Migrant

MQ-6 (Standard): Individual Questionnaire for Immigrant

MQ-7 (Standard): Household Socio-economic Characteristics Questionnaire
MQ-8 (Standard): Community Characteristics Questionnaire

711. Each of these Standard Model Questionnaires (MQs) is designed as a series of self contained modules, with each module dealing with a particular migration-related topic. The MQs provide the core set of questions needed to obtain population-based estimates of the determinants and consequences of international migration and mobility in the region. The MQs are designed to be manageable, economical, and to intrude as little as possible on the activities and privacy of families who are interviewed.

Box 5. MQ-5 (Standard): Individual Questionnaire for Forced Migrant. Details of Variables
Section 1: Migration Process
- Name of country of origin
- Date of birth, Age & sex
- Year of first move out of country of origin (CoO)
- Main reason for first move abroad
- Year of moving out of CoO as a refugee or for forced reasons
- Family members / relatives who left CoO with FM
- Other family members who joined FM in this country
- Number & names of countries stayed in since first move abroad before arriving in this country
- Main reason(s) for moving onwards from first country of asylum
- Main reason(s) for moving onwards from last country of asylum
- Ever applied for asylum in any country to the government or UNHCR
- Name(s) of countries in which FM applied for asylum
- Ever been recognized as a refugee
- Whether FM, when first moved abroad, knew of final destination & name of that country
- Factor(s) influencing decision to move to final destination before first move abroad
- Factor(s) determining chosen migratory route
- Financial source(s) for the migration at time of leaving CoO
- Whether FM had to pay money to anyone to get out of CoO or to move on to another country
- Whether any of those travelling with FM had to pay money to anyone to get out of CoO
- Difficulties FM encountered during outward journey

- Difficulties any of those travelling with FM encountered during outward journey
- Whether any of those travelling with FM died on the way to this country
- Whether difficulties relayed to family still in CoO
- Means of transportation used on migratory route
- Conditions under which FM would have stayed in any of the countries travelled through
- Conditions under which migrant would have stayed in any of the countries travelled through before arriving at current destination
Section 2: Situation in Host Country (THIS COUNTRY)
- Main reason(s) for choosing to move to this country
- Help received in asylum application by source
- Latest outcome of asylum application
- Current status in host country
- Type(s) of documents FM has in country
- FM perceptions about local population's attitude to refugees
- Reason(s) for FM perceptions
- Ever worked in CoO
- Status in employment in CoO
- Occupation in CoO
- Whether worked in six weeks before leaving CoO / Status in employment / Occupation
- Type of current job: permanent or temporary; full-time or part-time
- Days usually worked per week in current job
- Hours usually worked per day in current job
- Main production activity of current work place
- Payment (earnings) per day or per week or per month & currency
- Benefits provided by current employer
- Assistance needed to increase income
- Whether currently looking for work
- Reason(s) for not looking for work
Box 5. MQ-5 (Standar): Individual Questionnaire for Forced Migrant. Details of Variables, continued,
- Whether allowed to work in host country
- Access to and effective provision of public services (social protection)
- Assistance currently received from any person or organization in host country
- Whether receiving money from anyone living in another country
- Importance of money received for FM upkeep
- Whether sending money to anyone living in CoO or any other country
- Importance of money sent for their upkeep

- Mother tongue / Native language
- Knowledge of other language(s)
Section 3: Prospects and Intentions
- FM future plans
- Reason(s) for wanting to stay in host country
- Condition(s) required to return to home country / Meanwhile whether prefer to stay in this country or move to another country
- Reason(s) for wanting to move onwards to another country
- Name of country FM planning to move to
- Time-frame for planned move to another country
- Whether forced migrant felt discriminated against or harassed within the last 12 months on the basis of a ground of discrimination prohibited under international human rights law
- Awareness of risks of irregular migration
- Whether anyone residing with FM intending to move to another country/ Who? To what country? & When?
- Intention of members of FM family residing abroad to move to this country / Number intending to join FM
- FM advice to relatives/friends residing in CoO regarding moving abroad
Section 4: Marital Status and Reproduction
- Current marital status
- Age at and date of first marriage
- Whether married more than once
- Nationality of (first) spouse
- Reason for dissolution of first marriage
- Date first marriage dissolved by death of spouse
- Date first marriage dissolved by divorce
- Country of current residence of (former) first spouse
- Number of times married
- Age at and date of last marriage
- Nationality of (last/current) spouse at time of marriage
- Country of current residence of(last/current) spouse
- If Forced Migrant Male: Number of wives
- If Forced Migrant Female: Number of co-wives
- Number of own sons & own daughters living with FM
- Number of own sons & own daughters living elsewhere
- Among own sons living elsewhere, the number living in CoO & the number living elsewhere
- Among own daughters living elsewhere, the number living in CoO & the number living elsewhere

- Number of children born alive but now deceased
- Number of step-sons and step-daughters living with FM
- Number of own children born in host country
- Number of own children under 15 years of age and their country of current residence
- Ever used (or spouse) family planning & methods ever used
- Number of sons and daughters when first used family planning
- Currently using any method of family planning
- Family planning method(s) currently used
- Intention to use family planning methods in future
- Reason(s) for not using family planning methods
- Desire to have more children
- Preference of next child's sex
- Ideal total number of children
- Number of children wanted by spouse of Forced Migrant as compared with that wanted by Forced Migrant
Section 5: General Health Status
- Self-rated health status
- Health insurance
- Smoking every day any tobacco product
- Age when started to smoke on a daily basis
- Number of cigarettes smoked daily
- Number of 'shisha' sessions smoked daily/weekly/or only occasionally
- Use of smokeless tobacco
- Anyone smoked inside FM home in past seven days / Number of times
- Whether FM or any member needed to have access to health care from any source in past 3 months
- Health care received in past 3 months
- Reason for seeking health care
- Type of health facility provided care
- Total amount paid for all health care received in past 3 months
- Main reason for not receiving needed health care
- Whether diagnosed by a doctor of having a psychological condition
- Date of diagnosis of the psychological condition
- Currently receiving any treatment for psychological condition
- Length of time since having bad psychological problem
- Degree of resilience using general self efficacy scale

712. More information and the specimen questionnaires can be found at

<https://ec.europa.eu/eurostat/documents/52535/8125918/Manual+1+LIGHT+Model+Questionnaires/f8b70a20-aec8-26cf-5e03-ac47713bfc71>

5) TABULATION PLANS

713. The following tabulation plans for forced migrants are recommended under the MED-HIMS programme.

A. Tabulations of data gathered in the household interview

Table 1 Forced migrants households and population

Distribution of the households and population enumerated in the forced migration survey according to country of origin

Table 2 Household population by age, according to sex and nationality

Percent distribution of the population enumerated in the forced migration survey, by broad age groups, according to sex and country of origin

Table 3 Household headship and composition

Percent distribution of households enumerated in the forced migration survey, by sex of head of household, and household size, according to country of origin

Table 4 Educational attainment of household population (ages 10+)

Percent distribution of the population enumerated in the forced migration survey at ages 10 years and over by highest level of education attended, according to sex and country of origin

Table 5 Employment status of household population (ages 15+)

Percent distribution of the population enumerated in the forced migration survey aged 15 years or more, by employment status during the week preceding the survey, according to sex and country of origin

Table 6 Year of arrival in host country

Percent distribution of forced migrants by year of arrival in host country, according to country of origin

B. Tabulations of data gathered in the individual interview

Table 7 Age-sex distribution of forced migrants in the individual survey

Percent distribution of forced migrants selected for the individual interview, by age, according to sex

Table 8 Main reason for leaving country of origin

Percent distribution of forced migrants by the main reason for leaving country of origin for the first time

Table 9 Family members who accompanied forced migrants

Percentage of forced migrants who were accompanied by family members or relatives when leaving country origin for the first time

Table 10 Migratory route decision-making

Percentage of forced migrants who reported reasons specified for choice of migratory route

when they left their country of origin

Table 11 The journey to host country

Percent distribution of forced migrants by the number of countries visited before arriving in host country, according to country of origin

Table 12 Reason of moving onward from first country of asylum

Among forced migrants who arrived in current host country via one or more other countries, the percentage who reported reasons specified for moving onwards from the first country of asylum

Table 13 Difficulties encountered during migration journey

Among forced migrants who were confronted with difficulties during the migration journey, the percentage reporting specified type of difficulties encountered, according to country of origin

Table 14 Financing the migration journey

Percentage of forced migrants who reported specified sources of financing their journey from country of origin

Table 15 Main reason for coming to host country

Percent distribution of forced migrants by the main reason for coming to current host country

Table 16 Asylum applications

Percentage of forced migrants who applied for asylum, according to country of origin, and percent distribution of applicants for asylum by source of assistance

Table 17 Refugee status determination

Percent distribution of asylum applicants by refugee status determination, according to country of origin

Table 18 Identity documents

Percentage of forced migrants by type of identity documents they have in host country, according to country of origin

Table 19 Assistance received from any source in host country

Percentage of refugees who received specified types of assistance from persons or organizations in host country

Table 20 Work status

Percent distribution of forced migrants aged 15 years or more by current work status, according to country of origin

Table 21 Reason for not seeking work

Among forced migrants who were not working and not seeking work, the percentage who cited specified reasons for not looking for work

Table 22 Refugees plans for the future

Percent distribution of forced migrants aged 15 years or more by plans for the future, according to country of origin

Table 23 Conditions to move back to country of origin

Among forced migrants who reported planning to move back home under certain conditions, the percentage who reported specified conditions for returning

Table 24 Advice to relatives back home regarding moving abroad

Percent distribution of forced migrants by advice they would give to relatives and friends back in countries of origin about moving abroad

C. General health status tables

Table 25 Overall smoking tobacco status

Among persons aged 15 years and over, who were enumerated in the household survey, the percentage who ever smoked any tobacco product, the percentage who have stopped smoking any tobacco product, the percentage who currently smoke any tobacco product, the percentage who currently smoke cigarettes, and the percentage who currently smoke water pipe

Table 26 Age patterns of smoking tobacco

Among men aged 15 years and over, who were enumerated in the household survey, the percentage who ever smoked any tobacco product, the percentage who have stopped smoking tobacco products, the percentage who currently smoke any tobacco product, the percentage who currently smoke cigarettes, and the percentage who currently smoke water pipe

Table 27 Smoking tobacco status based on the individual survey

Among forced migrant men aged 15 years and over, the percentage who ever smoked any tobacco product, the percentage who have stopped smoking tobacco, the percentage who currently smoke any tobacco product, the percentage who currently smoke cigarettes, and the percentage who currently smoke water pipe, according to country of origin

Table 28 Age at starting smoking and number of cigarettes smoked per day

Among forced migrant men aged 15 years and over who currently smoke cigarettes, the percent distribution by number of cigarettes smoked per day, according to country of origin

Table 29 Passive smoking

Among male forced migrants aged 15 years and over, the percentage reporting that other persons had smoked in their home in their presence in the past seven days

Table 30 Health care

Percent distribution of forced migrants by type of health care received in past month, according to age and gender

b. SESRIC

OIC Statistical Capacity Building (StatCaB) Programme for NSOs by SESRIC

The member countries of the Organisation of Islamic Cooperation (OIC) have varying levels of capacities for producing statistical data in accordance with the relevant internationally agreed standards. The Statistical, Economic and Social Research and Training Centre for Islamic Countries (SESRIC), a subsidiary organ of the OIC, carries out a facilitating role in strengthening and improving the National Statistical Systems (NSSs) of OIC countries by contributing to their efforts in producing high-quality statistics and thus facilitating evidence-based policymaking.

In this context, SESRIC carries out a flagship statistical skills development initiative titled “OIC Statistical Capacity Building Programme (StatCaB)” which identifies statistical capacity needs of constituents within the NSS, particularly National Statistical Offices (NSOs). Based on these identified needs and capacities, a matching is done among OIC country experts to organise short-term statistics courses, study visits, and technical missions. In its purest form, these statistical activities are built on a South-South Cooperation approach.

Concerning the topic of refugees and internally displaced persons (IDP), eleven statistics courses and study visits have been organised by SESRIC focusing on Census Survey Design and Sampling Techniques (1), Human Settlements and Housing Statistics (1), Population and Demography (2), Population and Housing Censuses (4), Population and Migration Statistics (3) among others.

In addition to capacity building, SESRIC also facilitates the development of knowledge platforms and networks in the area of data and statistics. With reference to the UNSC Work Programme on Refugee and IDP Statistics, the Seventh Session of OIC Statistical Commission was held in May 2018 under the theme ‘Enhancing the Production, Dissemination, and the Use of Migration Statistics’. As a member of EGRIS, SESRIC hosted the Third Global Meeting of EGRIS at its Headquarters in Ankara, Turkey between 12 and 14 February 2019. Additionally, a meeting on Stateless Persons Statistics was hosted on 15 February 2019, in the margins of the 10th Regional Workshop on the Households International Migration Surveys in the Mediterranean countries between 14 and 16 February 2019 at the SESRIC Headquarters.

c. Harmonising Refugee Statistics in the European Union

714. In the European Union (EU), asylum and migration have become matters of Community responsibility on 1 May 1999 when the Treaty of Amsterdam entered into force. It became clear that developing, applying, monitoring and assessing a common asylum and migration policy require comprehensive and comparable data, collected and processed in accordance with a harmonised statistical methodology. It was established that there is a need to improve the quality of Community statistical collections and outputs, and the exchange and analysis of statistical information on asylum and migration.

715. At the time, there were considerable differences between Member States in terms of statistical information collected and produced, the existing national statistical methodologies and legal definitions of which persons are defined as migrants, asylum seekers or refugees. Therefore, extensive consultations followed with data suppliers and users to harmonise data collection. Based on the consultations, the Commission made a formal proposal for a framework regulation in 2005. The objective of the regulation was to establish a common framework for the collection and compilation of EU statistics on asylum and international migration. The proposal specified the data to be collected, the timetables to be applied, the definitions and the quality standards.

716. In order to convince Member States of the need for better, more detailed, more frequent and timelier statistics, it was important to underline that not only EU institutions would benefit from it but also Member States who could rely on these statistics when developing their policies and evaluating cross-country outcomes. Other users and the public would also gain from increased transparency in this area. In a policy area, as contested as asylum and migration policy, transparency is a particularly important factor, also in view of creating the necessary level of trust and confidence of the public in the actions and measures taken by the responsible authorities. The co-legislators, the European Parliament and the Council agreed on the final text of the regulation in 2007, resulting in Regulation 862/2007 on Community statistics on migration and international protection. The Regulation defines a 'refugee' in accordance with the 1951 Refugee Convention done at Geneva.

Harmonising Refugee Statistics in the European Union – Action Plan for Asylum and Migrant Statistics

In 2003, the European Commission presented an Action Plan, aimed at the development and improvement of Community statistics and their analysis in the field of asylum and migration. The document proposed a discussion about the form and main principles of possible future statistical legislation, and activities to enhance information exchange and promote decision-making. It emphasized that the success of these measures is, to a large part, dependent on the active co-operation of the respective authorities in the Member States: where national authorities fail to supply the relevant statistical data, the Commission is unable to produce statistics to the appropriate levels of timeliness and quality. The Action Plan acknowledged that better socio-economic information is needed for monitoring the integration of migrants and refugees and when forecasting future developments in the labour market. The collection of accurate and timely statistics can also improve the ability of the EU to concentrate its efforts where it is most needed and facilitate the allocation of resources.

In drafting the Action Plan, the Commission has been guided by the different users' needs, including EU institutions, governments and other Member State authorities, social and economic operators, academia and the public. Additionally, the Action Plan incorporated the principle of transparency: EU statistics are made publicly available on an online statistical dissemination database. Exceptions to this principle are made to protect the confidentiality of the individual, to avoid the disclosure of individual identity information. As an additional safeguard to protect individual confidentiality, asylum and migration statistics are rounded to the nearest five before publication.

d. UNHCR Agreements and Memoranda of Understanding for the transfer of data to Governments

717. The agreement template below is an example of standard agreement between the UNHCR and a national government to enable the transfer of data. The drafting notes included in the footnotes are included to provide guidance to readers on the preparation of the document. More information can be obtained from UNHCR.

**Agreement on the Transfer of Personal Data
of Refugees and Asylum-seekers**

(hereinafter referred to as this "Agreement")

between

The Government of {country name}⁶³

(hereinafter referred to as the "Government")

and

The Office of the United Nations High Commissioner for Refugees

(hereinafter referred to as "UNHCR")

Preamble

[Reaffirming the commitment of the Government to implement its obligations under the Convention relating to the Status of Refugees of 1951, the Protocol relating to the Status of Refugees of 1967 *{and applicable regional refugee instruments}*];⁶⁴

[Considering {the national law/decreree relating to the status of refugees and asylum-seekers}];⁶⁵

Recalling Conclusion No. 8 (XXVIII) – 1977 on the Determination of Refugee Status of the Executive Committee of the Programme of the United Nations High Commissioner for Refugees and Conclusion No. 91 (LII) – 2001 on the Registration of Refugees and Asylum-Seekers;

Acknowledging the importance of registration as a protection tool, including protection against *refoulement*, protection against forcible recruitment, protection of access to basic rights, family

⁶³**[The drafting notes in this document are for UNHCR guidance only and must be deleted prior to sharing the document externally. The green text highlights paragraphs that each operation is expected to customize at country level, in close cooperation with the relevant government counterpart.]** DRAFTING NOTE: The text highlighted in green indicates text which the UNHCR country office is expected to customize to their operation. If changes are made to the agreement beyond the text highlighted in green, the UNHCR country operation is requested to do so in track changes and to bring these changes to the attention of LAS and DIP when seeking clearance of the data sharing agreement, in accordance with para 6.2.3 of the UNHCR's Data Protection Policy.

⁶⁴ DRAFTING NOTE: Include this clause if applicable.

⁶⁵ DRAFTING NOTE: Include this clause if applicable.

reunification of refugees and identification of those in need of special assistance, and as a means to enable the quantification and assessment of needs and to implement appropriate durable solutions;

Considering the responsibility of States to take all necessary measures to register and document refugees and asylum-seekers on their territory as quickly as possible after their arrival, bearing in mind the resources available, and where appropriate to seek the support and cooperation of UNHCR;

Acknowledging the importance of establishing procedures for Refugee Status Determination and the value of UNHCR's participation in such procedures;

Recognizing that where UNHCR conducts registration and/or refugee status determination of refugees and asylum-seekers, the transferring of the personal data of refugees and asylum-seekers to the government of their host country needs to serve the purpose of enhancing their protection;

Recognizing, in this context, the right of every person under international human rights instruments such as the Universal Declaration of Human Rights and the International Covenant of Civil and Political Rights to be protected against the arbitrary or unlawful interference with his or her privacy and other fundamental principles for the protection of personal data. This includes to ensure that personal data is obtained and processed only in fair and lawful ways; that the purpose of the data processing is specified, legitimate and brought to the attention of the persons concerned; that the data is adequate, relevant and not excessive in relation to the purpose for which it is processed as well as accurate and, where necessary, kept up to date;

Recognizing, therefore, that the systematic sharing, electronically or otherwise, of personal data of refugees and asylum-seekers by UNHCR with States needs to be subject to data protection safeguards as contained in UNHCR's Policy on the Protection of Personal Data of Persons of Concern;

Acknowledging, furthermore, that the transfer of personal data must not compromise UNHCR's humanitarian and non-political character, jeopardize human rights or undermine the climate of trust and confidence which needs to exist between UNHCR and persons approaching it for protection and assistance;

[*Acknowledging {reference to applicable national privacy/data protection legislation};*]⁶⁶

[*Acknowledging and without prejudice to {Accord de Siège between UNHCR and the Government of the host country};*]⁶⁷

The Government and UNHCR (hereinafter referred to as the "Parties") agree as follows:

Article 1 – Definitions

For the purpose of this Agreement, the following definitions shall apply, unless the context otherwise requires:

⁶⁶ DRAFTING NOTE: Include this clause if applicable.

⁶⁷ DRAFTING NOTE: Include this clause if applicable.

- (a) “country” means *{country name}*.
- (b) “consent” means any freely given and informed indication of an agreement by the data subject to the processing of his/her personal data, which may be given either by a written or oral statement or by a clear affirmative action.
- (c) “data subject” means any refugee or asylum-seeker, whose personal data is subject to processing in the context of a registration and/or refugee status determination procedure, as well as any persons to whom the personal data refers.
- (d) “personal data” means any data related to an individual who can be identified from that data; from that data and other information; or by means reasonably likely to be used related to that data. Personal data includes biographical data (“biodata”) such as name, sex, marital status, date and place of birth, country of origin, country of asylum, individual registration number, occupation, religion and ethnicity, biometric data (such as a photograph, fingerprint, facial or iris image), as well as any expression of opinion about the individual, such as assessments of the status and/or specific needs.
- (e) “processing” means any operation, or set of operations, automated or not, which is performed on personal data, including but not limited to the collection, recording, organization, structuring, storage, adaption or alteration, retrieval, consultation, use, transfer (whether in computerized, oral or written form), dissemination or otherwise making available, correction, or destruction.
- (f) “refugee status determination” means the legal and administrative procedure undertaken by States and/or UNHCR to determine whether an individual should be recognized as a refugee in accordance with national and international law.
- (g) “registration” means the recording, verifying, and updating of information on refugees, asylum seekers and other persons of concern to UNHCR with the aim of protecting and documenting them and of implementing durable solutions.

Article 2 – Object and purpose

1. This Agreement regulates the transfer of personal data of refugees and asylum-seekers, generated by UNHCR through a registration and/or refugee status determination procedure, to the Government.
2. The overall purpose of the transfer of personal data pursuant to Article 3 and, if applicable, Article 4 of this Agreement, is to increase the protection of refugees and asylum-seekers in the country.

Article 3 – Personal data to be transferred^{68 69}

⁶⁸ DRAFTING NOTE: When determining which data sets to be transferred to the Government, please take note of the need for the transfer of data to fulfill a specific and legitimate purpose, be necessary and proportionate, and not exceed that purpose. While the listed data sets is what UNHCR would normally share with the host government, a data minimization review may be required to determine which data sets are shared in the specific operational context.

⁶⁹ DRAFTING NOTE: When concluding this Agreement, issues of who manages and regulates access rights to the database, has authority to rectify or erase data, etc., are crucial. If needed, please seek advice from the IMRS

1. UNHCR may transfer to the Government the following personal data (“basic bio-data”) of each refugee and asylum-seeker registered by UNHCR in the country and whose record status is active:

- (a) Name;
- (b) Date of arrival in the country;
- (c) Registration date;
- (d) Citizenship / Country of origin / Country of birth;⁷⁰
- (e) Place of birth;
- (f) Date of birth;
- (g) Sex;
- (h) Photograph;
- (i) Status under UNHCR’s Mandate (“refugee”, “asylum-seeker” or “not in need of international protection as a refugee”);
- (j) Current address [administrative level 1 or camp level only];⁷¹

2. UNHCR will make the appropriate arrangements to ensure that all data subjects are duly informed prior to {registration and/or refugee status determination}⁷² that the above information will be transferred to the Government.

Article 4 – Requests for transfer of additional personal data

1. Requests by the Government for the transfer by UNHCR of personal data of the refugees and asylum-seekers referred to Article 3(1) beyond the basic bio-data (“additional personal data”) shall be considered by UNHCR on a case-by-case basis, taking into account UNHCR’s mandate and its Policy on the Protection of Personal Data of Persons of Concern, the privileges and immunities of the United Nations, and the need to ensure the safety of UNHCR staff and other humanitarian personnel.

2. Any request for the transfer of additional personal data under this Article must be made in writing and must clearly state the reasons why such transfer is necessary and the purpose for which the requested personal data will be [used/processed]. Consent for disclosure shall be requested as necessary from the data subject with respect to whom the request relates.

3. In the event UNHCR agrees to a request for transfer of additional personal data, such transfer shall be governed by this Agreement, subject to any modifications or additions required or

and the DIP Data Protection Officer on the details of these arrangements.

⁷⁰ DRAFTING NOTE: Delete whichever is inapplicable.

⁷¹ DRAFTING NOTE: Delete if not appropriate.

⁷² DRAFTING NOTE: Delete one, if not relevant.

agreed by UNHCR.

Article 5 – Means by which personal data are transferred

1. The personal data will be transferred from UNHCR in {hardcopy and/or electronic format and/or by access to a UNHCR-owned database}.⁷³ The electronic format may be a database application, Extensible Markup Language (XML), Portable Document Format (PDF), a spreadsheet, or a word-processing document. If the personal data is transferred by means of the Government's access to proGres or another database owned by UNHCR, the following modalities and access limitations shall apply: [_____].⁷⁴

2. The means of transferring additional personal data under Article 4 of this Agreement shall be determined by UNHCR.

Article 6 – Specific purposes of personal data transfer⁷⁵

1. The Government may process the basic bio-data transferred in accordance with Article 3 of this Agreement only insofar as is necessary to achieve the following purposes:

- (a) Identifying persons who are in need of protection as refugees, as well as persons who are not in need, or no longer in need, of international protection;
- (b) Carrying out national immigration formalities, with due regard to the protection needs of the persons concerned;
- (c) Ensuring respect of the principle of *non-refoulement* and the general treatment of refugees and asylum-seekers in accordance with national and international legal obligations;
- (d) Issuing personal documents, such as documents necessary to prove identity and refugee status, to obtain access to assistance and services, and/or to exercise basic rights; travel documents; and certifications of birth, marriage, divorce and/or death;
- (e) Facilitating family reunification;
- (f) Implementing appropriate durable solutions, such as voluntary repatriation, local integration or resettlement to a third country;
- (g) Compiling statistical data.

2. Any additional personal data transferred in accordance with Article 4 of this Agreement may be used solely for the purposes specified in Article 6(1) or for any other purposes agreed in writing by UNHCR.

⁷³ DRAFTING NOTE: Please select what is agreed upon in the operation, including specific data transfer modalities, taking into consideration the safety of the data, the operational requirements, and the technology and resources available.

⁷⁴ DRAFTING NOTE: Please delete this sentence if no such access is granted. If access is granted, please specify the access modalities and limitations.

⁷⁵ DRAFTING NOTE: Please take note of the need for the transfer of data to fulfil a specific and legitimate purpose, in accordance with UNHCR's mandate, and by be necessary and proportionate to that purpose.

3. Personal data transferred under this Agreement may not be used by the Government for any purpose other than that specified in Article 6(1) and Article 6(2) except with the prior written authorization of UNHCR and/or, as appropriate, the consent of the data subject.

4. The Government shall take all organizational and technical measures to ensure that the personal data is processed only by its authorized personnel and institutions, whose use and access is required to achieve the purposes specified in Article 6(1) or agreed by UNHCR under Article 6(2).

Article 7 – Transferring data to third parties

1. Personal data transferred under this Agreement may not be transferred or otherwise disclosed by the Government to a third party except with the prior written authorization of UNHCR and/or, as appropriate, the consent of the data subject. Prior to any such authorized transfer or disclosure, the Government will enter into a binding agreement in which the third party agrees to the obligations undertaken by this Government in this Agreement, including (a) the using/processing of the personal data solely for the purposes described in Article 2(2) and Article 6(1) of this Agreement and (b) the prohibition on transferring or disclosing the personal data.

2. Under no circumstances shall personal data transferred under this Agreement be disclosed to the country of origin of the data subjects. The only exception to this concerns data processed in the context of a tri-partite agreement for voluntary repatriation under the auspices of UNHCR, conducted with the consent of the individual data subject(s).

Article 8 — Security of personal data

Both parties shall take organizational and technical measures to protect the personal data transferred under this Agreement against accidental or unauthorized destruction, accidental loss, unauthorized access, use, alteration or dissemination, and against all other unauthorized forms of processing.

Article 9 – Notification of personal data breach

1. The Government shall notify to UNHCR as soon as possible upon becoming aware of a personal data breach (as defined below), in particular if the personal data breach is likely to result in personal injury or harm to a data subject, and use, if appropriate jointly with UNHCR, its best efforts to take mitigating measures.

2. “Personal data breach” means a breach of data security leading to the accidental or unlawful/illegitimate destruction, loss, alteration, unauthorized disclosure of, or access to, personal data transferred, stored or otherwise processed.

Article 10 – Settlement of disputes

Any dispute between UNHCR and the Government arising out of or relating to this Agreement shall be settled amicably by negotiation or other agreed non-judicial mode of settlement (including arbitration).

Article 11 – Privileges and Immunities

Nothing in or relating to this Agreement is to be deemed a waiver, express or implied, of any privileges or immunities of the United Nations or of UNHCR, as a subsidiary organ of the United Nations.

Article 13 – General provisions

1. This Agreement shall enter into force on the date of its signature by both Parties and shall continue in force until terminated under Article 12(2).
2. This Agreement shall cease to be in force 30 days after either of the Parties gives notice in writing to the other Party of its decision to terminate this Agreement, except as regards the normal cessation of the activities of UNHCR in the country. The obligations under Articles 6, 7, 8 and 9 of this Agreement shall not cease with the termination of this Agreement.
3. This Agreement may be modified at any time by mutual written consent.
4. This Agreement is not intended to benefit or be enforceable by any third party.

IN WITNESS WHEREOF, the undersigned, being duly appointed representatives of UNHCR and the Government, respectively, have on behalf of the Parties signed this Agreement, in the English and *{country name}* language(s). For purposes of interpretation and in case of conflict, the English text shall prevail.

Done at {location} this {xx} day of {month} {year}

For the Government of
{country name}

{name of Government representative}
{title of Government representative}

For the Office of the United Nations
High Commissioner for Refugees

{name of UNHCR representative}
{title of UNHCR representative}

**PART III – ADDITIONAL TECHNICAL
MATERIAL**

PART III – PHASE 1A: Development of a stakeholder plan for displacement statistics

a. Roles and Questions

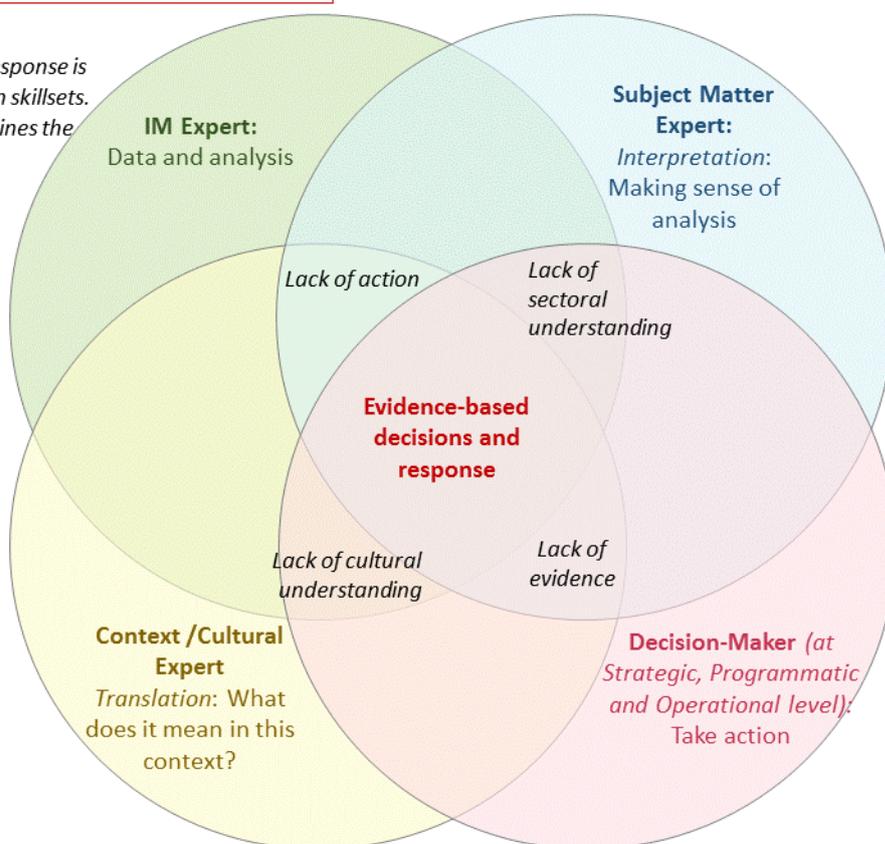
718. Four roles are needed for evidence-based decision-making: Decision-Makers, Sectoral Experts, Context/Cultural experts, Statistical Experts. Skillsets of each one and consequences of excluding one of these roles from the process is illustrated below.

Evidence-based decisions and response – Venn Diagram

Evidence – based decision making and response is generated by the interaction of four main skillsets. Excluding even only one of them undermines the veracity and usefulness of results.

The modalities and process of their interaction must be rigorous and predictable, for results to be useful and usable

- Discounting Information Management (IM) skillset results in lack of evidence.
- Discounting the subject matter or the cultural expertise prevents make sense of information (for a specific sector and in a specific culture/context).
- Excluding decision-makers results in lack of appropriate action.



Modified by EDAUUR - Grand Bargain Work Stream for Needs Assessment

719. 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:

1. Define Specific Information Needs,
2. Methodology Design,
3. Tools and Data Analysis Plan Design,
4. Analysis considered as one 'composite' step.

b. Define Specific Information needs: Roles & Responsibilities

720. **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.

721. **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.

722. **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.

723. **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.

724. 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):

- i. What is it that you have to do? (Decision Makers)
- ii. What are the decisions you need to make? (Decision Makers)
- iii. What information do you miss in order to make that decision? (Decision Makers)
- iv. How often should that information be updated, at a minimum, to be still usable? (Decision Makers, Sectoral Experts and Context Experts)
- v. Is that information already available/ accessible? (Sectoral Experts, Context Experts and Statisticians)
- vi. How will this information help in the decision-making (*What are logical flow & benchmarks*)? (Decision Makers)
- vii. 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)
- viii. Are any of these data already available/accessible? (Sectoral Experts, Context Experts and Statisticians)
- ix. What are the missing data we need to collect? (Sectoral Experts, Context Experts and Statisticians).

c. Methodology design: Roles & Responsibilities

725. **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.

726. **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.

727. **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.

728. **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):

729. **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

730. 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.

731. **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):

1. 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)
2. 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)
3. 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)
4. Questions and answers are included in a data plan, which may also include dissemination modality, source of data and more.
5. 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)
6. 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)

7. 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)
8. Assessment tool is designed, according to best practices (Statistician).

e. Analysis: Roles and Responsibilities

732. 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.

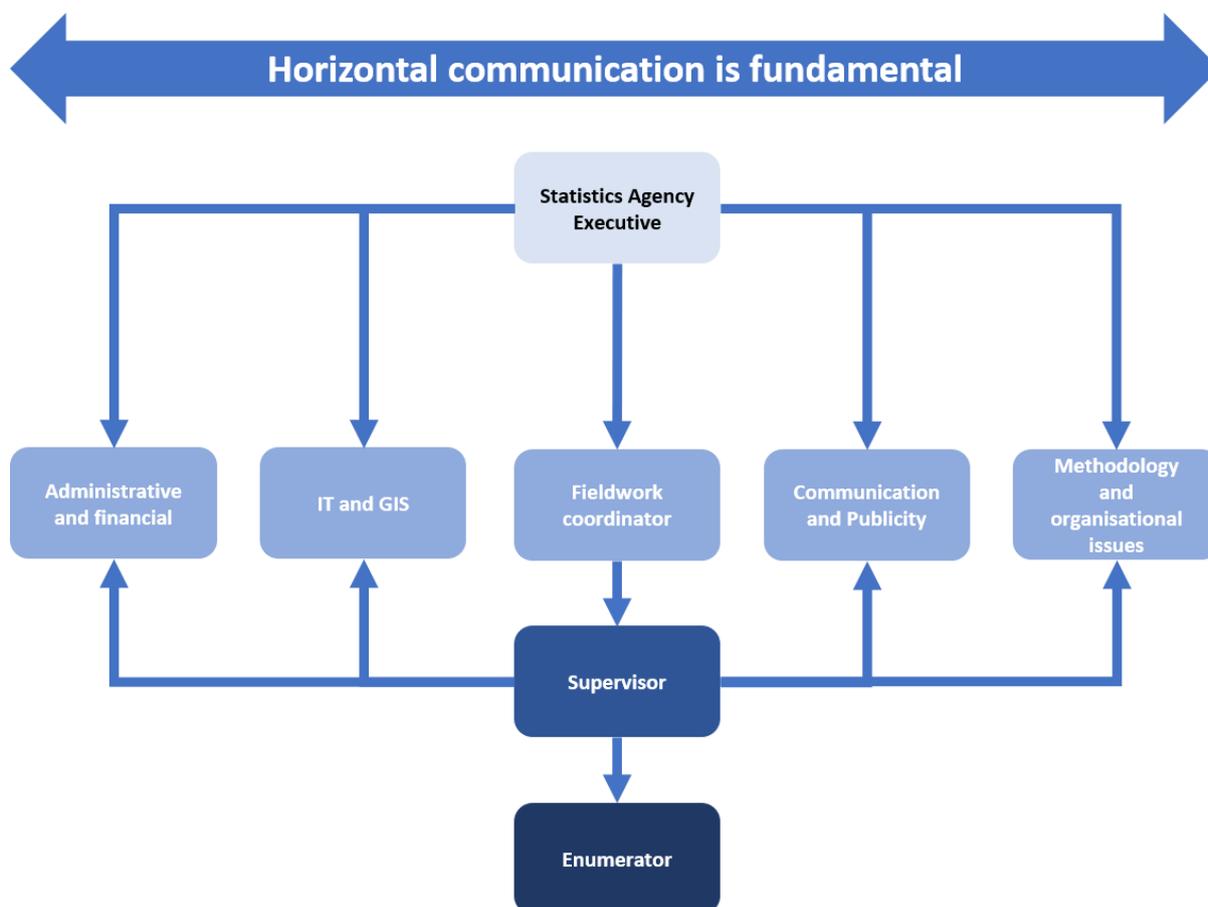
- **Descriptive analysis:** Describing data means to summarize and reduce large amount of data to a representation where it is easier to compare between them and identify the main points, important stories and relevant messages, e.g. a percentage, average, mean, mode, etc.
- **Explanatory analysis** 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)
- **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

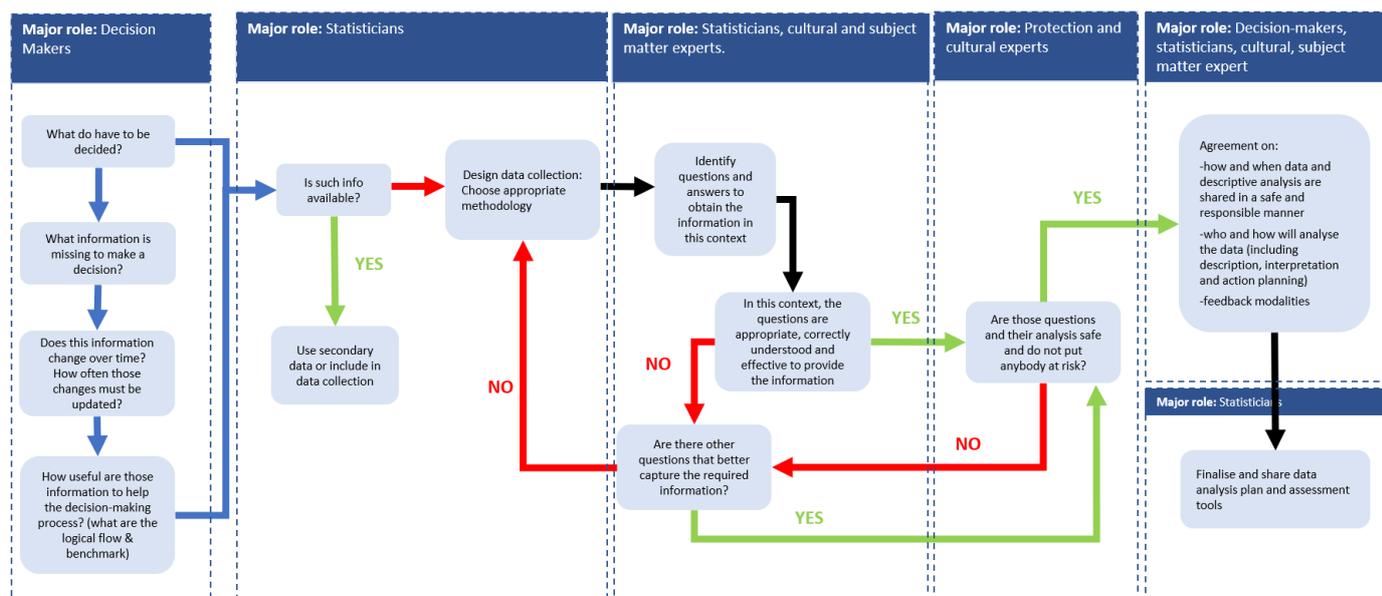
Makers, with support by Context- and Subject-matter experts)

733. 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 diagram below.



734. 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.

735. An activity tracking system is desirable. A Gantt chart similar to the illustration below 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.



Useful references and Links:

[1] Resolution of the data may be villages, sub-districts, districts, etc.). See: Aldo Benini, Documenting methods and data in rapid needs assessments, ACAPS, 2012.

[3] Useful tools to jointly identify most appropriate system for data collection include: PIM matrix, that can be adapted to other sectors (http://pim.guide/wp-content/uploads/2018/04/Quick-Reference-Flyer_Principles_Matrix_Process_2018-1.pdf); ACAPS “Direct Observation and Key Informant Interview Techniques for primary data collection during rapid assessments”: <https://www.acaps.org/direct-observation-and-key-informant-interview-techniques>; UNHCR overview of data collection methods and their use in UNHCR Needs Assessment Handbook, p 81: <http://needsassessment.unhcr.org/>

[4] Aldo Benini, Documenting methods and data in rapid needs assessments, 2012.

[5] For example, see ACAPS brief on questionnaire design: https://www.acaps.org/sites/acaps/files/resources/files/acaps_technical_brief_questionnaire_design_july_2016_0.pdf

[6] https://www.acaps.org/sites/acaps/files/resources/files/acaps_technical_brief_questionnaire_design_july_2016_0.pdf

PART III – PHASE 1 IDENTIFY NEEDS USING THE ADAPT TOOL

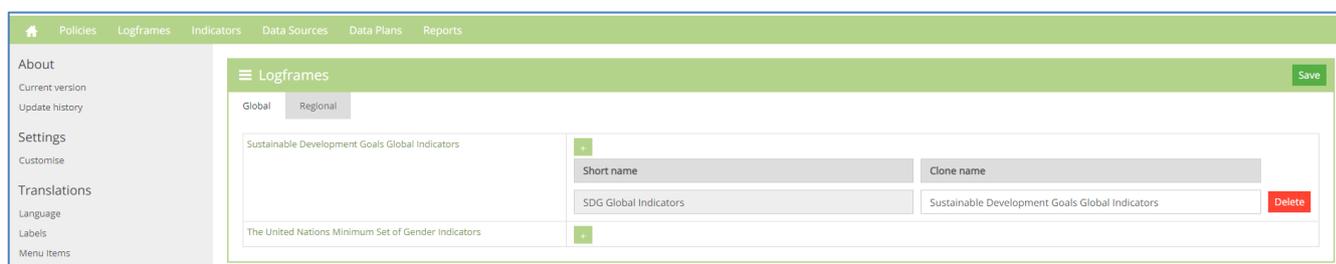
736. Use ADAPT Tool <https://adapt.paris21.org/auth/login>

a. Option 1. Cloning established indicator frameworks (logframes)

737. A list of indicators (from for example a log frame) that is relevant to refugee and IDP priorities in the local context is input into the tool. This list serves as the NSS's framework for reporting on displacement statistics, or the displacement statistics domain. As a next step, the pre-identified indicators that are available from national sources are loaded into the tool, which maps them against the log frame's requirements. A short description is added together with relevant information on applicability, availability, compliance, feasibility and overlaps with the log frames of other domains. The proposed indicator framework for basic statistics and tabulations in the IRRS and IRIS can be found in the ADAPT Library and can be cloned for use of any country.

738. If the NSO is interested in reporting on a national list of refugee and IDP-related indicators, ADAPT offers to upload customized log frames as outlined in the [ADAPT Manual](#) (Section 4).

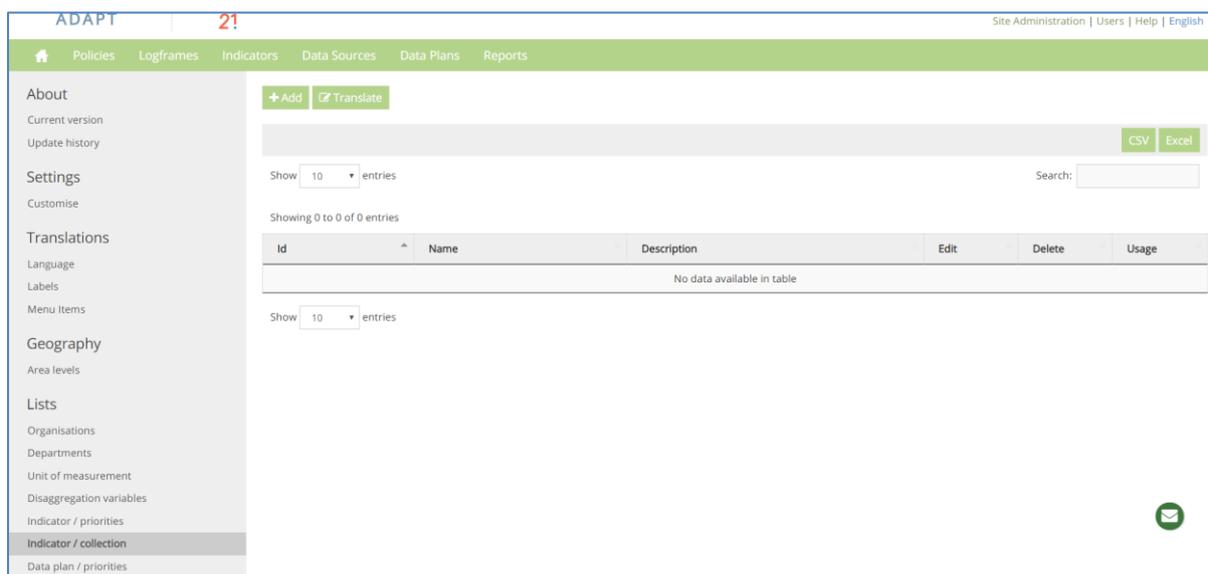
Table 0.1: How to add the IRRS or IRIS basic indicator framework as logframe in ADAPT



b. Option 2. Flagging indicators by collection themes (tags)

739. Indicators uploaded to ADAPT can be classified into themes (tags), such as refugee or IDP statistics. Collection themes can be generated manually, but one indicator can only be flagged using one tag. Flagging indicators by collection theme ensures that all indicators available have been recognized in the assessment of existent data gaps and in the reporting on refugee and IDP statistics.

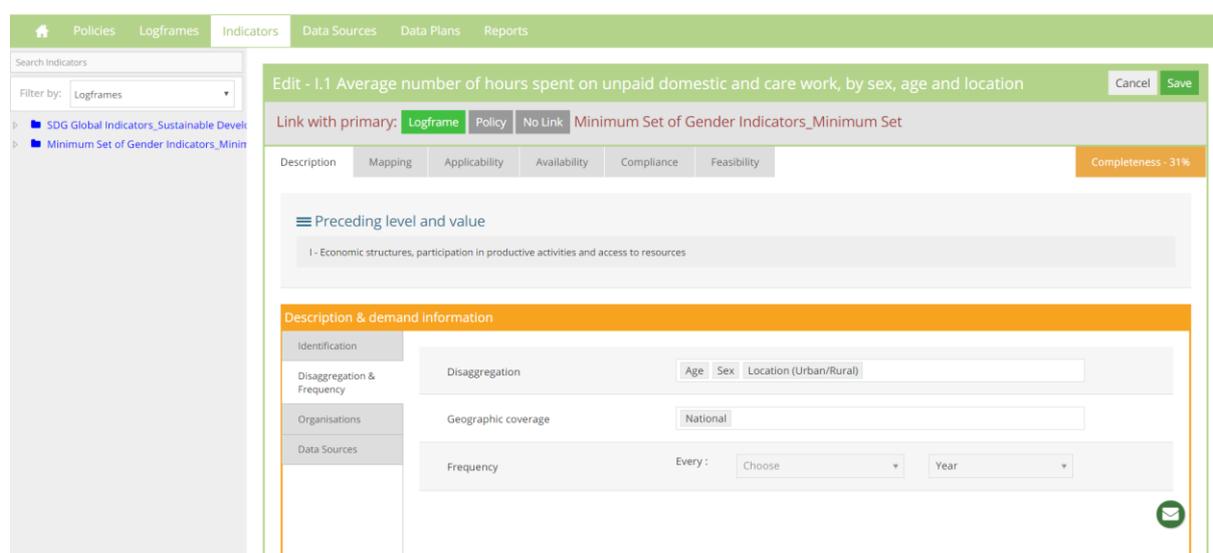
Table 0.2: How to create collection themes (tags) with a statistical domain in ADAPT



c. Option 3. Identification of indicators by disaggregation variable

740. ADAPT allows the users of the tool to note indicators where at least one disaggregation variable is related to refugee or IDP statistics, for example legal residential/international protection status. Disaggregation variables NSOs can be manually defined based on needs. Currently, ADAPT offers some disaggregation variables, such as age, sex and location, by default.

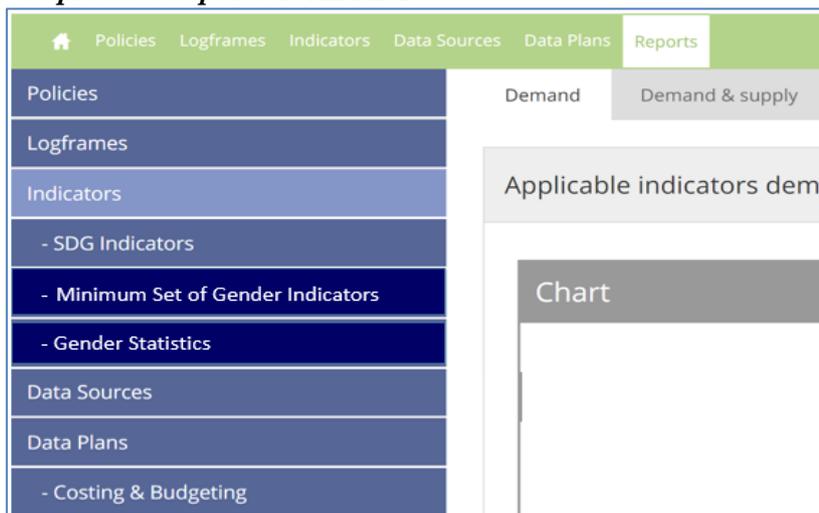
Table 0.3: How to add disaggregation variables to indicators in ADAPT



d. Refugee and IDP statistics output in ADAPT

741. Based on the information provided using the options above, the ADAPT tool produces two main reports on the status of refugee and IDP statistics, and identifies any data gaps at national level. The first report summarizes the status of refugee and IDP statistics in respect of the relevant national logframe, which could be based on the IRRS or IRIS tabulation indicator frameworks. 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 and that have been flagged using the defined tag.

Table 0.4: How to produce reports in ADAPT



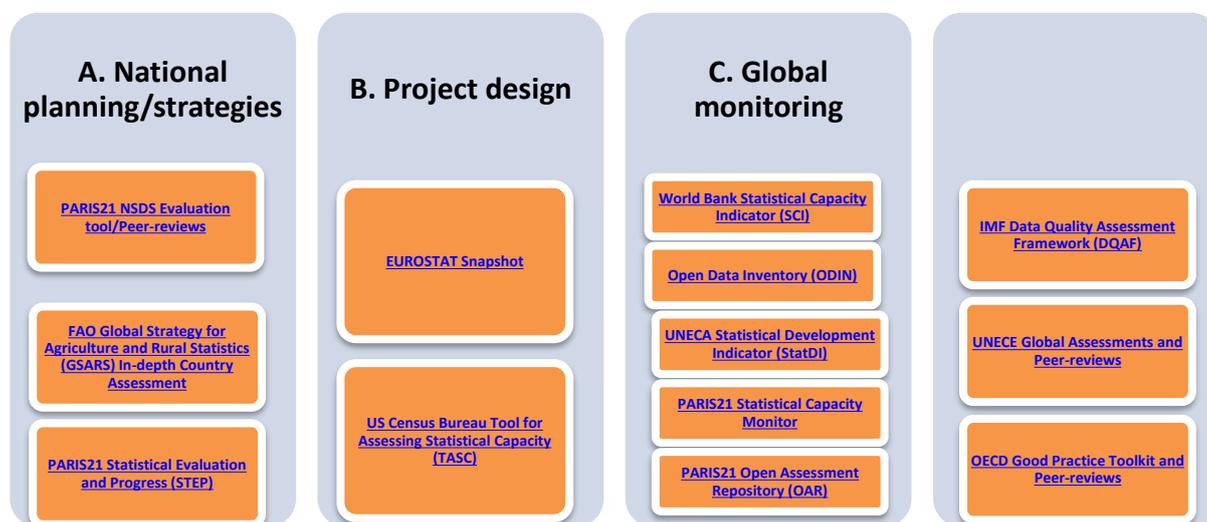
PART III – PHASE 1 NSS ASSESSMENT TOOLS

742. The resources and capacity requirements of establishing a new statistical series will need planning for, both nationally in statistical plans and budgets, and internationally in terms of supporting the funding and capacity needs for initiating a new product. The international community has historically heavily supported new statistical topics, and external funding may be needed for new statistics about displacement. Funding, from government or external partners will be needed and additional technical support should be considered. This would be part of an NSS assessment.

743. PARIS21 has identified 13 different NSS assessment tools which were categorized into four different types:

- A. Tools used by national authorities to inform national statistical planning/ strategies (NSDS),
- B. Tools for partners who want to “invest” in statistics to inform project design and monitoring,
- C. Tools for monitoring statistical performance internationally – global public goods, and
- D. Tools to assess quality and/or compliance with statistical codes of practice (see Figure 2).

Each assessment tool has been developed by a different agency (although some agencies have more than one tool, designed for specific contexts or needs) and each uses a different method for assessing country systems (PARIS21, 2018).



744. Part of the assessment would check whether any current data sources could meet user requirements, and establish the conditions under which they would be available, including any restrictions on their use. Any assessment of potential administrative or other non-statistical data sources, will use the following criteria to determine whether they would be suitable for use for statistical purposes.

- Define what data is needed and the preferred source,

- Availability of data sources in the country or collected by international agencies, to establish which source is best (e.g. specific registers maintained by administrations or organisations responsible for refugees/IDPs, databases maintained by humanitarian organisations supporting refugees/IDPs, general population registers, residence permit registers, tax registers, social security registers, register of border crossings),
- Specify what quality improvements are required when moving from operational data to qualify the data source as official statistics, or when including operational data as a component of a new statistical series. (see section 2.6 or data quality at design and 6.2 validate outputs page) see p63-65 of IRRS,
- Budgetary and capacity constraints,
- Identify any data sources (traditional, GIS or big data) to be used as possible sampling frames as this determines the subsequent phases for surveys.
- An assessment of the legal framework in which data would be collected or accessed, are any changes to existing legislation needed? This may need to be domain specific – i.e. IDP and refugees may be different cases. This will be picked up in more detail in later phases.
 - Data sharing and coordination arrangements.
 - Pay special attention to data protection issues:
 - Burden on respondents,
 - Sensitivities around data collection and management,
 - Being unable to return data to its source agency after processing for statistics (statistics laws),
 - Pay attention to the “do no harm” principle (see PIM principles here).

PART III –PHASE 2: SAMPLING METHODOLOGY

745. This is material which is more general than that included in Part I, and is not specific to refugees and IDPs. It is included for readers who would like to know more about sampling and should be read in conjunction with Part I.

2.4.2 Representativeness of the sample

746. **Representative sample:** To generalise the sample result to the population, we need to ensure the sample is representative of the population, i.e. each interviewed unit represents a certain number of similar units in the target population. In order to get representative samples, we will need to draw them randomly from the population, or, at least, with an element of randomness

747. A **non-representative sample** is a sample in which the distribution of the characteristics of interest is systematically different from the population from which it has been drawn. It is based on human choice rather than random selection, and it does not allow for generalisations to the whole population.

6) 2.4.3 DEFINING THE TARGET POPULATION AND FINAL SAMPLING UNIT

748. **The target population** is the complete set of final sampling units which the survey aims to collect information about. **Final sampling units** are the elements of the population the information is coming from, such as households, individuals, or groups of individuals such as communities. It is important to properly define the final sampling unit since it will determine all other steps of the sampling process. For example, when sampling refugees, careful consideration needs to be given to the question whether to use socio-economically defined households or registration units. The choice of final sampling unit is determined by the study purpose and the type of information that will be collected. To define the final sampling unit, it can help to answer the question “who or where should the data come from to be able to calculate the indicator of interest?”.

7) 2.4.4 SAMPLING DESIGN

749. Sampling design describes how the final sampling units are selected and reached for the study. It is determined by the substantive goals of the study, and by resource and logistical constraints. There are different forms of designs in sampling.

750. **Simple Random Sampling (SRS)** is a method of selecting n units out of the N such that any unit has exactly the same probability of being drawn. Requires a complete list of all final sampling units as a sampling frame. Logistical and resource constraints in obtaining such a list will often make more complicated sampling methods necessary for IDP and refugee populations. This kind of design is rarely if ever used, except in situations where the population is small and located in a small discrete area, as it tends to be inefficient and logistically difficult to organise.

751. **Cluster sampling:** This is where the final sampling unit is grouped in geographical clusters. A sample is drawn of clusters (usually villages or enumeration areas) from all available clusters and then all or a subset of the sampling units in each selected cluster is interviewed. If all the sampling units of a cluster are selected, this is known as one-stage cluster sampling. If a random selection of the sampling units in each cluster is interviewed, it is referred

to as two-stage cluster sampling. In some cases, the first-stage clusters (PSUs) contain not the final sampling units, but another form of clusters as the second-stage clusters. This is a third-stage cluster sampling design. For example, the primary sampling units in an urban setting could be enumeration areas, of which a random selection is drawn. In a second step, the selected enumeration areas are divided into blocks, of which again a random selection is drawn for each enumeration area. In a final step, a simple random sample of households is drawn in each selected block. The clusters should always be non-overlapping and complete. Furthermore, it is recommended that clusters are mutually homogeneous, meaning that clusters resemble each other with regards to their internal composition of sampling units. The big advantage of cluster sampling is that a full list frame of all the final sampling units (e.g. all households in a country of location) is not required. Only lists of the final sampling units for the higher-level sampling units (e.g. city blocks) that were selected are needed. Even this requirement can be dropped if the final sampling units are randomly selected using techniques such as GPS sampling, the EPI method (spin the pen/bottle), or community-built sampling frames. Another advantage is that if the clusters are geographical, then the final sampling units in a cluster are geographically close. This reduces the need to visit many different locations, thereby bringing down travel costs and time. The main disadvantage is that the cluster design decreases the efficiency of the sample, as the sampling units in a cluster tend to be similar to one another. This means that a higher sample size is required compared to simple random sampling (see also design effect).

752. Respondent-driven sampling (RDS), combines 'snowball sampling' (getting individuals to refer those they know, these individuals in turn refer those they know and so on) with a mathematical model that weights the sample to compensate for the fact that the sample was collected in a non-random way.

8) 2.4.5 CONSTRUCTING A SAMPLING FRAME

753. A simple definition of a sampling frame (syn. sample frame) is the set of source materials from which the sample is selected. The definition also encompasses the purpose of sampling frames, which is to provide a means for choosing the members of the target population that are to be interviewed in the survey. The ideal sampling frame can be thought of as a list (directly or indirectly) that contains all final sampling units of the target population.

754. In addition to the final sampling units such as individuals or households, it will often be necessary to define higher-level sampling units which describe groups of final sampling units. For example, if a host population around a camp lives in villages, the list of villages can be used in constructing the sampling frame (see “multi-stage sampling frame” for specifics). Villages would be higher-level sampling units in this case. The highest geographic level of sampling units are called primary sampling units (PSUs).

Different types of sampling frames

755. *List sampling frame:* A list sampling frame is a frame made up of a list of the target population units. It differs from an area sampling frame in that the area frame does not list the individual units, but simply the geographic divisions they are contained in. A list frame lists each sampling unit individually and can be easily used to draw a random sample from. A list sampling frame can either be obtained from existing data sources such as a census, refugee registration data or administrative records, or it can be constructed by a new listing of all final sampling units of the target population. In the latter case, the listing exercise will in many cases

require significant resources and needs to be planned well ahead of the main study.⁷⁶

756. *Area sampling frame:* An area sampling frame comprises the geographical units of the area under study (e.g. a camp, a country, a programme area) in a hierarchical arrangement. For example, a list of the districts in a programme area is an area sampling frame, and a full list of enumeration areas covering a country is an areas sampling frame. An area sampling frame differs from a list sampling frame in that it does not list the final sampling units (e.g. households) individually, but instead lists the geographical units they are contained in. Area sampling frames are often used as primary sampling units in multi-stage and cluster sampling schemes.

757. *Multi-stage sampling frames:* It will often not be possible to obtain or construct a list of the final sampling units. For example, it can be practically impossible to obtain a list of all refugee households living out of camp due to the difficulties in finding and accessing refugee households that are interspersed with the main population. The only reliable way to list all refugee households in this case would be a full census and determination which are refugee of IDP households during the census, a prohibitively expensive exercise if such data is not already available and up to date.

758. In most cases a full and reliable list will not be available, in such cases the sample can be drawn in two or more stages by first dividing the study area on non-overlapping and complete geographical sub-units (PSUs) and then selecting a sub-sample of final study units within each sub-unit. In many cases, a random sub-sample of the PSUs themselves will be drawn to further limit the logistical and financial resources needed to reach final sampling units. The sub-samples within the PSUs themselves can either be sampled from a list frame. In most surveys which occur within official statistics a full listing procedure of each PSU precedes the selection of the final sample of households. This listing should be carried out just before data collection, to ensure that the lists are up to date. This is particularly important in situations of displacement where the population is highly mobile.

759. In some cases, non-list random sampling procedure such as random walk or satellite-data based approaches. The example of sampling refugees in Turkey (see Part II of this Manual *Turkey*) constitutes such an approach.

760. **Two or more sampling frames in the same study:** If more than one type of target population is being studied in the same assessment, it will often be necessary to use two or more sampling frames. For example, households in the host population and those in camps will need to be constructed differently (e.g. list sampling frame from registration data for camp population, multi-stage sampling with communities as PSUs for host population).

⁷⁶ A household listing can be difficult if the location being listed is located in a conflict area. The individuals doing the listing may only be able to work in the field for short periods of time; for example, only during daylight hours. The work may need a facilitator, such as a local NGO to help the listers gain access to the territory. Time in the field and the predictable movements by interviewers may increase their exposure to robbery, kidnapping, and assault, and increase the likelihood that the local group in charge of security may object to the presence of outsiders in the community. Alternative techniques such as satellite mapping, segmentation, grid / geosampling, Qibla method, or random walk may also be considered (Himelein et al, 2016).

Himelein, Kristen, Stephanie Eckman, Siobhan Murray, and Johannes Bauer. 2016. "Second-Stage Sampling for Conflict Areas: Methods and Implications". Policy Research Working Paper 7617. World Bank, Poverty and Equity Global Practice Group, Washington, DC.

9) 2.4.6 DETERMINING SAMPLE SIZE

The results of sample surveys are always subject to some uncertainty because only part of the population has been measured and because of errors of measurement. The sample size affects the precision of the computed estimates. In general, the larger the sample size, the more precise the estimates will be.⁷⁷

761. Steps in estimating sample size:

1. Decide which indicators the sampling size calculation (also called power calculation) should be based on. The two main types of indicators are:
2. A categorical indicator of which the proportion of elements with a value for this indicator will be estimated. Examples: % of refugees experiencing a security incident within the past three months; % of households having an improved water source; % of displaced population planning to return to their place of origin within the next year.
3. The average or total value of a numeric indicator. Examples: Average household food expenditure within the last seven days; Total value of cash transfers received by population.
4. Estimate standard deviation of the indicator. This value might be known from previous studies or needs to be estimated. Formulas and rules of thumb to estimate the standard estimation if unknown:
 - a. For proportions, the standard deviation is $s = \sqrt{p \times (1 - p)}$, where p is estimated value of the proportion in the population. This might be known from previous studies in a comparable geographic area and population, from a pilot study, or estimated through an educated guess. In the absence of any information on p , it is recommended to set the value of the proportion to 50% ($p = 0.5$) which will lead to the most conservative (=highest) sample size. The standard deviation is $s = 0.5$ in this case.
 - b. For averages of a numeric variable, an initial guess of its standard deviation is required. As for proportions, previous comparable study data and pilot studies can provide good estimates. If nothing is known about the standard deviation of the indicator, it is

⁷⁷ It is important to note that the larger the sample size, the larger the possibility of non-sampling errors. Non-sampling errors include inaccuracies from causes as diverse as refusals, respondent fatigue, interviewer errors or the lack of an adequate sampling frame. Non-sampling errors are harder to predict and quantify than sampling errors. There are no standard formulae to compensate for non-sampling errors. Rather good planning, management and supervision of field operations are the most effective ways to keep them under control. Moreover, it is likely that management and supervision will be more difficult for larger samples than for smaller ones. Thus, one would expect non-sampling error to increase with sample size (Grosh and Munoz, 1996).

Grosh, Margaret and Juan Muñoz. 1996. A Manual for Planning and Implementing the Living Standards Measurement Study Survey. LSMS Working Paper 126. World Bank: Washington, DC.

recommended to determine the smallest and largest value the indicator could reasonably have in the population.

- c. For household food expenditure in the past seven days, this could for example be \$0 for the minimum and \$100 for the maximum. Establish the range by subtracting minimum from maximum and divide the obtained range by 6 to obtain an estimate for the standard deviation: Here 100 divided by 6 leads to a standard deviation of $s = 16.7$.

762. The larger the standard deviation, the larger the sample size required for a given margin of error and confidence interval. Note that the standard deviation uses the same unit as the indicator. Standard deviation is the square root of the variance which is sometimes quoted instead of the standard deviation. The standard variation describes how by how much any given member of a population (household, individual...) differs from a central value such as the arithmetic mean.

763. Margin of error (MOE) and confidence level: The margin of error and confidence level control the amount of the random variation underlying the results of a survey. For the purpose of sample size calculation, the margin of error is determined in advance and, together with the confidence level, expresses the desired precision of the estimate of the indicator of interest. The smaller the margin of error, the larger the sample size required.

764. Random variation stems from the fact that in a survey, not all elements of the population are being studied. The value of an indicator derived from a survey will therefore vary across different samples drawn from the same population. For example, let us assume that 35% of all 50,000 refugees in a camp had the intention of returning to their country of origin within the next year. We would find this proportion if we interviewed all 50,000 of them. However, due to budgetary constraints we have to limit ourselves to randomly sampling only a subset of the entire population. If we draw a sample of 1,000 individuals out of the 50,000 and asked them, we might find that 32% of these 1,000 want to return, instead of the true population value of 35%. If we draw another sample of 1,000 refugees, 37% might answer that they want to return. This variation around the true indicator value of 35% can be expressed formally through the margin of error and associated confidence level. The margin of error should be expressed on the same unit level as the indicator of interest, for example in absolute percentage points for a proportion, or USD/local currency for food expenditure. The confidence level expresses how certain we want to be that the value of the indicator is within the desired margin of error.

765. Example of a margin of error for a proportion:

We want to estimate the proportion of refugees with return intentions within a margin of error of 5% at a confidence level of 95%. This means that, assuming the true (but unknown) proportion among all refugees in a camp is 35%, we want to ensure that the value for the proportion we find in a sample of this population is between 30% and 40% (=35% plus or minus the margin of error of 5%) in 95% of all cases⁷⁸. This is equivalent to a 95%-confidence interval between 30% and 40%.

766. The larger the margin of error, the less precise the estimate of the proportion or average

⁷⁸ "Of all cases" means if we theoretically drew 100 different samples out of the target population, the value of the indicator in the sample data would be between 30% and 40% for 95 of these different samples. We will only draw one sample in reality, but this thought experiment helps to understand the margin of error and confidence level.

is. This gives less confidence that the survey's reported percentages are close to the "true" values of the indicator, that is the value of the indicator in the whole population. A margin of error can be calculated for each indicator produced from a sample survey, unless a nonprobability sample is used. The margin of error is half the confidence interval.

767. **Calculate the sample size of a simple random sample:** The minimum required sample size n for a simple random sample to estimate a generic proportion indicator p (assumed to be 50% in the population) with a desired margin of error MOE and 95% confidence can be calculated as follows⁷⁹:

$$n = \frac{p \times (1 - p) \times 1.96^2}{MOE^2}$$

768. The sample size calculation illustrated above can only be used in a Simple Random Sampling design where final sampling units can be selected completely at random (either from a list or in a field-based method such as a random walk procedure), and each element in the target population has exactly the same probability of being selected. For more complex designs such as cluster sampling, further corrections as described in the following steps need to be applied. It is recommended to enlist the help of a sampling expert for more complex designs.

1. Finite population correction (fpc): A downward adjustment of the sample size obtained in paragraph 768, to account for relatively small target population sizes. If the target population is small (rule of thumb: <10,000 final units), then fpc should be used to lower the sample size while maintaining the precision.
2. Design effect: The design effect (deff) is an upward correction factor of the sample size for cluster sampling designs. It measures the loss in estimation precision due to the clustering of the final sample units compared to Simple Random Sampling. From a statistical point of view it is defined as the ratio of the variance of the estimate obtained from the sample (from the complex design) to the variance of the estimate obtained from a SRS design of the same number of units.

769. **Clustering level** (in statistical terms, intra-cluster correlation): this is a measure of homogeneity within the selected clusters: the more similar the sampling units (individuals, household) within the cluster, the higher the intra-cluster correlation, and the larger the sample size needs to be.

770. **Number of samples per PSU.** PSU = Primary Sampling Unit, i.e. the cluster. How many units (individuals, households) will be selected in each selected cluster. As a general rule, selecting a large number of small clusters is better -other things being equal- in terms of precision than a small number of large clusters. But generally, this requires more resources.

⁷⁹ 1.96 in the formula is the $\left(1 - \frac{\alpha}{2}\right)$ percentile of the standard normal distribution with α being the significance level 0.05 or 1-the confidence level of 95%

PART III – PHASE 2C UNHCR Registration

771. This extract provides information on how UNHCR's registration data is being used by the agency and analysed at the individual, group and population levels. It provides a full explanation of the system. It should be read in conjunction with Phase 2C Heading - *UNHCR Registration Systems (see Part III for more information and Part II for a specimen data sharing agreement between UNHCR and national governments (page 224))*

a. Purpose of registration data

772. Registration data is required for designing, planning and delivering suitable programmes for the population of concern. Registration data analysis strengthens the overall operational intelligence for protection and solutions, helping to facilitate early identification of protection risks and highlights protection gaps as well as target resources for the greatest protection impact.

773. If, for example, persons of concern registration data show a certain percentage of children recorded are unaccompanied or separated, such data basically provides information about actions, such as recruiting a Child Protection Officer, concluding partnership agreement with a child protection partner to establish a child protection programme, conducting child protection training for relevant staff, discussing family tracing with ICRC, fundraising on this basis, etc. that should be taken to respond to the protection needs of this sub-group identified within the population.

774. Conversely, delayed or insufficient attention and analysis of registration data can lead to critical protection gaps that may seriously impact preparedness, negotiating position and resource capacity to respond to urgent protection needs.

775. Consider the following examples of some of the ways in which registration data can be leveraged for effective protection programming and the pursuit of innovative solutions for refugees and others of concern:

Focus of registration data analysis	Possible programme result
Skills and occupational profiles	Pursuit of training programmes and livelihood opportunities based on skills and occupational profiles to encourage and promote socio-economic empowerment.
School drop-out rate	Granular age disaggregation (e.g. 10-11, 12-13, 14-15, 15-16, etc.) identifying the critical ages for school drop-out informs focused group discussions and targeted surveys to understand risks and causes, and to establish programmes to address them.
Mapping of addresses combined with mapping of services available in geographical locations	Identification of gaps in essential services for refugees in the location where they live and advocate for appropriate interventions.
Number of persons with disabilities and disability types	Based on number and type of disabilities recorded, engage new partners, implement targeted assistance and strengthen access to livelihood activities.
Rates of arrest and detention of refugees in a city	If arrests are found to be more common in a

Focus of registration data analysis	Possible programme result
or in a certain part of a city	certain geographical area, advocacy and engagement efforts with law enforcement personnel may be targeted to promote recognition of and respect for refugee documentation and refugee rights.

b. Registration data management

776. Registration data can only be usefully analysed if the data collected at registration is complete, of a good quality and regularly updated. Relying on registration data of a ‘population’ from five years ago would not be likely to provide an accurate picture of population, in terms of group compositions, deaths and arrivals, updated locations, new protection and other needs, as well as changes in access to education, training and work. A good registration system should also include mechanisms to manage its data, making it reliable and effective for planning and programming.

777. **Data Coverage:** In order to leverage data for operational planning and programming, it is important that the right types of data are collected at the point of registration. While it is neither reasonable nor effective to collect all possible data at the point of registration, it is possible to define key data and proxy data fields that reliably inform specific programmes. A good registration system should also include mechanisms to manage its data, making it reliable and effective for planning and programming.

778. **Good quality data:** There are many reasons why registration data may be of poor-quality: data entry error, inadequate training for staff, ill-designed processes and insufficient oversight, cultural misunderstanding or weak standard operating procedures. In order to prevent and manage issues of data quality, several steps and processes can be taken: From a procedural perspective, standard operating procedures, training and regular monitoring and oversight help ensure that data collection is conducted in consistent and harmonized manner; from a technical perspective, regular checks should also be conducted against the registration data. For example, business validation rules that define specific inconsistencies prompt registration users to a data quality error (*e.g. sex male cannot be ‘sister’ to a female family member*).

779. **Regularly updated data:** Sound registration systems and processes require built-in mechanisms to review and update registration data and other information over time. Deep-rooted continuous registration processes will help to maintain accurate and representative data.

780. **Data security and data protection:** The personal data of persons of concern must be managed in a way that protects confidentiality and is consistent with privacy and data protection principles set out in UNHCR’s [Data Protection policy](#). To this end, data repositories (databases) should be designed so that access rights for each content material are determined individually per user, based on a “need-to-know” and “need-to-use” basis. A framework on the segregation of duties aims to clearly define and grant the level of access that corresponds with the user’s specific role and responsibilities, in line with the [Data Protection policy](#) (para. 4.2) and related [Guidance](#) (para. 6.3.8).

10) NEEDS ASSESSMENT DATA:

781. Needs assessments involve systematically gathering and analyzing information relating to the needs, conditions, and capacities of persons of concern in order to determine gaps between a current situation and agreed standards. Needs assessments help understand the conditions faced by populations of concern. Registration data can support needs assessments in multiple ways, including;

782. **As a source of population figures:** Needs assessments require population figures in order to ascertain the severity of a situation (i.e. number of people affected) and to calculate indicators (i.e. baseline population figures);

783. **To identify vulnerabilities related to family compositions:** Statistics on needs can be derived from family compositions, including unaccompanied children, households headed by single or female parents, and unaccompanied elderly persons.

784. **Vulnerabilities related to specific needs:** Registration records contain data on specific needs, e.g. serious medical conditions, children at risk and other types of specific needs which are relevant to needs assessments.

785. **Population sample frame:** If primary data collection will be undertaken in a Needs Assessment, population figures from existing database(s) can be used as the basis for calculating survey samples.

786. **Additional needs assessment data:** Additional questions asked at registration, for example, about refugees' living situations, may be directly relevant to needs assessments. This data can be collated, analyzed and incorporated into a Needs Assessment analysis.

c. Statistical Reporting

787. UNHCR's population, Registration and Identity Management Eco-System (PRIMES)⁸⁰ contains a Data Port element that uses registration data to generate reports, statistics and trends analysis based on global aggregate data. The Data Port was developed to maximize the potential of the data collected and contribute to global data analysis on displacement. The Data Port can be used by UNHCR and Partner Users of PRIMES tools; and will eventually be accessible to persons of concern as well as others looking for statistics on forced displacement. The tool is simple enough to be used by regular users and not exclusively technical users.

788. The Data Port can be used to generate real-time statistics including those required for UNHCR standard reports like the Annual Statistical Report (ASR), Resettlement Statistical Reports (RSR) and Statistical Overviews.

789. Registration data represents a key data source for identifying protection and socio-economic needs for targeted assistance and interventions. Furthermore, where additional information is required, registration data can help identify a suitable sub-set of the population to be sampled for surveys.

d. Registration and Assistance Management

790. Registration data is a key information source for designing and implementing assistance programmes for individuals, households and communities. It provides the aggregate population

⁸⁰ Suite of UNHCR interconnected (registration, identity management, case management, assistance management) tools that forms the basis of UNHCR's work on digital identity.

data needed for situation analysis and assessment and, together with other relevant sources, enables the targeting of assistance delivery, including food, shelter, energy, Water and Sanitation, healthcare, education services and cash-based interventions.

791. Registration data also provides the basis for identity management during assistance distribution. Individuals are verified against their registration record (biographic and biometric) at the point of distribution, and records of those receiving assistance updated after each distribution, ensuring accountability and reconciliation.

792. As such, registration staff support assistance management in several ways, including:

1. Supporting the identification of individuals most in need of assistance based on registration data analysis;
2. Preparing registration data for assistance activities, such as exporting data to the Global Distribution Tool (GDT)⁸¹ to enable identity management at distribution.
3. For assistance activities related to food security for persons of concern, the [2011 MoU between UNHCR and WFP](#), complemented by the [Cash Addendum \(May 2017\)](#) the [Joint Principles on Targeting \(December 2017\)](#) and the [Data Sharing Addendum \(September 2018\)](#)⁸² establish areas of cooperation between UNHCR and WFP, namely in the areas of preparedness planning, joint assessment of the eligibility of refugees and returnees for food assistance, the composition of the food basket or cash equivalent and the modality of the assistance to be distributed.

e. Assistance management tools

793. UNHCR has developed tools to facilitate assistance management processes and to ensure their integrity. **The Global Distribution Tool (GDT)** is UNHCR's corporate tool for identity management and assistance tracking at the point of assistance distribution. The GDT draws on registration data to manage attendance and distribution of assistance in food, Non-food Items (NFI) and cash-in-hand scenarios. The GDT uses UNHCR biometrics to verify (returning results in under 5 seconds) refugees' identities during distribution, speeding up the process and at the same time minimizing avenues for fraud and preventing unauthorised collection. The GDT also allows for better management of information with built-in reports that can be easily generated, including on who has collected assistance and which commodities have been distributed. This information then feeds back to proGres, helping to maintain the accuracy of registration data, together with other information recorded, such as presence/absence and data changes updated at the point of distribution. The GDT is also used by UNHCR's partners, including governments and other UN Agencies.

794. **CashAssist** is UNHCR's corporate tool for cash assistance management that is integrated with proGres, and is a component of PRIMES. CashAssist provides a transparent platform for UNHCR and partners to create and send secured payment instructions to financial service providers. It draws on registration data in proGres and feeds financial and other data from the assistance activities back to proGres, e.g. the cash amount, forms of payment and

⁸¹ Part of the PRIMES Biometrics portfolio, enabling a crucial distribution-focused functionality to operations using biometric.

⁸² The 2018 Addendum on Data Sharing sets out a global framework for the sharing of personal and non-personal data between the two agencies. It simplifies the procedure for requesting data on persons of concern and specifies the data elements that can be shared via this simplified procedure. The Addendum aims to facilitate timely provision of data for programme efficiency and protection, and to reduce duplication in data collection activities.

other payment information, presence or absence of individuals and certain other changes declared and recorded in the CashAssist application.

f. Targeting for assistance

795. Targeting aims to ensure that persons of concern are supported with the most appropriate interventions to address their needs, reinforce their capacities, and exercise their rights. Registration data provides an important overview of the demographic, geographic, specific needs and other characteristics of the population of concern and should be a key data source when considering different targeting approaches and defining targeting criteria.⁸³ There are three main targeting approaches applied in UNHCR operations, each making differentiated use of registration data:

796. **Administrative targeting** is where an administrative party (e.g. UNHCR) not directly representing the community or individuals of concern, defines the eligibility criteria and carries out the targeting process. For example, UNHCR, another UN agency or partner or a government entity may define the eligibility criteria and carry out the targeting process. Administrative targeting is used for RSD, resettlement and some cash-based interventions. Administrative targeting can also be adopted to support risk-prevention objectives, for example, identifying groups and individuals at risk of child labor. Geographical targeting and demographic targeting are forms of administrative targeting that utilize combinations of data elements, many of which are found in the biodata of individual records in proGres (e.g. address, age, gender, dependency ratio, specific needs).

797. **Community-based targeting** is a targeting method in which a group of community members or community leaders decide who in the community should benefit. Where possible, communities define the eligibility criteria based on information they possess as well as information provided to them by UNHCR and other stakeholders. This process is then validated through the collection of additional data or verified through existing data. In participatory approaches, communities may be involved in generating the lists of persons of concern that can be cross-checked and validated by the implementing agency.

798. **Self-targeting**, or self-selection, does not involve the active selection of participants, but relies on programme incentives and conditions to motivate the population to participate. In order to receive targeted assistance, willing participants must express their interest by participating in a programme. Voluntary repatriation is an example of self-targeting. Self-targeting is also utilized in educational programmes, for example high school graduates self-referring to a scholarship programme coordinated by UNHCR.

799. **WFP and UNHCR** have established [*joint principles for targeting assistance to meet food and other basic needs to persons of concern*](#) and have undertaken to collaborate on targeting activities including jointly developing eligibility criteria. The two agencies have also undertaken to share information and results between them when a targeting activity is conducted by one or the other agency alone.

g. Registration data for Livelihood Planning and National Development

800. The Joint Data Centre (JDC), a collaboration between the World Bank and UNHCR,

⁸³ In combination with other sources, e.g. socio-economic assessments, protection monitoring and other relevant available data.

supports the alignment of data and information between humanitarian and development actors, pursuant to the Global Compact on Refugees. It also supports the World Bank's IDA 18 programme, which dedicates World Bank grants to refugee populations and their host communities.

801. Registration data in proGres is 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). When complemented with additional socio-economic household surveys and assessments, this data can provide UNHCR and its partners, as well as the World Bank and its partners, with demographic and welfare information on persons of concern to support tailored programming and planning.

802. In addition, demographic information from registration data can be included in annual Household Budget Surveys conducted by national statistical bodies, ensuring that the specific needs and circumstances of refugees are duly reflected in national development planning. Furthermore, verification and continuous registration activities can be leveraged by UNHCR, the World Bank and their respective partners, to conduct socio-economic survey samples of refugees. The information collected during these exercises can be collated and extrapolated to provide analysis for targeted livelihood programs for refugees as well as their host communities, and as well as to help inform local development programs and policies in host countries.

803. UNHCR and the World Bank also collaborate on the anonymization of registration data for open data access. Open data access via [UNHCR's Microdata Library](#) allows for high-quality analysis and research on refugee and other populations of concern, including for the purposes of identifying gaps in the implementation of services to populations of concern as well as informing broader national development and humanitarian planning.

UNHCR registration data variables

Category	Core data elements	Value to UNHCR
Group pre-registration	Group size Age cohort/sex breakdown Name of Group Focal Point Country of origin of Group Focal Point Specific Needs within the group Unique group identifier Pre-registration data (optional) GPS coordinate	Organizing movements of populations Facilitating assistance distribution, particularly in emergencies Scheduling for individual registration Establish the general profile of a population including demographic characteristics and prevalence of specific needs, which are used to establish initial assistance and protection responses, particularly in emergencies
Individual Emergency	Individual names (full name) Sex	Issuing individual identity documents and entitlement documents as necessary

<p>Registration (IER)</p>	<p>Date of birth</p> <p>Relationship to group focal point</p> <p>Marital status</p> <p>Country of origin</p> <p>Specific needs</p> <p>Photo (strongly recommended)</p> <p>Biometrics (recommended)</p> <p>Date of arrival</p> <p>Registration date</p> <p>Group unique identifier</p> <p>Legal status</p> <p>Current CoA address (if available)</p> <p>Phone number (if available)</p> <p>Consent/prohibition to share information</p>	<p>Ascertaining individual identity</p> <p>Identifying persons with specific needs</p> <p>Targeting and referral to assistance and services</p> <p>Reliable planning and statistics</p> <p>Use in non-emergency contexts for a limited time to address backlog and other challenges resulting in excessive waiting times for registration and access to assistance.</p>
<p>Individual Basic Registration (IBR)</p>	<p>Individual names (full name)</p> <p>Name of father and mother</p> <p>Sex</p> <p>Date of birth</p> <p>Country of birth</p> <p>Relationship to group focal point</p> <p>Marital status</p> <p>Country of origin</p> <p>Citizenship(s)</p> <p>CoA address and telephone number (current)</p> <p>Date of arrival</p> <p>Registration date</p> <p>Specific needs</p>	<p>Issuing individual identity documents and entitlement documents as necessary</p> <p>Ascertaining individual identity</p> <p>Identifying persons with specific needs</p> <p>Targeting and referral to assistance and services</p> <p>Reliable planning and statistics</p> <p>(In non-emergency contexts) for a limited time to address backlog and other challenges resulting in excessive waiting times for registration and access to assistance</p>

	<p>Legal status</p> <p>(Highest) Education level</p> <p>(Last) Occupation</p> <p>Religion</p> <p>Ethnicity</p> <p>Photograph</p> <p>Biometrics(recommended)</p> <p>Consent/prohibition to collect, process, use and share information</p>	
<p>Individual Comprehensive Registration (ICR)</p>	<p>In addition to IBR</p> <p>Additional personal names</p> <p>Place of birth</p> <p>Complete education information</p> <p>Complete occupation/skills information</p> <p>Complete languages information</p> <p>Reasons for flight (for RSD only)</p> <p>Reasons for unwillingness/inability to return</p> <p>Complete relatives' information (spouse and children first, followed by other relevant relatives)</p> <p>Complete documents information (Government-issued, UNHCR-issued, others)</p> <p>Address details (CoA, Country of Origin, others as relevant)</p> <p>Travel details (for RSD only)</p>	<p>Protection monitoring and protection case management, including RSD processing and resettlement activities and pre-identification for complementary pathways (e.g. humanitarian admission programmes, family reunification and opportunities for skilled migration, labor mobility and education).</p> <p>Providing a baseline for all protection programming, including prevention of statelessness among refugee children (monitoring issuance of birth certificates), education and livelihood programming.</p> <p>Targeting for assistance.</p> <p>Other individual intervention.</p>
<p>Individual Enhanced Registration (IEhR)</p>	<p>Collection by registration staff of certain data elements in addition to the individual comprehensive registration data elements, including additional data relevant to targeting for assistance, a specific protection intervention, birth registration, RSD processing, resettlement, return or local integration. It is not about asking the greatest number of questions but rather nuancing the data set to meet 'particular' data needs.</p>	

PART III – 2D Data Integration Techniques

2.3 Design Collection

804. Because of the variety and the varying coverage of the possible sources containing statistical information on refugee and internally displaced persons (IDPs), 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 ESS 2020 vision project in the area of data integration witness of the growing interest for data integration.

805. In particular, data integration can answer several challenges raised when compiling official statistics on refugee and IDPs by:

1. Improving the coverage of target groups,
2. Improving data availability, especially socio-economic variables related to refugee and IDP populations,
3. Improving accuracy through the crossing of different sources in order to assess refugee and IDP status.

806. The objective of this part of the Manual is to present the main steps related to data integration as well as an overview on the possible data integration techniques which can help and guide practitioners in NSOs interested in applying data integration to refugee and IDPs statistics. As far as possible, the provided information will be illustrated by existing good practices in NSOs and will be completed by more detailed technical information and references about the presented methods.

807. The first step is to clearly define the output objectives of integration and to agree this with the providers of the data sources to be used. This is followed by a discussion of the several issues that NSOs will face if they want to apply data integration techniques; these issues relate to the legal environment and data protection legislation, the relationships with external data providers, the required IT resources and the skills needed. The next section addresses issues pertaining to the main sources which could be the object of data integration and is followed by the main data integration statistical methods. These methods are illustrated by examples of good practices. Since any data integration process actually covers all the phases of the GSBPM, practitioners and data compilers can refer to the other parts of the Compilers' Manual for more detailed information on the statistical tasks common to all statistical processes, such as the quality assessment or the design of the process.

2.3.1 Output objectives of data integration

808. The output objectives of data integration should be defined at the very first stage. The feasibility and quality of the foreseen outputs need also to be assessed during this phase. When establishing contacts with possible data providers establishing outputs has to be undertaken simultaneously, in order to have a clear vision and understanding of the data requirements between all stakeholders. Among possible output objectives which could be achieved through data integration, the following list shows the main objectives:

1. Production of new variables of interest (socio-economic variables related to

refugees and IDPs)

2. Improve statistical classification (profiling)
3. Editing and imputation
4. Improve the coverage (detection of duplicates, use of capture/recapture methods)
5. Produce geographically very detailed data (integration with GIS)
6. Improve sampling frame for surveys (Administrative sources, New digital sources)
7. Quality assessment

809. The advantages and limitations of integrating data from different sources are set out below in *The advantages and limitations of integrating data from different sources* are set out below in **Error! Not a valid bookmark self-reference.** The starting point is the administrative register of individuals, to which can be added other data. These other data sources can improve quality, add variables and improve coverage and where new digital sources are used add geographic and movement details.

810. Table 2.17: Advantages and disadvantages of integrating data from different data sources. The starting point is the administrative register of individuals, to which can be added other data. These other data sources can improve quality, add variables and improve coverage and where new digital sources are used add geographic and movement details.

Advantages and disadvantages of integrating data from different data sources

Sources	Administrative registers	General population surveys	Specific surveys (Refugees, IDPs)	Operational data	New digital sources (Satellite images, mobile phones, GIS)
Administrative registers	Determination of status Add variables Improve coverage Improve quality	Add variables	Add variables Improve coverage Improve quality	Improve coverage Determination of status	Geographic context, e.g. proximity to nearest school or hospital Movement patterns, e.g. daily commutes Detailed geographical breakdown Improve Coverage Improve timeliness
General population surveys			Add variables	Add variables	

Sources	Administrative registers	General population surveys	Specific surveys (Refugees, IDPs)	Operational data	New digital sources (Satellite images, mobile phones, GIS)
Specific surveys (Refugees, IDPs)			Improve coverage Improve quality	Improve sampling frame Improve coverage Improve quality	Improve sampling frame Detailed geographical breakdown
Operational data					Estimation
New digital sources (Satellite images, mobile phones, GIS)					Estimation Improve coverage Geographical information New indicators Sustainability and accessibility

2.3.2 Data availability, accessibility and other requirements

811. The design of the integration methodology constitutes one of the main tasks to be realised in data integration. Besides the technical tasks, good collaboration with external data providers and other involved stakeholders is of high importance when designing the data collection, production system and workflows. Very often, the sources that are considered for data integration are not under the direct responsibility of statistical authorities. The first requirement is to ensure the data's availability and accessibility, as well as to have in place the capacities to use them (see also Phase 1.5 Check Data Availability).

812. This process is obviously strongly dependent on the national context, in particular the legal environment as well as the status of the various statistical authorities. The following paragraphs will describe in more detail the main tasks to achieve taking into account the potential national context. Assessment of the points set out below should be undertaken at the very first stage and help to construct a business plan for data integration.

a. Legal environment and data protection legislation

813. 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.

814. A guarantee of data confidentiality, ethical standards, as well as the legal environment obliges NSOs to guarantee the confidentiality of data, either for dissemination or when sharing data with other providers. In particular, NSOs shall avoid direct identification or indirect identification of the respondent (statistical unit). Direct identification means identification of the respondent from their formal identifiers (name, address, identification number), whereas indirect identification means inferring a respondent's identity by a combination of variables or characteristics (e.g. age, gender, education etc). Concerning dissemination, statistical disclosure control can use methods like tabular data protection for aggregate information on respondents presented in tables (use of suppression, rounding, randomisation and/or interval publication) and micro-data protection (using local suppression, sampling, global recoding, top and bottom coding, rounding, rank swapping and micro-aggregation). Concerning the storage of data, physical protection guarantee that the data is securely stored and not accessible to anyone without explicit authorisation. When coming to data integration (micro-data sharing with external data providers), the use of anonymisation techniques (see also g. Security environment discussed below) shall guarantee the confidentiality of the data shared.

b. Cooperation with external data providers

815. Political willingness and technical readiness are essential. Considerations of data security tend to dominate, and a data sharing agreement is a requirement, this should identify the data points to be shared, and the modalities of data transfer. The following issues are recommended;

1. Building trust, depoliticizing the discourse.
2. Define cooperation with data providers and other involved stakeholders (administration, NGOs, other) through clear MOUs (“data used for this and only this”, safeguard of confidentiality by statistical authorities). What are the precise requirement and limits of use of the data, access to micro or macro data?
3. Undertaking a comprehensive mapping of the systems and of the data.
4. Ownership of the results by statistical authorities.

c. IT Resource requirements

816. Data integration has specific IT requirements which are as follows;

1. Necessary software capable of integrating data sources.
2. Compatible data format for receiving and sharing data, for extraction.
3. Secure connections to allow data sharing (VPN, API...).

d. Skills needed

817. In addition to IT data integration needs high level skills which go beyond the statistical;

1. Software architect and engineer to look at integration part,
2. Database developer to then look at structure,
3. Data management expert,
4. Legal advisor,
5. Statisticians/data scientists for building and implementing integration methods.

e. Statistical requirements

818. The availability of unique identifiers (UID) for statistical units or the possibility to build them is necessary. These may be obtained from

- 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
- Building of a synthetic UID based on data matching
- Probabilistic data matching
- Respect of the legislation on data protection

819. There should be a quality assessment of the chosen data integration methods including;

- Assessment of the data linkage quality.
- Assessment of the estimation quality when integrating various data sets.
- Assessment of the applicability and relevancy of the chosen method.
- Documentation on the methods used in order to guarantee transparency and clarity of produced statistics.

f. Collaboration with external data providers and other stakeholders

820. External data providers should be included in the business process in the following areas;

1. Presentation to external stakeholders of the foreseen output (data to be collected, flows, stocks and other variables to be produced, analytical purposes),
2. Mapping in collaboration with data providers between the information contained in

various sources and the official statistics needs,

3. Design of the data collection process in cooperation with external data providers (including metadata),
4. Design production systems and workflows in cooperation with external data providers (software, connection between various stakeholders, IT resources, etc), and,
5. Sustainability of the envisaged process.

g. Capacity requirements

821. Data integration requires considerable statistical capacity including;

1. In house skills which vary depending on the adopted systems but in general which include data science and statistical modelling skills, necessary IT skills to exchange and store data in a secure and efficient way, as well as database administration skills including data mapping documentation, extraction, transformation and load (ETL), database performance tuning for load and retrieval of data.
2. Capacity building
3. Costs.

Access requirement according to various sources

Sources/Access requirements	Legal environment	Availability	Sustainability	Statistical requirements	Required resources (Cost, IT environment, technical skills)
Censuses	Statistical law	Micro/macro data	Depends on statistical authorities	Low Frequency UID	Costly
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
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 (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

e. Security environment

822. When exchanging or sharing data from multiple data providers, this process shall be in conformity with the data protection legislation. It means in particular that any data exchange is secure, and that micro-data confidentiality is respected. In addition, confidentiality of data disseminated must also be respected. Secure data exchange can be realised through data encryption, whereas hashing techniques can be used for ensuring micro-data confidentiality. Lastly, confidentiality of data for processing and analysis can be done with pseudonymisation techniques. Below is presented a short presentation of data encryption, pseudonymisation and anonymisation techniques.

2.3.3 Confidentiality

a. Data encryption

823. In order to secure data exchange, encryption techniques can be used. Encryption techniques consist of making a message unreadable for those who do not have the key to make it readable again. Two families of encryption techniques can be distinguished:

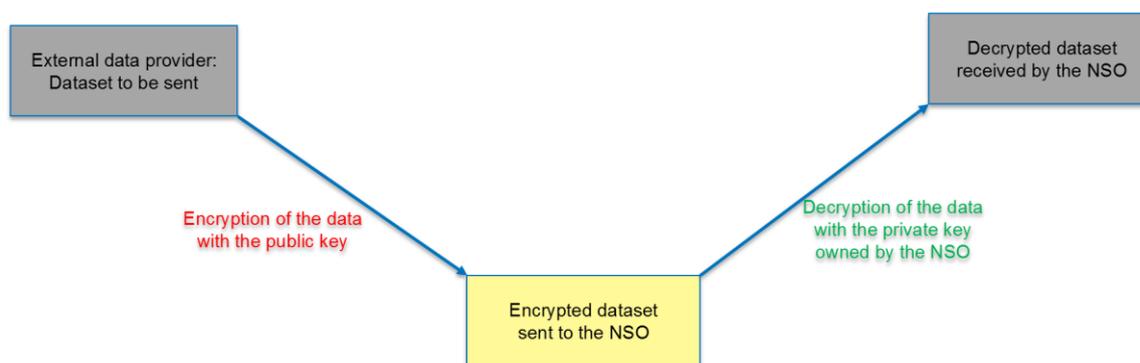
- Symmetric methods: the same key is used for encrypting and decrypting.
- Asymmetric methods: two keys are used, a public key and a private key.

824. In Symmetric methods, the sender and the recipient of the encrypted message must have the same key, which is kept secret. These methods have several drawbacks:

- The key must be exchanged in a secure way.
- If a NSO wants to exchange data with several external data providers, it will require a secret key for each provider, which implies being able to manage a high number of keys.

825. Asymmetric encryption methods overcome this problem. These techniques are based on pairs of keys, a public one and a private one. The public key may be disseminated widely, whereas the private key is known only by the owner. The encrypting function (public key) shall not be revertible unless the user has a secret code (private key) which enables the decryption of the encrypted message. Effective security only requires keeping the private key private; the public key can be openly distributed without compromising security. In such a system, any person can encrypt a message using the receiver's public key, but that encrypted message can only be decrypted with the receiver's private key.

Asymmetric encryption



b. Pseudonymisation

826. Pseudonymisation can be defined as a data de-identification procedure by which personally identifiable information fields within a data record are replaced by one or more artificial identifiers, or pseudonyms. A single pseudonym for each replaced field or collection of replaced fields makes the data record less identifiable while remaining suitable for data analysis and data processing. Pseudonymisation can be for example a way to comply with the European Union's new General Data Protection Regulation requirement for secure data storage of personal information. Pseudonymised data is a revertible process that can always be restored to its original state with the addition of information which then allows individuals to be re-identified, while anonymized data can never be restored to its original state.

827. Data masking is the standard solution for data pseudonymisation. Using masking, data can be de-identified and de-sensitised so that personal information remains anonymous in the context of support, analytics, testing, or outsourcing.

*c. Data hashing*⁸⁴

828. Hashing techniques are function aiming at generating a fixed-size fingerprint (or signature) from any data whatever the size. A hash function allows any data to be mapped to an element of a finished set, whose cardinal is very large. This means that there is no inverse operation that will allow the initial data to be retrieved from the fingerprint. Moreover, the distance between two fingerprints of two data is independent of the distance between these data. A minimal difference between two data leads to two very different fingerprints. In contrast, two very different data may have similar or identical fingerprints.

829. For example, hashing the chains “Dupont” and “Dupond” by the SHA256⁸⁵ function gives the following fingerprints:

SHA256(~Dupont~)	=
3bde3a5999601d8fa7b6bcc6bfdd2ee6a9fb473043d9768fbf8274b5936ef4d2	
SHA256(~Dupond~)	=
535a7594e59be910df06483d24371c7697854fa84d8ed8c0f400126edc25af3a	

830. A good hash function presents a low risk of collision, which means that for different data of a similar size, the probability of having the same result after using a given hash function is extremely low. It should also be added that collisions that could be introduced by hashing are minor in comparison with problems of homonymity, which can occur in practice when clear text data such as surnames, first names and dates of birth are manipulated.

831. As the hash function has no inverse function, processing by hashing is said to be irreversible. However, if the hash function is known, it is possible to retrieve original data from its signature, thanks to so-called dictionary attacks. In order to deal with such attacks, a solution therefore consists in modifying the chain before applying the hash function either by adding a secret key or by applying a secret function to the data.

832. There are several “standard” hashing functions⁸⁶ (MD5, SHA1, SHA256, SHA512, ...), which are continually being studied for their resistance to attacks. For example, in 2016, the French agency for the security of information system recommended the use of the SHA-256 method.

833. Hashing techniques applied to unique or quasi-unique identifiers make it possible for anonymised files to be linked. It is also worth mentioning the use of double hashing when files are to be linked, for example, from several data providers. It is necessary to hash the identifier fields in the same way (with the same secret key) in all of the files to allow linkage according to these fields. However, the establishment (NSO) that receives these files carries out a second hashing with a second secret key so as to render the obtained dataset anonymous vis-à-vis the external data provider. Hashed data can then be used either for deterministic or probabilistic

⁸⁴ Securitising data linkage in French public statistics <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5053094/>

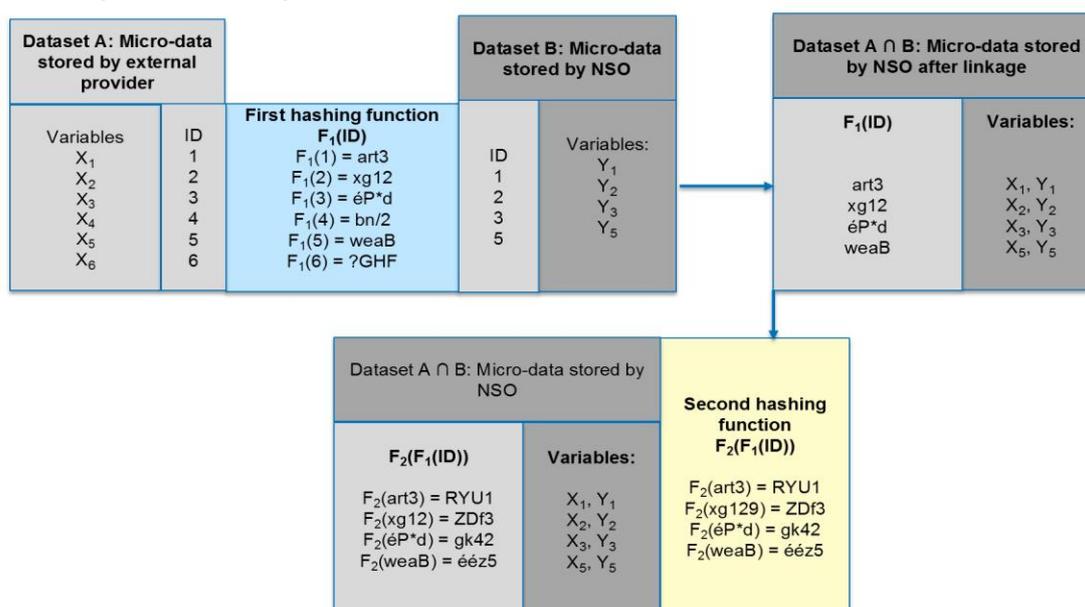
⁸⁵ A cryptographic hash (sometimes called 'digest') is a kind of 'signature' for a text or a data file. SHA-256 generates an almost-unique 256-bit (32-byte) signature for a text.

⁸⁶ National Institute of Standards and Technology. Secure Hash Standard (SHS). Federal Information Processing Standards Publication; 2012. <http://csrc.nist.gov/publications/fips/fips180-4/fips-180-4.pdf>

linkage methods.

834. However, data hashing does have drawbacks. Indeed, if the slightest error is made in entering the name, for example, the hashed data for the misspelled name will be completely different from that obtained for the correct name. Use of upstream normalization procedures before hashing may limit such problems (e.g. lower case only, suppression of accents, etc ...). Moreover, for the same reason to calculate an edit distance (for example, Levenshtein distances, or Hamming distances) to use in deterministic linkage.

Data hashing when linking micro-data:



835. Finally, given the irreversible nature of hashing, it is important to keep the unhashed data, as they may be exploitable only with data that underwent the same hashing process.

2.3.4 Design, processing and analysis

836. This section describes the different tasks required by data integration and provides a description and references on the methods that could be applied. This part is mainly based on the work done within the framework of the ESS Vision 2020 ADMIN (Administrative data sources), in particular the one on data integration and estimation methods⁸⁷. Nevertheless, this section shows as far as possible how integration and estimation methods could be applied to the field of refugee and IDP statistics.

a. Data editing and imputation

837. When using multisource data, inconsistencies at micro level are frequently encountered. Integration of data sources at micro-level may give rise to composite records that consist of a combination of values obtained from several different sources. For instance, a register could

⁸⁷ https://ec.europa.eu/eurostat/cros/content/wp2-statistical-methods_en

be combined with values obtained from a survey for the same units (obtained by record linkage); or from an integration of several surveys with non-overlapping units, in which case a unit from one source is matched with a similar (but not identical) unit from another source. Records with values obtained from different sources can also arise as a consequence of item non-response and subsequent imputation, in which case the two sources are the directly observed values versus the values generated by the imputation method. In all these cases the composition of a record by combining information obtained from different sources may lead to consistency problems because the information is conflicting, in the sense that edit rules that involve variables obtained from the different sources are violated. The purpose of editing conflicting micro-data is to achieve numerical consistency in the first instance and, ultimately, statistical consistency of the resulting aggregates.

838. For example, a survey provides an aggregated value A for the total income received from social allowances received by person x; this value A should be equal to the sum of two types of allowances B (health) and C (education) collected through administrative sources. If A is not equal to the sum of B and C, this inconsistency should be removed by using micro integration methods⁸⁸ like prorating, minimum adjustment methods, or generalised Ratio Adjustments. In the case of data on refugee and internally displaced persons, such aggregation rules when integrating several sources should be rather rare, but their existence should be checked on a case by case basis, and data compilers should take them into account if necessary. Micro integration or editing conflicting micro-data needs to be distinguished from variable harmonisation (later on described) which applies to situations where there exist similar versions of the same target variable, for instance due to different definitions across the sources.

b. Creation of joint statistical micro data

1) DATA LINKAGE: COMBINING DATA FROM SEVERAL SOURCES BELONGING TO THE SAME UNITS

839. The integration of data at micro level with the main purpose of accurately reflecting the individual's real-world circumstances at individual micro level, even when the data are stored in different sources of various types, is known as record linkage. Ultimately, the record linkage process creates a micro dataset where the corresponding separate observations are combined into joint statistical data drawn from information which originate in multiple sources. Record linkage is also referred to as object identification, record matching, entity matching, entity resolution, or reference reconciliation. In the case of the National Statistical Institutes (NSIs) the joint use of statistical and administrative sources is a product of a rationalization of all the available sources in order to reduce costs, response burden and, most of all, to enrich the information collected in order to produce high quality statistics.

840. An important distinction between record linkage and statistical matching needs to be made, the former regards the "fusion" of sources composed mainly of the same units, partially or completely overlapping, e.g., in the case of integration of administrative registers and sample surveys, while the latter concerns the integration of different units, e.g. derived from different sample surveys.

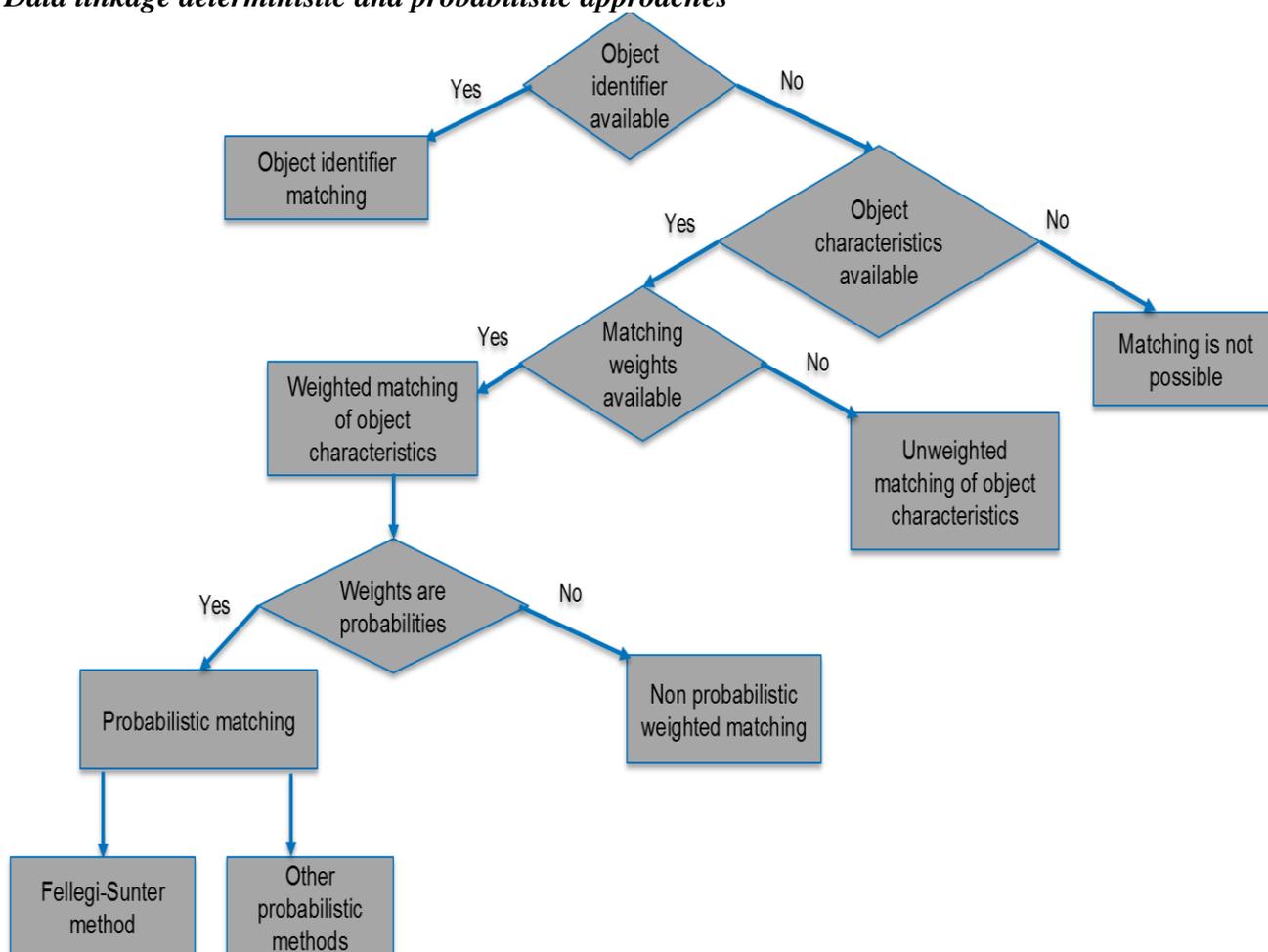
841. The record linkage problem requires statistical estimation methods when unique identifiers are not available for all the units in the sources, or the unique keys are affected by

⁸⁸ Pannekoek, J. and Zhang, L.-C. (2015) Optimal adjustments for inconsistency in imputed data. *Survey Methodology*, vol. 41, pp. 127-144.

errors, or a unique composite key can be derived from a combination of error-prone variables. The presence of errors in the unit identifiers and the use of statistical estimation methods for linking data files, may introduce linkage errors in the integrated data. This point needs to be taken into consideration by data compilers' during the design of the data linkage process. Actually, a standard statistical estimation approach can produce inaccurate results in case of linkage errors, which requires the use of statistical methods in order to evaluate and adjust for linkage errors.

842. The deterministic record linkage approach that links records if, and only if, there is a full agreement of unique or common identifiers - the matching variables; or if it has satisfied some *a priori* defined specific criteria. All the data couples that present the same values on the selected matching variables or satisfy the chosen criteria are assigned with certainty to the set of matches. On the basis of this approach, the uncertainty in the matching procedure between two different databases is minimized but the linkage rate could be very low; the linkage quality can be assessed only by means of accurate and expensive clerical reviews or by means of re-linkage procedures.

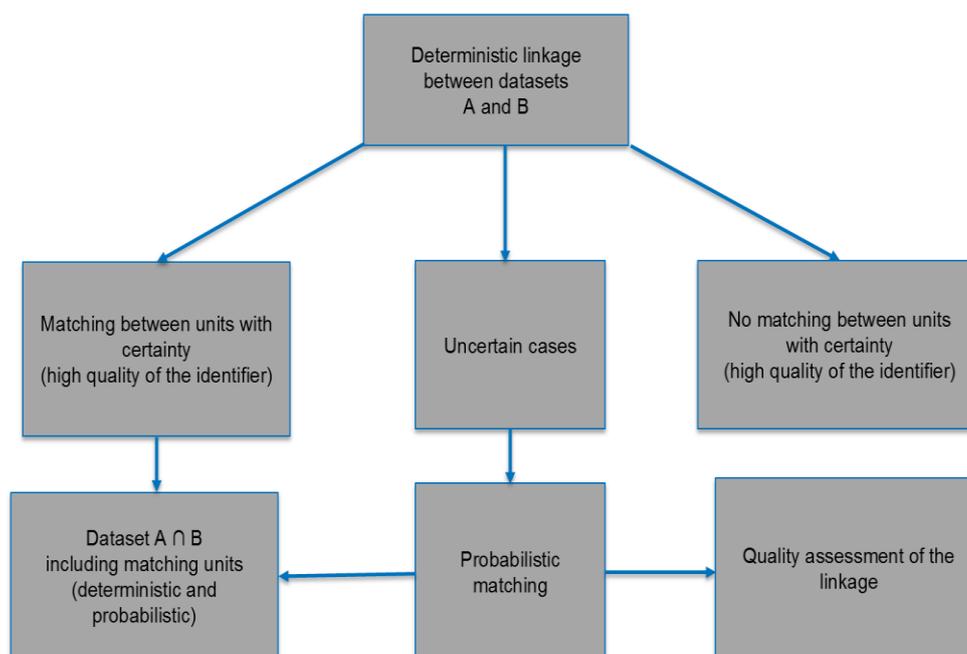
Data linkage deterministic and probabilistic approaches



843. Compared with the deterministic approach, the probabilistic one can solve problems caused by bad quality data and can be helpful when variables that are differently spelled,

swapped or misreported are stored in the two data files. The probabilistic record linkage approach is formally based on a statistical methodology and allows also to evaluate the linkage errors, calculating the likelihood of the correct match. Generally, the deterministic and the probabilistic approaches can be adopted jointly in a two phase process: firstly the deterministic method is performed, choosing the high quality variables, then the probabilistic approach is adopted on the residuals which are the units not linked in the first step, the joint use of the two techniques depends on the aims of the whole linkage project.

Deterministic linking between datasets



Possible estimation methods:

844. Estimation methods can follow two different types of approaches:

1. The standard approach for probabilistic record linkage proposed by Fellegi and Sunter (1969) and Jaro (1989), which define the record linkage problem as a classification one, where the matching status is unknown and needs to be estimated by means of a decision/classification rule.
2. The Bayesian approach (Fortini et al, 2001, Tancredi and Liseo, 2011) where prior distribution of the linkage probabilities and the number of matches and posterior distributions are simulated using MCMC. More recently, Steorts et al (2014) a different Bayesian approach, based on a parametric model for categorical data that addresses matching different files, detecting simultaneously duplicate records within the lists. The pattern of matches and non-matches was represented as a bipartite graph, in which records are directly linked to the true but latent individuals which they represent while they are only indirectly linked to other records.

845. Another topic is the simultaneous linkage of more than two sources, the so called multiple record linkage (Sadinle et al.). The longitudinal nature of some linkage processes,

which involve updated version of administrative data has to be taken into account as well. In these cases, the statistical estimation methods are sometimes confused with IT solutions as incremental record linkage Gruenheid et al (2014), and parallel record linkage Christen et al (2002).

846. In connection with data linkage, the impact of linkage error and the related adjustments are largely studied in recent in Chambers 2009, whereas in official statistics, some research activities are devoted to solving recent issues related to the use of record linkage techniques. One of the active research fields is the privacy preserving record linkage (PPRL), related to the use of pseudo-anonymized data, so as to preserve data from privacy issues and be able to link data provided in a more protective framework; see Fienberg, (2010) and Christen (2012) and for the most comprehensive survey Vatsalan et al, (2013).

References:

- Chambers R. (2009). *Regression Analysis of Probability-Linked Data*. Official Statistics Research Series 4.
- Christen, P. (2012). *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*. Heidelberg: Springer, Chapter 8: Privacy Aspects of Data Matching.
- Christen P., Hegland M., Roberts S., Nielsen O. M., Churches T., Lim K. (2002). *Parallel computing techniques for high-performance probabilistic record linkage* available at <http://datamining.anu.edu.au/linkage.html>
- ESSnet Statistical Methodology Project on Integration of Survey and Administrative Data (ISAD), Report of WP2. *Recommendations on the use of methodologies for the integration of surveys and administrative data, 2011*, available at https://ec.europa.eu/eurostat/cros/content/deliverables-wp2di-integration-applications-isad-wp2_en
- Fellegi I P and Sunter A B (1969) *A theory for record linkage*, *Journal of the American Statistical Association*, Vols. 64: 1183-1210.
- Fienberg S. Hall R. (2010) *Privacy-preserving record linkage*, in: *Privacy in Statistical Databases*, Springer Lecture Notes in Computer Science, vol. 6344, Corfu, Greece, 2010, pp. 269–283.
- Fortini M., Liseo B., Nuccitelli A., Scanu M. (2001), *On Bayesian record linkage*, *Research In Official Statistics*, 4, 185–191.
- Gruenheid A. Dong X. L., Srivastava D. (2014) *Incremental Record Linkage*. *Proc VLDB Endowment* 7 (9):697-708.
- Jaro M.A. (1989), *Advances in record-linkage methodology as applied to matching the 1985 Census of Tampa, Florida*. *Journal of the American Statistical Association*, 84, 414–420.
- Sadinle M., Hall R., and Fienberg S. *Approaches to Multiple Record Linkage* available at <https://www.cs.cmu.edu/~rjhall/ISIPaperfinal.pdf>
- Shlomo, N., & Skinner, C. (2010). *Assessing the protection provided by misclassification-based disclosure limitation methods for survey micro-data*. *Annals of Applied Statistics*, 4(3), 1291-1310. DOI: 10.1214/09-AOAS317. Publication link: [fbb6d0a6-32fd-41c0-9aac-6db410555bab](https://doi.org/10.1214/09-AOAS317)
- Steorts, Hall, Fienberg (2014) *SMERED: A Bayesian Approach to Graphical Record Linkage and De-duplication, a new approach in the Bayesian framework*, *Journal of Machine learning research*.
- Tancredi, Liseo (2011) *Hierarchical Bayesian approach to record Linkage and population size problems*. *The Annals of Applied Statistics*, 2011, Vol. 5, No. 2B, 1553—1585.

- Vatsalan, Christen, Verykios (2013) *A taxonomy of privacy-preserving record linkage techniques Information Systems* 38 (2013) 946–969.

2) STATISTICAL MATCHING: INFERENCE OF JOINT DISTRIBUTION FROM MARGINAL OBSERVATIONS

List of possible methods:

847. Concerning micro-objective, the most popular statistical matching methods are based on imputation techniques. They are mainly based on cold deck methods (see Singh et al., 1993). The resulting data set can be used for inferential purposes on the joint (conditional) distribution of the pair of variables of interest only if the conditional independence assumption between the pair of random variables of interest given the conditional ones holds. This assumption seldom applies, with the remarkable exception of the case where the pair of variables of interest is highly correlated with one of the conditional variables.

848. In order to deal with the uncertainty caused by the lack of joint information on the pair of variables of interest, one can select, based on maximum likelihood or Bayesian approach, a set of equally plausible estimates corresponding to all those joint distributions that are compatible (or that can co-exist) with the maximized identifiable distributions, namely the marginal distribution of the conditioning variables and the two conditional univariate distributions of the pair of random variables of interest given the conditional ones. This set of estimates is studied in D’Orazio et al (2006a, 2006b) in a maximum likelihood context. In a non-proper Bayesian case, this set of estimates is studied in Rubin (1986) and extended to a proper Bayesian context in Raessler (2002).

849. When dealing with statistical matching related to samples drawn according to complex survey designs, one can refer to Rubin (1986, with the so called concatenated sample) and Renssen (1998, with the incomplete and synthetic two-way stratification).

References:

- Ahfock D., Pyne S., Lee S.X., McLachlan G.J. (2016) *Partial identification in the statistical matching problem. Computational Statistics & Data Analysis, Volume 104, December 2016, Pages 79–90*
- Conti P.L., Marella D. Scanu M. (2015). *Statistical matching analysis for complex survey data with applications. Journal of the American Statistical Association, to appear (available on line)*
- D’Orazio M., Di Zio M., Scanu M. (2006a). *Statistical Matching: Theory and Practice. John Wiley & Sons, Chichester, ISBN: 0-470-02353-8*
- D’Orazio M., Di Zio M., Scanu M. (2006b). *Statistical Matching for Categorical Data: Displaying Uncertainty and Using Logical Constraints, Journal of Official Statistics, 22, 137-157.*
- D’Orazio M., Di Zio M., Scanu M. (2009). *Uncertainty intervals for non-identifiable parameters in statistical matching. ISI conference, Durban (South Africa).*
- D’Orazio M., Di Zio M., Scanu M. (2016). *The use of uncertainty to choose matching variables in statistical matching. 8th International Conference on Soft Methods in Probability and Statistics (SMPS), Rome, September 12-14 2016.*

- Di Zio M., Vantaggi, B. (2017). *Partial identification in statistical matching with misclassification. To appear in International Journal of Approximate Reasoning.*
- Donatiello D., D'Orazio M., Frattarola D., Rizzi A., Scanu M., Spaziani M. (2016) *The role of the conditional independence assumption in statistically matching income and consumption. In press, Statistical Journal of the IAOS*
- Fosdick, B. K., De Yoreo, M. and Reiter, J. P. (2016), *Categorical data fusion using auxiliary data, Annals of Applied Statistics, to appear.*
- Marella D. Conti. P.L., Scanu M. (2013) *Uncertainty analysis for statistical matching of ordered categorical variables. Computational Statistics and data Analysis, 68, 311, 325*
- Marella D. Conti. P.L., Scanu M. (2016) *How far from identifiability? A systematic overview of the statistical matching problem in a non-parametric framework. Communications in Statistics – Theory and Methods, in press. DOI#10.1080/03610926.2015.1010005. Journal ISSN 1532-415X*
- Memobust (2014) *Module Micro-Fusion - Statistical Matching. https://ec.europa.eu/eurostat/cros/content/handbook-methodology-modern-business-statistics_en*
- Memobust (2014a) *Module Micro-Fusion - Statistical Matching Methods. https://ec.europa.eu/eurostat/cros/content/handbook-methodology-modern-business-statistics_en*
- Okner, B.A. (1972) *Constructing a new data base from existing micro-data sets: the 1966 merge file. Annals of Economic and Social Measurement 1(3), 325–342.*
- Paass, G. (1986) *Statistical match: evaluation of existing procedures and improvements by using additional information. In G.H. Orcutt, J. Merz and H. Quinke (eds) Microanalytic Simulation Models to Support Social and Financial Policy, pp. 401–422. Amsterdam: Elsevier Science.*
- Raessler, S. (2002) *Statistical Matching: A Frequentist Theory, Practical Applications and Alternative Bayesian Approaches. New York: Springer-Verlag.*
- Reiter, J. P. (2012), *Bayesian finite population imputation for data fusion, Statistica Sinica, 22, 795 - 811.*
- Renssen, R.H. (1998) *Use of statistical matching techniques in calibration estimation. Survey Methodology 24, 171–183.*
- Rubin, D.B. (1974) *Characterizing the estimation of parameters in incomplete-data problems. Journal of the American Statistical Association 69, 467–474.*
- Rubin, D.B. (1986) *Statistical matching using file concatenation with adjusted weights and multiple imputations. Journal of Business and Economic Statistics 4, 87–94.*
- Sims, C.A. (1972) *Comments on Okner (1972). Annals of Economic and Social Measurement 1(3), 343–345*
- Singh, A.C., Mantel, H., Kinack, M. and Rowe, G. (1993) *Statistical matching: Use of auxiliary information as an alternative to the conditional independence assumption. Survey Methodology 19, 59–79.*

Statistical matching: inference of joint distribution from marginal observations

850. In statistical matching (SM), also referred to as data fusion, the ultimate aim is to provide estimates on variables observed in distinct data sources by exploiting the available common information (variables observed in both data sources). The data to be joined is characterized by representing the same population, but not necessarily having common units. Thus, statistical matching can be framed as an imputation problem in which a missing target

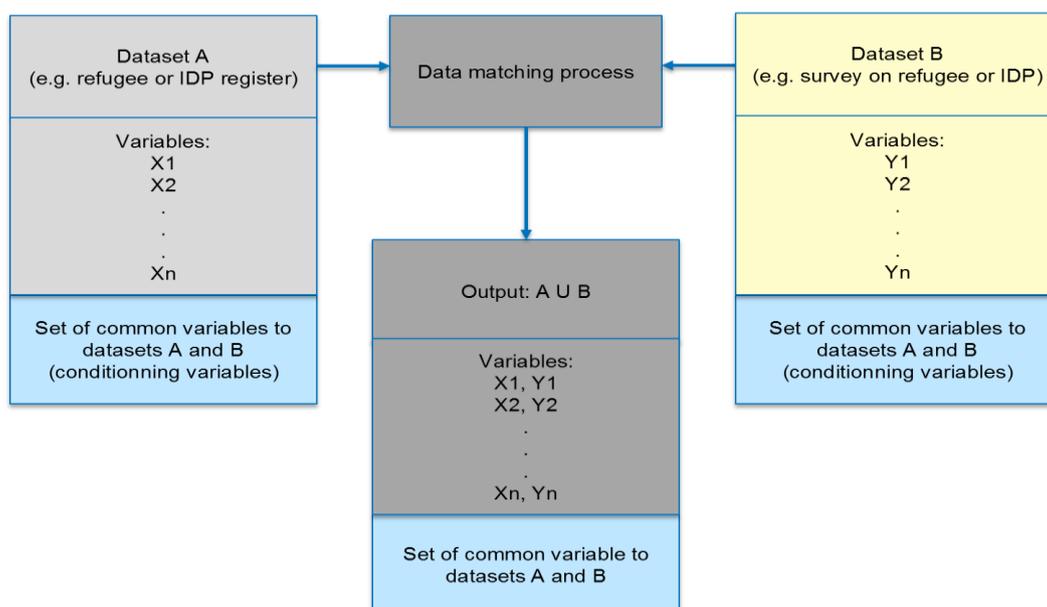
variable for units in a recipient data set is predicted based on information on similar units (“statistical twins”) observed in a donor data set.

851. A statistical matching problem can be defined according to its objective and data availability:

1. (objective) knowledge on a (possibly conditional) joint distribution function of a couple of random variables, or on some of the parameters representing the (conditional) joint relationship between the two random variables, or on a data set representative of the (conditional) joint distribution of the two variables, is requested;
2. (available data) the two random variables of interest are not available jointly in any datasets, but it is possible to use two random samples, each one observing one of the two random variables, both observing all the conditioning variables. Furthermore, the two samples should be representative of the same population, but the overlap between the two observed sets of units is either null or consists of subsets of units that are not representative of the population itself.

852. The two most important outputs that statistical matching can produce are: either macro (estimation of a joint distribution or of a relationship parameter) or micro (a data set with complete observations on the variables of interest, i.e. both the couple of random variables and the conditioning variables). It is worthwhile noting that the macro approach to statistical matching may be also be considered as a ‘Multisource estimation at aggregated level’. In the field of refugee and IDP statistics, the micro-objective (fusion of two datasets) is the most frequent possible use of data matching.

Micro-objective: fusion of two datasets



853. It should be pointed out that even if a unique personal identifier had been available in the two samples, it would not have not been useful because the data sets to integrate do not overlap. Identification is not requested for the statistical matching problem. This statement

should be considered as a warning: a statistical matching problem does not correspond to the identification of the most similar unit to the one to integrate (as stated in Okner, 1972, p. 327), hence adaptations of the record linkage methods for statistical matching purposes via the use of imputation procedures should be generally avoided (see Sims, 1972).

854. In fact, the main source of error in a statistical matching problem is not the linkage error, as in record linkage, but the possible introduction of specific relationship models between the variables never jointly observed by means of the statistical matching method itself. In a micro approach, the data set resulting from such an activity becomes representative of the imposed model, possibly very different from the truth. The typical imposed model in this framework is the statistical independence between the never jointly observed variables given the conditional ones. This is, for instance, the imposed model by any imputation procedure filling in the missing variables in the two observed samples by a function of only the conditioning variables.

855. In order to avoid a result that can be completely misleading, the new approaches to statistical matching aim at introducing a new form of uncertainty that goes beyond sample variability: uncertainty due to lack of joint knowledge on the couple of variables of interest. (see Part 3 for more detail and a list of possible methods with references)

2.3.4 Alignment of populations and measurements

a. Alignment of populations: Unit harmonisation

856. When data originating from multiple sources are used, one needs to transform objects into statistical units referring to the target population. This is a mandatory requirement valid both for usages based on a single data source, and for cases based on multisource data. After a preliminary analysis aimed at verifying such a consistency, it is sometimes necessary to start a process of harmonisation or alignment of statistical units. For example, when combining survey data with administrative registers, to create a living household as the ideal unit for refugee or IDP statistics, one needs to make use of units such as person, family, residence address, study or workplace address, etc. The different units need first to be aligned with each other, in order to improve the accuracy of the living household created based on them.

Possible estimation methods:

857. No general statistical methods are available, while ad-hoc deterministic methods are generally used. For instance, statistical units are derived using certain deterministic derivation rules based on available information.

b. Alignment of measurements: Variable harmonisation

858. When administrative data are used, one needs to transform attributes of the objects into measurements referring to the concepts to be measured, that is the target variables. After a preliminary analysis aimed at verifying such a consistency, it is sometimes necessary to start a process of harmonisation or alignment of measurements.

859. When data from multiple sources are combined, differences in definition can occur between variables in different sources. In particular, variables in an administrative data source

are defined according to the administrative purposes of the register owner. These definitions may differ from those of the target variables for statistical purposes. For example, when the variable 'age-group' is observed in two data sources with different groups. A technique for obtaining a unique variable 'age' is needed. This apparently deterministic task can involve statistical estimation method (e.g., classification techniques) whenever a well-defined mapping is not known.

860. In case differences in variable definitions occur between data sources, these variables need to be harmonised during data integration. That is to say, for each unit in the integrated data set, the values of the target variable according to the desired definition need to be estimated from the observed values that are available.

2.3.5 Multisource estimation at aggregated level

861. This task refers to the phase of production of estimates by using integrated multisource administrative data. Estimation methods in this case must deal with the problem that data are not obtained according to a random sampling design, and that observations may refer to specific subsets of units of the population (under-coverage), or out-of-scope units (over-coverage). In addition, when integrated multisource data are used, specific problems may arise. They are concerned with the problem of consistency and coherence of estimates.

a. Population size estimation: multiple lists with imperfect coverage of target population

862. When each one of the multiple sources has imperfect coverage of the target population, including both under- and over-coverage, statistical methods for population size estimation are needed. The most common approaches relate to capture-recapture (CRC) methods. In estimating a population size based on several sources, the misalignment between the scope of the administrative data and that of the statistician poses several methodological challenges and sets us apart from a classical capture-recapture setting. For instance, it is often useful to develop methods taking into account the dependence among data sources, the fact that some data sources refer to a specific subpopulation, and that data may contain units out-of-the target population that are not deterministically identifiable.

863. 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 face this task is the capture-recapture (CRC) method, which was originally been developed to estimate the size of animal populations (Fienberg, 1972; Bishop, Fienberg & Holland, 1975; IWGD MF, 1995). 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.

864. 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. Difficult to reach refugees or IDPs through surveys will most likely not be recorded in administrative registers, which can break the third assumption and imply a problem of under-coverage. 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, applying the CRC method to each identified specific group before proceeding to an aggregation of the results found for each group. Lastly, the fifth hypothesis could be assumed to be verified in most of the cases.

865. The violation of these assumptions can lead to serious bias in the CRC-estimation of the population size. CRC is in particular sensitive to violation of the assumption in the case of a low implied coverage, i.e. the second register overlaps greatly with the first register and adds relatively few new records to it. So, several extensions of the CRC method were proposed in order to face problems connected to violation of the basic assumptions, we can divide them into two group: methods aiming at improving the CRC-method; alternatives methods.

866. While the class of CRC model was explicitly designed and developed to estimate the under-coverage, recently the estimation of the over-coverage has emerged as an important subject when studying population size estimation methods, in particular when the multiple lists are collected for administrative purposes. In fact, the risk that the administrative data contain units out-of-the target population, as well as duplicated units, is higher with respect to survey data collected for statistical purpose.

867. There are different approaches for measuring and/or integrating a measurement of the over-coverage into population size estimation, some of them dealt with different types of over-coverage separately, and therefore make separate adjustments to the population estimates (e.g. Statistics Canada, 2015 and ONS, 2012). Other methods have been developed in alternative to the CRC approach, both for the evaluation of the under-coverage and for the over-coverage.

1) EXAMPLE: BASIC CAPTURE RECAPTURE METHOD

868. Let assume two different sources for a giving area (survey and administrative sources for example), the following data are collected:

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

Based on those 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:

$$N^* = (N_{10} + N_{11}) / [N_{11} / (N_{01} + N_{11})] = 360 / (140/1000) = 3\ 571$$

With $N_{10} + N_{11}$ = the number of IDPs identified during the survey

and $N_{11} / (N_{01} + N_{11})$ = the share of identified IDPs during the survey in the number of identified IDPs in the administrative list.

The variance of the estimator in this case would be:

$$V(N^*) = N^* N_{01} N_{10} / N_{11}^2 = 7500 \times 960 \times 260 / 40^2 = 56\ 414$$

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

Possible estimation methods:

869. The standard statistical method for evaluating the population size is based on the CRC model (also known as the Petersen or Lincoln-Petersen model). Extensions of this methodology that take into account the violation of the basic assumptions, aiming at improving the CRC model, include:

- methods connected to the correction for linkage errors (Ding and Fienberg, 1994; Di Consiglio and Tuoto, 2015);
- methods accounting for the over coverage (e.g. Zhang, 2015, Zhang and Dunne, 2017);
- methods dealing with partially overlapping populations (e.g. Zwane et al., 2004);
- methods relaxing the independence assumptions, considering more than two lists;
- methods dealing with heterogeneity in capture probabilities.

870. Alternatives methods with respect to the CRC model and its extensions have also been proposed in order to deal with over coverage and partially overlapping populations; for instance, a latent class modelling approach is described in Di Cecco et al (2016).

871. Rasch models have been proposed in order to deal with dependencies between sources and heterogeneity of captures.

872. Finally, it is interesting to cite the Bayesian approaches to the estimation of the population size. Besides the Bayesian capture-recapture model (Ghosh and Norris 2005), in this field two main groups of methods can be identified: the former is mainly related to the record linkage topic and its outcome in population size estimation (Steorts et al 2014, Tancredi and Liseo, 2011), the latter is connected to the use of Bayesian approaches in order to evaluate and projecting demographic stocks and flows in human population (Raftery et al 2012, Bryant and Graham, 2013).

873. Most of the cited approaches are also object of ongoing research in official statistics.

References:

Baffour, B., J.J. Brown, P.W.F Smith, (2013). *An investigation of triple system estimators in censuses. Statistical Journal of the International Association for Official Statistics*, vol. 29, pp. 53-68

Bakker, B.F.M., & P. Daas, 2012, *Some Methodological Issues of Register Based Research, Statistica Neerlandica*, vol. 66, nr. 1, pp. 2-7

Bartolucci, F. and Forcina, A. (2001), *Analysis of Capture-Recapture Data with a Rasch Type Model Allowing*

- for Conditional Dependence and Multidimensionality, *Biometrics*, 57, 714–719
- Bishop, Y., Fienberg, S., & Holland, P., (1975). *Discrete multivariate analysis, theory and practice* New York: McGraw-Hill.
- Bryant and Graham, (2013). *Bayesian Demographic Accounts: Subnational Population Estimation Using Multiple Data Sources*. *Bayesian Analysis*, 8,3, pp.591–622
- Brown, J.J., O. Abott & I.D. Diamond, (2006). *Dependence in the 2001 one-number census project*. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 169, 883-902
- Di Cecco, D., Di Zio, M., Filipponi, D., Rocchetti, I. (2016). *Estimating population size from multisource data with coverage and unit errors*. *Proceedings of the ICES V 20-23 June 2016 Geneva*
- Di Consiglio L., T. Tuoto, (2015). *Coverage Evaluation on Probabilistically Linked Data*, *Journal of Official Statistics*, vol. 31, nr. 3, 2015, pp. 415–429
- Ding, Y. and S.E. Fienberg, (1994). *Dual System Estimation of Census Undercount in the Presence of Matching Error*. *Survey Methodology*, vol. 20, pp. 149–158
- Fienberg, S., 1972, *The multiple recapture census for closed populations and incomplete 2k contingency tables*. *Biometrika*, vol. 59, 409-439.
- Gerritse, S. C., P.G.M. van der Heijden & B.F.M. Bakker, (2015a) *Sensitivity of Population Size Estimation for Violating Parametric Assumptions in Log-linear Models*. *Journal of Official Statistics*, vol. 31, no. 3, pp. 357-379 <http://dx.doi.org/10.1515/JOS-2015-0022>
- Gerritse, Susanna C. , Bart F. M. Bakker & Peter G. M. van der Heijden, (2015b). *Different methods to complete datasets used for capture-recapture estimation*, *Statistical Journal of the IAOS*, vol. 31, no. 4, pp. 613-627, 2015 (doi 10.3233/SJI-150938)
- Gerritse, Susanna C. , Bart F. M. Bakker, Daan B. Zult & Peter G. M. van der Heijden, (2016). *The effects of linkage errors and erroneous captures on the population size estimation* (submitted)
- Ghosh S.K., Norris J.L. (2004). *Bayesian Capture-Recapture Analysis and Model Selection Allowing for Heterogeneity and Behavioral Effects*, *NCSU Institute of Statistics, Mimeo Series 2562*, pp. 1-27
- IWGDMF (International Working Group for Disease Monitoring and Forecasting), 1995, *Capture- recapture and multiple record systems estimation. Part 1. History and theoretical development*. *American Journal of Epidemiology*, 142, 1059-1068
- ONS, 2012, *2011 Census: Over-count estimation and adjustment*. Available at <http://www.ons.gov.uk/ons/guide-method/census/2011/census-data/2011-census-data/2011-first-release/first-release--quality-assurance-and-methodology-papers/overcount-estimation-and-adjustment.pdf>
- Raftery, A.E., Li. N., Ševčíková , H., Gerland, P. and Heilig, G.K. (2012). *Bayesian probabilistic population projections for all countries*. *Proceedings of the National Academy of Sciences* 109:13915-13921.
- Statistics Canada, (2015). *Census Technical Report: Coverage*. Available at <https://www12.statcan.gc.ca/census-recensement/2011/ref/guides/98-303-x/index-eng.cfm>
- Steorts, Hall, Fienberg (2014) *SMERED: A Bayesian Approach to Graphical Record Linkage and De-duplication, a new approach in the Bayesian framework*, *Journal of Machine learning research*
- Van der Heijden, P.G.M., J. Whittaker, M. Cruyff, B. Bakker & R. van der Vliet, (2012). *People born in the Middle East but residing in the Netherlands: invariant population size estimates and the role of active and passive covariates*, *Annals of Applied Statistics*, vol. 6, no. 3, pp. 831-852
- Wolter, K.M. (1986). *Some coverage error models for Census data*, *Journal of the American Statistical Association*, vol. 81, pp. 338-346
- Zhang, L.-C., (2015). *On modelling register coverage errors*. *Journal of Official Statistics*, vol. 31, nr. 3, pp. 381-396
- Zhang, L.-C. & Dunne, J. (2017). *Trimmed Dual System Estimation*. In *Capture-Recapture Methods in Social and Medical Sciences* (eds. D. Böhning, J. Bunge and P. v. d. Heijden), Ch. 15, pp. 237-262.
- Zwane, E. N., Van der Pal-de, B., Van der Heijden, P.G.M., (2004). *The multiple-record systems estimator when registrations refer to different but overlapping populations*, *Statistics in Medicine*, 23, pp. 2267-228

2.3.6 Reconciliation

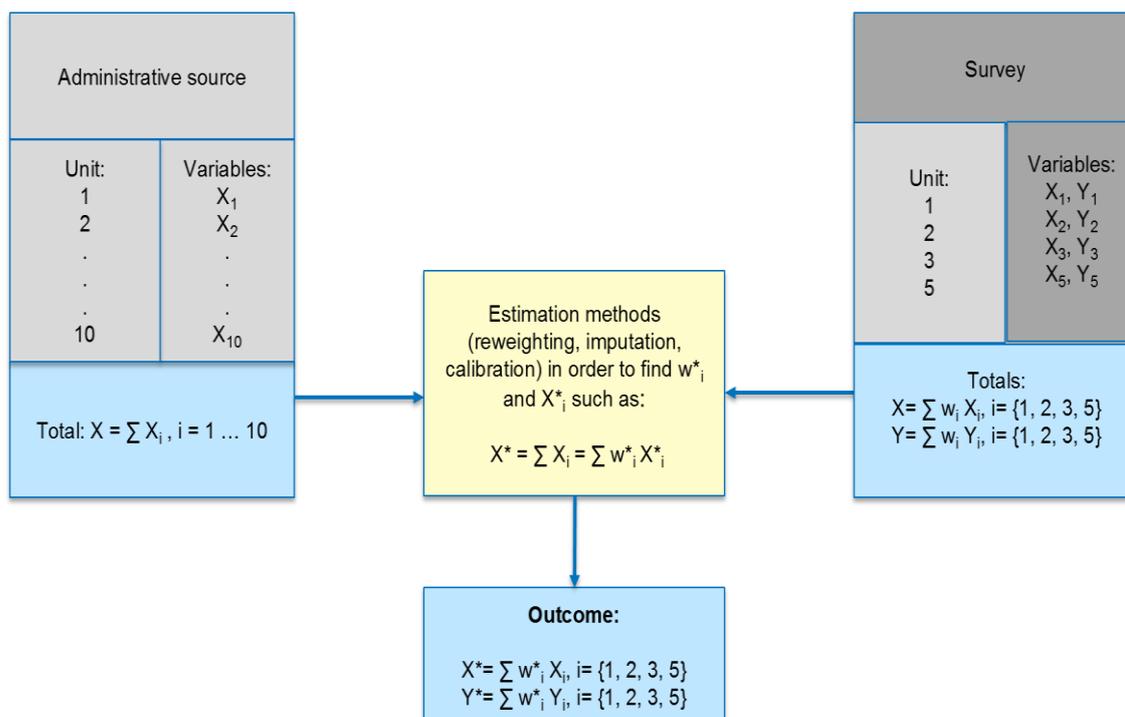
874. Different estimates for the same phenomenon could lead to confusion among users of these figures. Many NSIs have therefore adopted a one-figure policy. According to this one-figure policy, estimates for the same phenomenon in different tables should be reconciled, meaning that they should be equal to each other, even if these estimates are based on different underlying data sources.

875. When using a mix of administrative data sources and surveys on which to base estimates upon, obtaining one estimate for the same phenomenon may become problematic as for different (combinations of) variables data on different units, e.g. different persons, may be available.

876. This means that different estimates concerning the same variable may yield different results, if one does not take special precautions. For instance, if one uses a standard weighting approach to produce estimates, where one multiplies observed counts or values with surveys weights, one may get different estimates based on two samples, because different units and hence different survey weights are used in the two samples.

877. In principle, these differences are merely caused by “noise” in the data, such as sampling errors. So, in a strictly statistical sense, different estimates concerning the same variables are to be expected and are not a problem. However, different estimates would violate the one-figure policy and form a problem for the users.

Example of reconciliation:



878. The example in the figure above can be extended to the case where total by age group and by gender. In this case totals shall be estimated at the lowest level of aggregation (age group by gender) before proceeding incrementally to further estimations for each of the upper level of aggregations.

879. A similar problem affects estimates when a time component characterises data sources.

A common problem of macro-integration arises when times series data of the same target variable exist with different frequencies in different sources, e.g. quarterly (or yearly) administrative data vs. monthly survey data. The data of the lower frequency may be based on a larger or more reliable set of data, in which case one may fix these and adjust the frequency time series accordingly, where one would like to keep the adjustments (to be defined later) as small as possible.

List of Estimation methods:

880. When micro-data and aggregated data have to be estimated and reconciled, several methods are available, such as repeated weighting (RW), repeated imputation (RI), mass imputation and macro-integration.

a. Repeated weighting

881. In the RW approach a separate set of weights is assigned to sample units for each table of population totals to be estimated. In this approach the tables to be estimated are estimated sequentially and each table is estimated using as many sample units as possible in order to keep the sample variance as low as possible. The combined data from administrative data sources and surveys are divided into rectangular blocks. Such a block consists of a maximal set of variables for which data on the same units has been collected. The data blocks are chosen such that each table to be estimated is covered by at least one data block. Item non-response in a block is assumed to be treated beforehand by means of imputation.

882. How a table is estimated depends on the available data. Data from an available administrative data source covering the entire population can simply be counted or added. Data only available from surveys are weighted by means of regression weighting. In that case weights must be assigned to all units in the block to be weighted. For a survey one usually starts with the inverse inclusion probabilities of the sample units (i.e. the design weights), corrected for response selectivity. These weights are then further adjusted by calibrating them to previously estimated totals. For a data block containing the overlap of two surveys, one usually begins with the product of the standard survey weights from each of the surveys as starting weight for each observed unit, and then corrects these starting weights by calibrating to totals known from administrative data sources and previously estimated totals.

883. When estimating a new table, all margins of this table that are known or have already been estimated for previous tables are kept fixed to their known or previously estimated values, i.e. the regression weighting is calibrated on these known or previously estimated values. This ensures that the margins of the new table are consistent with previous estimates.

b. Mass imputation

884. In the mass imputation approach, one imputes all variables for all population units for which no value was observed. This leads to a rectangular data set with values for all variables and all population units. After imputation estimates of totals can be obtained by simply counting or adding the values of the corresponding variables.

885. The approach relies on the ability to capture all relevant variables and relevant relations between them in the imputation model, and to estimate the model parameters sufficiently accurately. Given that all relevant variables and all relevant relations among them can be captured accurately by the imputation model, the approach is very straightforward. However,

it is generally impossible to capture all relevant variables and relations in the imputation model, simply because there are not enough observations to estimate all model parameters accurately.

886. For rich data sets with many variables, especially if not all tables to be estimated are specified beforehand, it seems that mass imputation cannot be applied successfully. For data sets with a limited number of variables and where all tables to be estimated are specified beforehand mass imputation seems to a viable option, however.

c. Repeated imputation

887. RI is the equivalent of RW. The important difference is how estimates are produced: in the case of RW by means of a weighting method, in the case of RI by means of an imputation method. Like RW, RI is a sequential approach where tables are estimated one by one. For some variables in a table estimates may have already been produced while estimating a previous table. Similar to RW, these variables are then calibrated to the previously estimated totals.

888. A prerequisite for applying RI is an imputation method that succeeds in preserving the statistical aspects of the true data as well as possible, that is able to satisfy specified edit rules and that is able to preserve previously estimated totals.

889. For the time being, the RI approach has mainly been applied to small data sets, not on large data sets arising in practice. Evaluations on large data sets remain still to be carried out. Software for the RI approach is not yet generally available.

d. Macro-integration

890. Macro-integration is the process of reconciling statistical figures on an aggregate level. These figures are usually in the form of multi-dimensional tabulations, obtained from different sources. When macro-integration is applied, only estimated figures on an aggregated level are adjusted. The underlying micro-data are not adjusted or even considered in this adjustment process. The main goal of macro-integration is to obtain a more accurate, consistent and complete set of estimates for the variables of interest. Several methods for macro-integration have been developed, such as the methods by Stone, Champernowne and Meade (1942), Byron (1978), Sefton and Weale (1995) and Magnus, Van Tongeren and De Vos (2000).

891. Traditionally, macro-integration has mainly been applied in the area of macro-economics, in particular for compiling the National Accounts, for example to adjust supply and use tables to new margins (see Stone, Champernowne and Meade 1942). Also, applications in other areas have been studied, namely for the reconciliation of tables of Transport and Trade Statistics, for the Census, and for combining estimates of labour market variables. Possible application to refugee and IDP statistics are still being studied.

Possible example for IDP statistics drawn from two different regional registers:

Number of IDPs changing place of residence from previous region of residence to new region of residence reported by region of origin

Region of origin / Region of arrival	R1	R2	R3	Total (b)	Balance (b) – (a)
R1	n/a	6	10	16	6
R2	5	n/a	8	13	1
R3	5	6	n/a	11	-7
Total (a)	10	12	18	40	0

Number of IDPs changing place of residence from previous region of residence to new region of residence reported by region of arrival

Region of origin / Region of arrival	R1	R2	R3	Total (b)	Balance (b) – (a)
R1	n/a	4	8	12	2
R2	6	n/a	8	14	5
R3	4	5	n/a	9	-7
Total (a)	10	9	16	35	0

892. The starting point of macro-integration is a set of estimates in tabular form. These can be quantitative tables, for instance a table of the number of IDPs having change residence from a region to another one like in the two tables above. If the estimated figures in these tables are based on different sources (regional registers), and the tables have margins in common (red figure in the tables above), these margins are often conflicting because update of registers in the regions can vary. In this case, aggregated figures in red need to be balanced.

893. The macro-integration approach often consists in a constrained optimization problem (see, e.g., Stone, Champernowne and Meade 1942, and Byron 1978). A target function, for instance a quadratic form of differences between the original and the adjusted values, is minimized, subject to the constraints that the adjusted common figures in different tables are equal to each other and additivity of the adjusted tables is maintained. Inequality constraints can be imposed in these quadratic optimization problems.

e. Macro-integration with a time component

894. The general issue behind macro-integration with a time component occurs when both low frequency data (yearly data) and high frequency data (quarterly or monthly data) are available for the same variable. Lower frequency data is usually based on a larger or more reliable set of data, and those data is often kept fixed. High frequency data are then adjusted to them, where we would like to keep the adjustments (to be defined later) as small as possible. In the context of the combination of administrative data with survey data, an example is that a

nearly complete set of administrative data are available on a yearly basis whereas survey sample data on the same variable are available on a quarterly basis.

895. For macro-integration with a time component there are three different types of methods: benchmarking, temporal disaggregation and nowcasting.

896. **Benchmarking:** The methods of benchmarking concerns the situation where the high and low frequency series concern the same variable. The idea is that some difference between the original and the adjusted figures is minimized subject to equality or inequality constraints. Maybe the most basic method to deal with benchmarking is a method that tries to minimize the adjustments to the original level estimates for each of the time periods. When each of the level estimates are adjusted with the same relative factor, this is referred to as pro-rating. Pro-rating leads to the so-called step-problem. The step-problem means that the adjusted (benchmarked) series may have a disproportional large adjustment in the transition from one low-frequency period to the next. To avoid step-problems methods have been developed that take account of the changes in the original time-series.

897. One well-known method that tries to preserve the changes in the original high-frequency series is the growth rate preservation method (GRP) by Causey and Trager (1981). That method minimizes the squared difference between the original period-to-period change and the corresponding adjusted growth rate. Another approach is the movement preservation method (MP) by Denton (1971). The most commonly applied variants of this method minimize the squared differences between adjustments of two subsequent time periods, also referred to as first differences. This can be further refined into minimizing additive or relative differences. A modification of the relative difference model, as proposed by Cholette (1984), is most popularly applied in practise (often called modified Denton). Mathematically, MP is easier to apply than GRP, because MP deals with a standard linearly constrained quadratic optimization problem, while GRP solves a more difficult linearly constrained nonlinear problem (a ratio of two estimators) that can be efficiently solved by an interior-point-algorithm.

898. Note that benchmarking can also be applied to link two related time series, when there is a change of methods or revision in classifications between the two series.

899. **Temporal disaggregation:** Methods for temporal disaggregation are very close to those of benchmarking. Temporal disaggregation is used for the situation that one has a low frequency series of a target variable that one wishes to disaggregate into a higher frequency series, but there are no data valuable for the target variable on this high frequency. Instead, for the disaggregation one uses data of other variable(s) that one considers to be indicative of the high-frequency changes of the target variable. An example of such a method is given by the seminal paper of Chow and Lin (1971). This method minimizes the adjustments to the original level estimates for each of the time periods, so the disadvantage is that it leads to the already mentioned step-problem.

900. **Nowcasting:** An important problem with the estimation of the high-frequency series is that the benchmark of the low frequency series is not yet available when the preliminary estimates of the high frequency series are made, this is referred to as the timeliness issue. One of the means to cope with this timeliness issue is to pre-adjust the high-frequency series (e.g. Fortier and Quenneville, 2007). Another means to handle the timeliness issue is to use nowcasting. Nowcasting aims to provide an estimate for the present or recent past, based on a limited amount of data on the present. In particular, often information is used from sources that are available earlier, or with higher quality. Nowcasting can be done in many different ways,

for instance by using an ARIMA model or by using a structural time-series model (STM).

References:

- Boonstra, H.J., C.J. De Blois and G.J. Linders (2011), *Macro-Integration with Inequality Constraints an Application to the Integration of Transport and Trade Statistics*. *Statistica Neerlandica* 65, pp. 407-431.
- Byron, R.P. (1978), *The Estimation of Large Social Account Matrices*. *Journal of the Royal Statistical Society A* 141, pp. 359-367.
- Coutinho, W., T. de Waal and N. Shlomo (2013), *Calibrated Hot Deck Imputation Subject to Edit Restrictions*. *Journal of Official Statistics* 29, pp. 299-321.
- De Waal, T. (2016), *Obtaining Numerically Consistent Estimates from a Mix of Administrative Data and Surveys*. *Statistical Journal of the IAOS* 32 pp. 231–243.
- Houbiers, M. (2004), *Towards a Social Statistical Database and Unified Estimates at Statistics Netherlands*. *Journal of Official Statistics* 20, pp. 55-75.
- Kroese, A.H. and R.H. Renssen (2000), *New Applications of Old Weighting Techniques; Constructing a Consistent Set of Estimates Based on Data from Different surveys*. In: *Proceedings of ICES II*. American Statistical Association, Buffalo NY, pp. 831-840.
- Kroese, B., R.H. Renssen and M. Trijsenaar (2000), *Weighting or Imputation: Constructing a Consistent Set of Estimates Based on Data from Different Sources*. *Netherlands Official Statistics* 15, pp. 23-31.
- Pannekoek, J., N. Shlomo and T. de Waal (2013), *Calibrated Imputation of Numerical Data under Linear Edit Restrictions*. *Annals of Applied Statistics* 7, pp. 1983-2006.
- Särndal, C.E., B. Swensson and J. Wretman (1992), *Model Assisted Survey Sampling*. New York: Springer-Verlag.
- Shlomo, N., T. de Waal and J. Pannekoek (2009), *Mass Imputation for Building a Numerical Statistical Database*. UN/ECE Work Session on Statistical Data Editing, Neuchâtel, Switzerland.
- Stone, R., D.G. Champernowne and J.E. Meade (1942). *The Precision of National Income Estimates*. *Review of Economic Studies* 9, pp. 111-125.
- Whitridge, P., M. Bureau and J. Kovar (1990), *Mass Imputation at Statistics Canada*. In: *Proceedings of the Annual Research Conference, U.S. Census Bureau, Washington D.C.*, pp. 666-675.
- Whitridge, P. and J. Kovar (1990), *Use of Mass Imputation to Estimate for Subsample Variables*. In: *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, pp. 132-137.
- Cholette, P. (1984). *Adjusting sub-annual series to yearly benchmarks*, *Survey Methodology*, 10, 35-49.
- Chow, G.C. and A. Lin (1971), *Best Linear Unbiased Interpolation, and Extrapolation of Time Series by Related Series*. *Rev. Economics and Statistics* 53, pp. 372-375.
- Cholette, P. (1984). *Adjusting sub-annual series to yearly benchmarks*, *Survey Methodology*, 10, 35-49.
- Chow, G.C. and A. Lin (1971), *Best Linear Unbiased Interpolation, and Extrapolation of Time Series by Related Series*. *Rev. Economics and Statistics* 53, pp. 372-375.
- Daalmans J., di Fonzo, T., Mushkudiani, N. and R.P. Bikker (2016). *Removing the Gap between Annual and Sub-Annual Statistics based on Different Data Sources*. *Proceedings of the Fifth International Conference on Establishment Surveys, June 20-23, 2016, Geneva, Switzerland: American Statistical Association*.
- Dagum, E.B. and Cholette, P. (2006). *Benchmarking, Temporal Distribution and Reconciliation Methods for Time Series Data*. New York: Springer-Verlag, *Lecture Notes in Statistics*, volume 186.
- Denton, F.T. (1971), *Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization*. *Journal of the American Statistical Association* 66, pp. 99-102.
- Di Fonzo, T. and M. Marini (2005), *Benchmarking a system of Time Series: Denton's movement preservation principle vs. data based procedure*. University of Padova, Available at http://epp.eurostat.cec.eu.int/cache/ITY_PUBLIC/KSDT-05-008/EN/KS-DT-05-008-EN.pdf
- Durbin, J. and B. Quenneville (1997). "Benchmarking by state space models". *International Statistical*

Review 65, 23–48.

- *Harvey (1989), Forecasting, structural time series models and the Kalman filter, Cambridge University Press.*
- *Stone, R., D.G. Champenowne and J.E. Meade (1942), The Precision of National Income Estimates. Review of Economic Studies 9, pp. 111-125.*

2.6 Design production systems and workflow

‘This sub-process determines the workflow from data collection to dissemination, taking an overview of all the processes required within the whole statistical production process, and ensuring that they fit together efficiently with no gaps or redundancies. Various systems and databases are needed throughout the process. A general principle is to reuse processes and technology across many statistical business processes, so existing production solutions (e.g. services, systems and databases) should be examined first, to determine whether they are fit for purpose for this specific process, then, if any gaps are identified, new solutions should be designed. This sub-process also considers how staff will interact with systems, and who will be responsible for what and when’. (UNECE,2013)

2.6.1. Introduction

901. This section deals with fitting together the various components of the design into a coherent whole. It also includes the design of the quality assurance processes through which statistics should be evaluated for quality.

902. The gathering and harmonisation of data on displaced people are the key to international comparative analysis and policy work on forced displacement. The timely, accurate statistical information on IDP, asylum and refugee statistics enables a wide range of policy answer in the rapidly evolving forced displacement caused by humanitarian crises.

903. The objectives of this IDP, asylum and refugee-related statistical production system is to improve efficiency of data and metadata collection, validation, processing, storage and dissemination. It is also a target of a well-functioning statistical production system to improve quality by eliminating errors and incoherencies, and by shortening the statistical publication cycles. The accessibility and visibility of the asylum related statistical outputs should also be a key objective.

904. In order to achieve these targets a modern statistical production system should utilize of advances of modern information technologies and standards.

2.6.2. The statistical production environment

905. Before developing a statistical production system, the data sources of this system should be determined. Based on their supervising body in the national data production system, there are two types of data sources which can feed these systems.

906. The first type is part of the national statistical system, this is the administrative data source where the data owners are an integrated part of the national public administration. In this case the country's government with its public administration is present and a national legislation is available which regulates the asylum procedure or IDP registration. Parallel to the national legislation are specialized administrative bodies or custodian agencies responsible for the IDP or asylum procedure and for the collection of the administrative data related to the

procedures.

907. In this case the statistical data are by-products of the administrative procedure. In an optimal situation the IDP or asylum procedures are supported by IT-systems, and case-management systems where the data are stored and can be queried. This case-management system is interlinked with a statistical sub-system, it means that the case-management system should have an interface where the statistical queries can be set. In this production system chain the case-management system is the input for the statistical subsystem, it is not part of the statistical data production system.

908. The statistical sub-system should ensure the ideal technical environment for data production and for the metadata production environment as well. The statistical production system should have also a link to the data storage warehouse, which is a dedicated part of the statistical production system. In the storage system the data are stored, if possible, in so called cubes, these cubes should have a link to the dissemination system which enables the statistical data publication via user interfaces. According to the methodology of the data inquiry we can differentiate between static systems where the data can be requested via static tables e.g. Excel tables or html tables. The other way is to enable the compilation of user-friendly, interactive statistical tables.

909. The second type of data source describes an IDP or asylum system where the national legislation is not so well developed as in the first case, and a dedicated governmental custodian body for the asylum procedure is not functional. Refugee status-determinations can be undertaken by UNHCR in contrast to status determination carried out by a government. In the latter case the responsible agency for the asylum procedure is an international UN organisation such as UNWRA or UNHCR. The data source is the administrative system of this international organization which collects data related to the IDP or asylum procedure and also disseminates statistics at the international level.

910. If such an asylum related statistical system is available, the international organization serves as the producer of the input for the national public administration. A memorandum of agreement could enhance the cooperation between organizations of public administration and the UNHCR which is directly involved in providing shelter for asylum seekers and refugees and in data collection (see Part 2 *UNHCR Agreements and Memoranda of Understanding* for an example of this).

‘UNHCR collects data on persons in concern i.e. refugees, asylum-seekers, internally displaced persons (IDPs), returnees and stateless persons. The data collection includes information on country of origin and of asylum, sex age and location or settlement such as camp or centre. Furthermore, information about protection and living conditions is also gathered with variables such as education, health, nutrition as well as information regarding sexual and gender-based violence.’⁸⁹

911. The main obstacle to data collection is the lack of harmonised definitions and measurement criteria which can occur in cases when the country did not comply with the 1951 Geneva Convention relating to the Status of Refugees (See IRRS Chapter 2), or in the case of

⁸⁹SIGMA0108, The UN Refugee Agency – global provider of asylum data, <https://www.unhcr.org/en-my/4868e61e4.pdf>, downloaded: 29th May 2019

IDPs observe the *Guiding Principles on Internal Displacement*⁹⁰ The lack of harmonisation can be especially problematic if the translation of definitions of the convention into national law is not done in a systematic and harmonised way, and not in observance of the Recommendations set out in the IRRS or IRIS. Other factors which can make the data collection difficult are differences in quality and periodicity.

2.6.3. The statistical database environment

912. When establishing a statistical production system there are also some criteria related to the technical details and functionalities of the system which should be analysed more detailed. It is recommended to develop a common hosting environment for production databases, the idea behind the common production system is that it strengthens standardisation of the statistical data production and it minimises the number of tools used in statistical activities and reduces training and support requirements.

913. It is recommended that the statistical data production system should include a toolkit which enables the management of statistical data, this toolkit should include different facilities, such as:

- data migration⁹¹,
- security management,
- database administration (management of user access rights)
- data collection,
- data validation,
- data-query toolkit,
- metadata management system.

914. When planning a statistical data production system, it is essential to design a centralized metadata storage system which enhances the efficiency of metadata preparation, storage, access, management and dissemination. The centralized metadata system addresses the problems of fragmented metadata located in numerous databases. The metadata sub-system of the statistical production system should meet some basic expectations, such as:

- the metadata system should have a common structure, defining common metadata items,
- all statistical data should have appropriate metadata,
- it should cover a broad group of user needs,
- the metadata must be consistent across the different subject matter areas.
- The metadata system should accommodate any kind of metadata related to the corresponding statistical data.

⁹⁰ E/CN.4/1998/53/Add.2, 11 February 1998, UN Commission on Human Rights, Report of the Representative of the Secretary-General, Mr. Francis M. Deng, submitted pursuant to Commission resolution 1997/39. Addendum: Guiding Principles on Internal Displacement, available at: <https://goo.gl/GJU2Sz>

⁹¹ Source: OECD Statistical Information System (SIS) <https://www.oecd.org/sdd/oecdstatisticalinformationsystems.htm>, downloaded: 29th May 2019

Eurostat's Single Integrated Metadata Structure (SIMS)⁹²

The Single Integrated Metadata Structure is the dynamic inventory of statistical concepts used for quality and metadata reporting in the European Statistical System.

The Single Integrated Metadata Structure streamlines and harmonises metadata and quality reporting in the European Statistical System (ESS) and it decreases the reporting burden on the statistical authorities by creating the framework for “once for all purposes”. It is an integrated quality and metadata reporting framework, which ensures a flexible and up to date system where future extensions are possible.

The system enables that the same concept names and the same quality indicators are used in the different ESS metadata structures, it also enhances the consistency with the SDMX statistical standards as it stands in the SDMX Content-oriented Guidelines.

It has to be underlined that data users and data producers have different needs on statistical information and this has to be reflected by the quality reports that are addressed to them. The distinction between the user and producer oriented quality reporting is assured through the Single Integrated Metadata Structure.

Based on this approach there are two different quality reporting systems available within the SIMS, the user-(U) and producer (P) –oriented quality reporting. The short user-oriented quality report is implemented through the improved visibility and readability of the quality related concepts and the producer-oriented quality report (P) is implemented via the ESQRS⁹³ report structure.

2.6.4. Statistical dissemination process

915. It is essential that the design of the statistical production system (the statistical production database) should interlinked with the dissemination system, this system enables the establishment of a modernised tool for traditional statistical publications and interactive data products.

916. For the dissemination system the data source is the statistical production system, the production database, it increases the efficiency of statistical dissemination and reduces the risk of mistakes, enables more timely publication of statistics.

917. The improved statistical dissemination system ensures a common layout and reduces

⁹² Source: European Commission, Technical Manual of the Single Integrated Metadata Structure (SIMS): <https://ec.europa.eu/eurostat/documents/64157/4373903/03-Single-Integrated-Metadata-Structure-and-its-Technical-Manual.pdf>, downloaded: 03rd June 2019

⁹³ ESS Standard for Quality Reports Structure

the number of different software applications in use. The recommended statistical publication tool should have an interface for managing publication contents definition and it is also necessary to have an interface for generating XML output file in order to combine statistical data and metadata.

918. A workflow model should be developed which monitors the complex statistical data production from the production of statistical data to the production of metadata, through the storage of the statistical data. This complex model also monitors the statistical channels used for monitors.

2.6.6 Indicator system for measuring the design production systems and workflow⁹⁴

919. In order to measure the effectiveness of the statistical production system and workflow an indicator system should be established which can measure the effectiveness of the statistical production with indicators like, soundness of the implementation, the cost effectiveness, accuracy and reliability, timeliness and punctuality and last but not least accessibility and clarity.

920. A suitable indicator system can be seen in the table below, in the following this indicator system is presented: the indicator soundness of implementation measures the quality of the workflow in the design and production system phase of the GSBPM structure. It identifies the different sub-processes being part of this phase. The soundness of implementation should take into consideration the types of the data which are processed (admin data or survey type data) in this phase, because it affects the specifications of the indicator system.

921. It should be also outlined that the cost of designing a statistical production system plays an important role, it is necessary to reduce the costs of the different software tools. In order to achieve this goal, the best solution is the use of common applications. The products should be disseminated (see also Phase 7 *Dissemination*) on an appropriate platform, e.g. on the home page of the body responsible for statistical data production which can also increase the costs of the statistical production but using optimal software and technical solutions this cost can be reduced.

922. The indicator 'accuracy and reliability' measures the ratio of quality indicators used as performance indicators, the source data, integrated data, intermediate results and statistical outputs should be regularly assessed and validated.

923. The next indicator, 'timeliness and punctuality' defines the time frame which is available for the production in the different sub-processes.

924. According to the last indicator in the list, 'accessibility and clarity', the number of visitors of the home page or other social media products of the statistical data producer is calculated. During the performance of this statistical data production system a quality indicator system can ensure a permanent feedback on the production system.

⁹⁴ Source: Modernstats; Quality Indicators for the Generic Statistical Business Process Model (GSBPM) - For Statistics derived from Surveys and Administrative Data Sources
https://statswiki.unece.org/download/attachments/185794796/Quality%20Indicators%20for%20the%20GSBPM%20-%20For%20Statistics%20derived%20from%20Surveys%20and%20Administrative%20Data%20Sources_Final.pdf?api=v2, downloaded: 29th May 2019

Quality indicators for measuring the design of production systems

Quality dimension	Indicator
Soundness of implementation	Percentage of identifies and documented GSBPM processes
	Specifications for production systems and workflow take into consideration the type of data being processed (survey data or data from administrative data sources)
Cost effectiveness	Percentage to which corporate solutions (e.g. tools, processes, technologies) are reused in subsequent phases
	Percentage to which responsibilities for subsequent phases and sub-processes have been set
	Estimated cost for producing and disseminate designed outputs
Accuracy and reliability	Percentage to which quality indicators are planned to be calculated for subsequent sub-processes of GSBPM
	Percentage of quality indicators used as performance indicators
Timeliness and punctuality	Planned time frame for subsequent phases and sub-processes
Accessibility and clarity	The number of visitors of the home page and on the social media
	Number of metadata consultations
	The number of consultations of data tables

Phase References and Links

- Online Template for metadata documentation (European SIMS system); <https://ec.europa.eu/eurostat/data/metadata/metadata-structure>
- Links to automated quality check for integrated data “Holoclean” software; <http://www.holoclean.io/>. See also Phase 5.1 of this manual.
- Metadata example: <https://www.ssb.no/en/befolkning/statistikker/flyktninger>
- Guidance for Data Integration for Measuring Migration; <http://www.unece.org/fileadmin/DAM/stats/publications/2018/ECECESSTAT20186.pdf>

PART III – PHASE 4 – COLLECT

925. This section discusses a typical Training needs plan and is a continuation of Phase 4.2 and links to *paragraph 34343*.

a. Training Needs Plan

926. The **training activities** need to be tailored to the role of each of the field teams;

- i. THE SURVEY MANAGER AND FIELD MANAGER are assumed to be professionals already knowledgeable about surveys in general. Thus, the only training they need is in the special characteristics survey. This training occurs on the job as they prepare the survey in collaboration with experts on displacement statistics in other countries.
- ii. For the DATA MANAGER, the training needs are usually more specific but are also accomplished on the job. About two to four weeks of close collaboration with people who have developed and applied integrated data management techniques for other complex surveys are usually required. Training is both theoretical and practical. The conceptual framework includes criteria for survey data consistency, error levels (range checks and consistency checks), design of a dictionary of variables, file management for data entry and analysis, and questionnaire design techniques for effective data management. The practical part of the training consists of translating the questionnaire structure into either a set of linked data entry screens, graphically designing several of these screens, and defining the most important range and consistency checks for CAFÉ or selecting the best CAPI software for the particular survey conditions. Thus, during the training, part or all the data entry program is prepared.
- iii. SUPERVISORS. Some Supervisors will be trained on the job, as they will take part in the project from the early stages of field testing and will actively participate in preparations for the survey. However, some aspects of the job should be presented formally, through training sessions and in the Supervisor manual. These include: sample design, contents and design of the survey, structure of the interviews, structure of the management team, structure of the field teams, quality control criteria, coding, and household replacement criteria.
- iv. INTERVIEWERS AND DATA ENTRY OPERATORS. There should be sufficient time to comprehensively train the Interviewers and, if used, Data Entry Operators. A two or three-day event is never enough time to thoroughly prepare the field workers. The training period for LSMS surveys is much longer than for other surveys (which tend to average less than one week) for two reasons. Training of all personnel is key to the process of reducing non-sampling errors. Second, CAPI questionnaires are far more complex than paper questionnaires, so more training is required to achieve a given level of understanding. The training must cover the basic structure of how to understand and use the questionnaire, but it must go much further.

927. **CAPI training** should target three different audiences, most often separately, so that each group knows how to use their tools to perform their role.

928. Interviewers should be trained, not just on the content of the questionnaire, but also on how to use the CAPI instrument for data collection. Early introduction to the CAPI questionnaire is recommended to ensure adequate familiarity with the devices, the software, and the questionnaire. Plenty of time (at the very least one day of field practice) should be allotted to hands-on practice with the CAPI instrument via self-completed questionnaires or mock interviews. By the end of the training, the Interviewer should be well-versed in all stages of the survey questionnaire process.

929. Some CAPI software have built-in supervisory capabilities, and when this is the case, Supervisors should be properly trained on the part of the CAPI software that allows them to effectively monitor the progress and quality of their team's work. Ideally, Supervisors should also be well trained on conducting interviews with CAPI to ensure they can help Interviewers deal with potential issues in the field.

930. **Staff at headquarters** should be trained on how to make survey assignments, monitor progress, and assess data quality. To troubleshoot potential data collection problems effectively, they also need to have a reasonable understanding of Interviewer mode and Supervisor tools. To achieve this objective, the headquarters team should be part of all survey activities, from initial CAPI application development, through pilots and trainings, and on to the main data collection. This way, the headquarters team can progressively gain familiarity with all the components of the CAPI software and be exposed to the potential problems that could arise during the main data collection.

931. **The training programs for Data Entry Operators and Interviewers** should be coordinated when using paper questionnaires. When fine tuning the training program, a common session for both Interviewers and Operators may be useful. This reflects the importance of doing the actual required tasks as part of the training and the fact that the work of all staff will be coordinated once the survey is fielded.

932. In order to ensure that uniform criteria and instructions are conveyed, surveys should strive for centralized training. This is also a reason to keep the number of teams small. This entails close coordination and monitoring of the different lecturers. In more extreme situations with large numbers of Interviewers, decentralized training is the only possible option. In these cases, training must be done in two steps: A group of trainers is trained first, so that they can later train others in different locations. All these factors should be carefully considered when planning a course of training because of the need for suitable rooms, audiovisual equipment, and so forth. Other logistical arrangements for training include lodging and transportation for the trainees who come from outside.

933. The training plan should emphasize practice interviews with households. This is the only way to discover whether Interviewers are really understanding what they are meant to learn. Not even practice interviews with each other will be as useful. Moreover, the Interviewers will at first be rather shy with households and need time to get over that before the survey starts. Parts of the practice interviews should be observed by trainers, their assistants, or Supervisors, to help detect where the Interviewers are having problems.

934. Depending on the number of field staff to be trained, the experience and skills of the core staff, and language constraints, the training may be conducted by the core staff,

international technical assistants, or a combination of the two.

935. The data entry program should be close to its final form at the moment of training, though fine-tuning and debugging are almost always necessary during the training period because the data from the actual questionnaires completed by the trainees reveals situations that were not foreseen during program development.

936. The importance of Interviewer training can hardly be overstressed. It is important that the Interviewer teams understand both the role of the questions in analysis and the role of the Interviewers (to administer the questionnaires as designed).⁹⁵

4.2.4 Training Manuals

937. The main written materials used for training field teams (Supervisors, Interviewers, and Data Entry Operators) are the questionnaires and the field manuals. It is recommended to produce many more manuals than needed for training because apart from their obvious use as a support for field operations the manuals are also valuable tools for the survey analysts.

938. The basic contents of each kind of manual is described in the following paragraphs.

1) SUPERVISOR MANUAL.

939. This manual should start by making explicit the objectives, methodology, and organization of the survey. It should then specify the Supervisor's responsibilities, duties, and the way the Supervisor should be connected to the survey's core management team and to the statistical agency's regular organization.

940. Another chapter of the manual should be devoted to the procedures to be carried out in each cluster, including the public relations tasks needed to ensure cooperation from the local authorities and the selected households. The difficulties in locating the selected households and ways to deal with refusals and other forms of non-response (as well as selection and documentation of replacements) should also be made clear.

941. Sections in the Supervisory manual should address the relationship of the Supervisor and the Interviewers, including procedures for preparing questionnaires for both survey rounds, and the use of the supervision forms for Interviewer evaluation, questionnaire verification, and check-up interviews. The latter should include detailed instructions on dealing with problems that might be found.

942. The manual should also specify procedures for coding any open questions in the questionnaires, including the codes needed to do so.

943. An important part of the Supervisor manual should be devoted to data entry when paper questionnaires are used. It should explain how and when the questionnaires are to be given to the Data Entry Operator and how to interpret the data entry printouts along with the rest of the inconsistencies signalled by the operator on the questionnaires. The manual should also explain how the data entry diskettes are to be sent to the survey core management team - by

⁹⁵ Both the data entry program and good supervision from the central office should help to detect such mistakes early in the survey and correct them before data collection gets very far.

Wi-Fi, by USB key, etc.

2) INTERVIEWER MANUAL.

944. The fundamental objectives of the Interviewer manual are to provide concepts and definitions, to define field procedures, and to ensure uniform criteria in the few parts of the questionnaire that are not self-explanatory. The manual should include general sections on the survey's objectives and methodology, the attitudes and behavior expected from the Interviewer, the relationship between the Interviewer and the Supervisor, the structure of the questionnaire, the conventions used in the questionnaire's design and interpretation of the data entry program outputs, and specific sections on each module of the questionnaire.

945. In many surveys the Interviewer manual contains a list of all questions in the questionnaire, with detailed instructions on how to ask and record the answers to every single one (for instance, 'Question 4 (Gender). Record the gender of the respondent, using code '1' for male and '2' for female", and so forth). Such an exhaustive approach can be both tedious and useless given the length of the questionnaires and the fact that they are pre-coded and have explicit skip patterns. Instead, the manuals should focus on clarifying important concepts that the Interviewers need to understand.

3) DATA ENTRY OPERATOR MANUAL.

946. When paper questionnaires are used, this manual should explain in great detail the role of the operator in the field operational setup and how outputs of the program (e.g., on-line messages and printouts) are to be transferred from the operator to the team Supervisor. Contrary to what might be expected, this manual needs to make very little reference to the computer or the data entry program. The use of the latter should be intuitive enough not to need further explanation.

4.3.2 Developing Supervision Forms

947. Three of the tasks of the team Supervisors should be supported by written documents, known as supervision forms. These are (1) Interviewer evaluation, (2) questionnaire verification, and (3) check-up interviews. The forms are intended to give these tasks formal definition, as opposed to loosely defined responsibilities left to the Supervisor's personal initiative, and to make it possible to supervise the Supervisors themselves (e.g., make supervision tasks verifiable by the survey core staff). Guidelines for the design of these forms are given below, with examples taken from the Pakistan Integrated Household Survey.

948. **Interviewer Evaluation.** The purpose of Interviewer evaluation is to monitor the performance and attitudes of the Interviewers. At least once a week (more often for weak Interviewers), the Supervisor should sit in on interviews conducted by each of the Interviewers in order to observe that they are administering the questionnaire correctly.

949. The Supervisor witnesses an Interviewer strictly as an observer and should not talk to the Interviewer or the respondent. The Interviewer should be informed that he or she is not allowed to ask for advice during the interview and that the Interviewer must behave as though the Supervisor were not present. The Interviewer evaluation form allows the Supervisor to make notes on any questions or concepts that the Interviewer may have difficulty asking or understanding. It should be filled in on the spot before the details of the interview are forgotten.

950. Interviewer evaluations also offer the chance to spot weaknesses in the questionnaire and suggest improvements for future versions. The form might also contain space for making note of problems or difficulties in the interviewing process, particularly with respect to inappropriately worded questions, concepts that are unclear to the respondent, or questions that are not answered because they are too personal or too sensitive.

951. **Questionnaire Verification.** The purpose of this operation is to ensure that the paper questionnaire is completely filled out; that is, that everyone who should have been interviewed has replied and that every section is complete. Verification should be done after the paper questionnaire is completed, before the Supervisor leaves the area and before the questionnaires are given to the Data Entry Operator.

952. A Questionnaire Verification Form should be designed to assist the Supervisor in this task. It should be filled out for all questionnaires after each round of the survey. If problems are found in the questionnaire, it should be returned to the Interviewer with instructions to correct them immediately before leaving the area.

953. Questionnaire verification is not supposed to replace exhaustive quality controls to be performed later by the data entry program, but rather serves as an early warning of major omissions that could be amended by sending the Interviewer back to the household before the team leaves the area.

954. Typical items to be considered in questionnaire verification are:

- **MANDATORY SECTIONS.** Some sections, such as socio demographic characteristics and housing, should be present in all questionnaires. Other sections may be present in certain locations but not in others.
- **COMPLETENESS OF INDIVIDUAL SECTIONS.** Depending on age, sex, or some other characteristic, certain sections of the questionnaire may or may not have to be completed. For instance, all women 15 to 49 years old, but no men, should answer the fertility section.
- **COMPLETENESS OF LISTS.** If the exhaustive approach is used to scan item lists in certain sections of the questionnaire, then all Yes/No questions should be completed, and a series of answers should follow each item marked "Yes."
- **FILTER QUESTIONS AND OTHER MAJOR SKIPS.** Some questionnaire sections are headed by "filter" questions, which indicate if the section is applicable or not to a particular household. The questionnaire should be consistent about the structure and use of skip patterns.
- **CHECK-UP INTERVIEWS.** The purpose of the check-up interview is to confirm that the Interviewer is indeed interviewing completely and accurately. The check-up interviews convey the importance of accuracy and completeness to the Interviewer. This reinforcement is important in maintaining high standards, even among diligent Interviewers. The check-up interview may also reveal any unsatisfactory Interviewers so that corrective action may be taken. These random revisits tend to be ignored or neglected. However, they are the best way to ensure effective interviews and are a standard procedure in all serious marketing research surveys.

955. It is generally considered acceptable to conduct check-up interviews in 15 to 25 percent of the households. The check-up interviews should not take longer than 15 minutes. It should be kept in mind that a difference in a response from the re-interview and a response from the original interview does not necessarily mean that the Interviewer is not doing a careful job. Respondents may provide different information at different times and sometimes the respondents contacted by the Interviewer and the Supervisor may not be the same. However, numerous differences indicate the need to follow up with the Interviewer regarding possible causes.

956. Supervisors fill out a check-up interview form to document the results of the re-interview. This ensures that the double-checking is thorough and impartial. It also allows the headquarters staff to effectively supervise the Supervisors. The most important things to control for in the check-up interview form are the questions for which certain answers may represent substantial differences in interview time later. "Rounding up" a woman's age a little bit may make her ineligible to answer the fertility section. More subtle omissions are not considering an illness serious enough to be reported in the health section, nor the purchase of small amounts worth the inclusion of certain items in the list of expenditures. Apart from those, the check-up interview form may include certain observational records (like some the building materials in the housing section) and other questions that are deemed unlikely to change between the interview and the check-up interview.

957. Interviewers should be made aware that some check-up interviews will take place, though of course they should never know in advance the households where these will be conducted. In marketing research surveys, which are always brief field operations, it is also considered that the contents of the check-up interview (that is, the questions that will be re-asked) should be kept secret from the Interviewers.

4.3.5 Duties of Survey Team Members

958. The composition of a typical survey team is described below;

1. **SURVEY MANAGER.** The Survey Manager should have decision making authority. This person coordinates the questionnaire's design, maintains communication with the technical assistance suppliers and data users, sets up the activities leading up to the survey in liaison with the existing statistical structures, and manages the implementation of the survey itself, and ensures that data documentation and dissemination procedures are put in place.
2. **DATA MANAGER.** The Data Manager designs and develops the data entry program and has input into the data entry aspects of the questionnaire design. This person writes the data entry manuals, selects and trains the Data Entry Operators when CAFÉ is used provides specialized training for Interviewers using CAPI techniques, prepares the data bases for analysis, and helps prepare tabulations and graphs for the analyses.
3. **FIELD MANAGER.** The Field Manager designs and supervises the sampling procedures and the household listing operation and is responsible for preparing the pilot survey and the field test. This person also designs the field operational procedures and the field manuals and is responsible for selecting and training the field staff. When the survey is fielded, this person implements the central supervision of the teams. This includes reviewing the various written supervision

instruments described below and occasionally conducting the same kind of observation and double-checking of interviews as the field Supervisors.

4. SUPERVISOR. As the primary person responsible for the quality of information collected in the field, the Supervisor is the most important member of the field team. In order to manage the field work effectively, the Supervisor must have a thorough understanding of the tasks required of each member of the team. The Supervisor should be able to respond to specific interviewing problems that may arise in the field and may on occasion need to perform interviews personally if any of the team's regular Interviewers falls ill or becomes otherwise unavailable. The Supervisor's main responsibilities include:
 - i. OVERALL FIELD SUPERVISION, COORDINATION, AND MONITORING OF DATA COLLECTION ACTIVITIES. An important part of this task is coordinating the work of the Interviewer in each household and of the male and female Interviewers when staff of both genders are needed. This is particularly important when an exchange of questionnaires between them becomes necessary. In addition, the Supervisor may on occasion have to assist the Interviewers in locating households and ensuring their willingness to respond to the survey. If necessary, Supervisors will select replacement households in line with criteria determined for the survey as a whole by the central managers.
 - ii. PUBLIC RELATIONS. The Supervisor should establish contact with local authorities in each area visited by the survey, and deliver letters of introduction, specially prepared brochures, and any other materials and information that might be necessary to ensure their cooperation.
 - iii. PREPARATION OF THE QUESTIONNAIRES. The Supervisor ensure that each Interviewer has the materials to complete the questionnaire in each household.
 - iv. COMPLETING ANY SUPPLEMENTAL QUESTIONNAIRES. The Supervisor must complete any supplemental questionnaires that are done during the field work. This may include questionnaires that look at facilities used by the respondents; or questionnaires for the communities in which the population of interest are located.
 - v. MONITORING, REVIEW, AND EVALUATION OF THE QUALITY OF FIELD INTERVIEWS. The Supervisor is expected to routinely observe interviews without advance notice to the Interviewer. The Supervisor should give immediate feedback based on established criteria for evaluating Interviewers.
 - vi. QUALITY CONTROL OF COMPLETED PAPER QUESTIONNAIRES.⁹⁶ Once data are collected for each round of the survey, the Supervisor should check that the Interviewer's writing is legible, skip patterns were followed,

⁹⁶ CAPI questionnaires have this quality control built into the programming of the questionnaire.

and the instructions in the questionnaire were followed.

- vii. **CHECK-UP INTERVIEWS.** The Supervisor should also revisit randomly selected households in each location to verify that the Interviewers have visited the household and to cross-check some of the information provided by the household.
 - viii. **CHECKING THE DATA ENTRY PRINTOUTS.** The Supervisor should compare the printout with the data on the paper questionnaire and should check errors in data that were detected by the data entry program. Either the Supervisor or one of the Interviewers should revisit the household, if possible, to correct the errors.
 - ix. **MANAGEMENT OF PERSONNEL, EQUIPMENT, AND VEHICLES.** The Supervisor is responsible for managing the team's support staff (i.e., the Data Entry Operator and the driver). The Supervisor should ensure that the staff work efficiently to provide efficient and trouble-free data collection and be responsible for the proper handling and care of computer equipment and vehicles. In certain cases, the Supervisor may be also be responsible for managing the team's finances, including the monthly payroll of salaries and bonuses.
 - x. **EXCHANGE OF INFORMATION BETWEEN CENTRAL SURVEY STAFF AND FIELD TEAMS.** As the main channel of communication, the Supervisor sees to it that any advice or instructions from the central management staff is relayed to and followed by the field team and that the central staff is regularly informed of the progress of data collection.
5. **INTERVIEWERS.** The Interviewer's main responsibilities include:
- i. **ESTABLISHING CONTACT WITH THE HOUSEHOLDS.** With the help of the Supervisor, the Interviewer must first introduce him or herself to each household and explain the survey's objectives and methodology in simple terms. The Interviewer should explain that the household was selected at random, along with many other households in the country, to help planners understand the people's living conditions. It should then be made clear that all information will be kept confidential.
 - ii. **SELECTING INDIVIDUAL RESPONDENTS.** The Interviewer should try by all means to interview each adult member personally, if possible, in private. This may require visiting the household several times during the survey period or going to the farm or place of business of the respondent.
 - iii. **CONDUCTING THE INTERVIEWS.** The Interviewer should conduct the interview in accordance with good survey practice. For example, they should be polite and exhibit a neutral attitude toward whatever answers the respondent gives. They should respect the wording on verbatim questions and follow the flow dictated by the skip pattern, without any variation.

- iv. **PROBING.** Probing for responses may be necessary, either because of explicit instructions in the questionnaire or in order to help respondents when they cannot answer exactly. The latter may be necessary, for instance, to record approximate birthdates.
6. **DRIVER.** The driver not only provides transportation from headquarters to the location itself but will help ferry team members to the various households, farms, and markets they may need to visit. When these are spread out and the Supervisor and Interviewers all need to visit two or more places in one day, the driver will be kept very busy.
7. **DATA ENTRY OPERATOR.** The Data Entry Operator enters the data from each round of the interviews immediately following data collection when paper questionnaires are used. This person revises all errors and inconsistencies flagged by the data entry program, corrects those that are an outcome of his or her own mistakes or omissions, and produces printouts with the rest of the errors so that they can be reviewed by the Supervisor in time for corrections to be made in the field.

ENDNOTES

REFERENCES

- Aalandslid, V., Lund, K.-A., & Berglund, F. (2014). The Population Census as a Tool to Capture Statistical Information on Forcibly Displaced Persons. *UNHCR Statistical Yearbook*, 14th Edition, 21-25.
- Benček, D., & Strasheim, J. (2016). Refugees welcome? A dataset on anti-refugee violence in Germany. *Research & Politics*, 3(4).
- CCSA. (2013). *Principles Governing International Statistical Activities*.
- Da Costa, R. (2006). Rights of Refugees in the Context of Integration: Legal Standards and Recommendations. *Legal and Policy Research Series (POLAS/2006/02)*.
- Donatiello, G., D'Orazio, M., Frattarola, D., Rizzi, A., Scanu, M., & Spaziani, M. (2014). Statistical matching of income and consumption expenditures. *International Journal of Economic Sciences*, 3(3), 50.
- D'Orazio, M., Di Zio, M., Scanu, & Mauro. (n.d.). *Statistical Matching: Theory and Practice*. Wiley.
- Dupriez, O., & Boyco, E. (2010). Dissemination of Microdata Files: Principles, Procedures and Practices. *IHSN Working Paper(005)*.
- Edmunds, R. (2005). *Models of Statistical Systems*. PARIS21/OECD.
- EU. (2004, November 19). Brussels: Justice and Home Affairs.
- Eurofund. (2016). *Sixth European Working Conditions Survey – Overview Report*. Luxembourg: Publications Office of the European Union,.
- Eurostat. (2011). *European Code of Practice*.
- Eurostat. (2014). *Technical Guidelines for the Data Collection Under Art. 6 of Regulation 862/2007 - Statistics on Residence Permits*.
- Government of Canada. (2014, October 9). *Directive on Open Government*. Retrieved September 6, 2017, from Government of Canada: <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=28108§ion=html>
- Hou, F., & Chen, J. (2002). Unmet needs for health care. *Health Reports*, 13(2), 23.
- Hughes, C., Zagheni, E., Abel, G., Wiśniowski, A., Sorichetta, A., Weber, I., & Tatem, A. J. (2016). *Inferring migrations, traditional methods and new approaches based on mobile phone, social media, and other big data: Feasibility study on Inferring (labour) mobility and migration in the European Union from big data and social media data*. European Commission.
- ISI. (1985). *ISI Declaration on Professional Ethics Adopted in August 1985*. Retrieved September 6, 2017, from ISI: <https://www.isi-web.org/index.php/news-from-isi/151-ethics1985>.

- Jacobsen, K. (n.d.). The forgotten solution: local integration for refugees in developing countries. *New Issues in Refugee Research*(45).
- Juran, S., & Snow, R. (2017, January-March). The potential of the 2010 population and housing census round for international migration analysis. *Migration Policy Practice*, VII(1).
- Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons Inc.
- Medstat-program . (2013). *Model Questionnaires (questionnaire N°5 on forced migrants) MED-HIMS Surveys, Manual 1*.
- Medstat-program. (2013). *Sampling Plans for MED-HIMS Surveys*.
- OECD. (2002). *Measuring the Non-Observed Economy: A Handbook*. OECD.
- OECD. (2007). *Jobs for Immigrants (Vol. 1): Labour Market Integration in Australia, Denmark, Germany and Sweden*. Paris: OECD Publishing.
- OECD/EU. (2015). *Indicators of Immigrant Integration 2015: Settling In*. Paris: OECD Publishing.
- PARIS21. (2017). *Guidelines for the elaboration of a National Strategy for the Development of Statistics*.
- Rea, A., & Wets, J. (2014). *The Long and Winding Road to Employment An Analysis of the Labour Market Careers of Asylum Seekers and Refugees in Belgium*. Brussels: Academia Press.
- Telford, J., & Dobak, J. (2016). Understanding Australian Migrant Outcomes : Making better use of administrative data. *United Nations Economic Commission for Europe (UNECE) and Eurostat work session on Migration Statistics*.
- Toga, A., & Dinov, I. (2015). Sharing big biometric data. *Journal of Big Data*, 2(7).
- UN Global Pulse. (2014). Estimating Migration Flows Using Online Search Data. *Global Pulse Project Series, 4*.
- UN Statistical Commission. (2012). *Guidelines for the Template for a Generic National Quality Assurance Framework*.
- UN/ECE Statistical Commission. (2000). *Terminology on Statistical Metadata*. Geneva: United Nations.
- UNECE. (2007). *Managing Statistical Confidentiality and Microdata Access - Principles and Guidelines of Good Practice*. New York and Geneva: United Nations.
- 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)*.

- UNHCR. (2001). *Counting Forcibly Displaced Populations: Census and Registration Issues. Symposium on Global Review of 2000 Round of Population and Housing Censuses: Mid-Decade Assessment and Future Prospects*. UNSD.
- UNHCR. (2015). *Report of the International Conference on Refugee Statistics 7-9 October 2015, Antalya, Turkey*. Antalya: UNHCR.
- UNHCR. (2017). *Global Trends Forced Displacement in 2016*. Geneva: United Nations High Commissioner for Refugees.
- United Nations. (1998). *Recommendations on Statistics of International Migration Revision 1*. New York: United Nations.
- United Nations. (2003). *Handbook of Statistical Organization Third Edition*. New York: United Nations.
- United Nations. (2004). *United Nations Demographic Yearbook review - National reporting of international data - Implications for international recommendations*. Retrieved from <https://unstats.un.org/unsd/demographic/products/dyb/techreport/migration.pdf>
- United Nations. (2008). *Principles and Recommendations for Population and Housing Censuses Revision 2*. New York: United Nations.
- United Nations. (2008). *Principles and Recommendations for Population and Housing Censuses Revision 2*. Department of Economic and Social Affairs, Statistics Division. New York: United Nations. Retrieved from http://unstats.un.org/unsd/publication/seriesM/seriesm_67Rev2e.pdf
- United Nations. (2014). *Principles and Recommendations for a Vital Statistics System Revision 3*. New York: United Nations.
- United Nations. (2014). *Report of Statistics Norway and the Office of the United Nations High Commissioner for Refugees on statistics on refugees and internally displaced*. Economic and Social Council, Statistical Commission. New York: United Nations.
- United Nations. (2014). *Resolution adopted by the General Assembly on 29 January 2014, 68/261 Fundamental Principles of Official Statistics*. General Assembly. New York: United Nations.
- United Nations. (2015). *Fundamental Principles of Official Statistics Implementation Guidelines*. United Nations.
- United Nations. (2015). *Fundamental Principles of Official Statistics Implementation Guidelines*. United Nations.
- United Nations. (2015). *Principles and Recommendations for Population and Housing Censuses Revision 3*. Department of Economic and Social Affairs, Statistical Division.
- United Nations. (2015). *Report of Statistics Norway, the Turkish Statistical Institute, Eurostat and the Office of the United Nations High Commissioner for Refugees on progress in the work on statistics on refugees and internally displaced*. Economic and Social Council, Statistical Commission. New York: United Nations.

- United Nations. (2015). *Statistical Commission Report on the forty-sixth session (3-6 March 2015)*. Economic and Social Council, Statistical Commission. New York: United Nations.
- United Nations. (2015). *Trends in International Migrant Stock: The 2015 Revision*.
- United Nations. (2016). *Resolution adopted by the General Assembly on 19 September 2016*. General Assembly. New York: United Nations.
- United Nations. (2017). *Handbook on Measuring International Migration through Population Censuses*. Statistical Commission, Department of Economic and Social Affairs. New York: United Nations.
- Van der Erf, R., Heering, L., & Spaan, E. (2006). A Cohort Approach to Measuring the Asylum Procedure. In M. Poulain, N. Perrin, & A. Singleton, *Towards Harmonised European Statistics on International Migration* (pp. 289-320). Louvain: Presses Universitaires de Louvain.
- Zagheni, E., Garimella, V., & Weber, I. (2014). Inferring international and internal migration patterns from twitter data. *Proceedings of the 23rd International Conference on World Wide Web* (pp. 439-444). ACM.