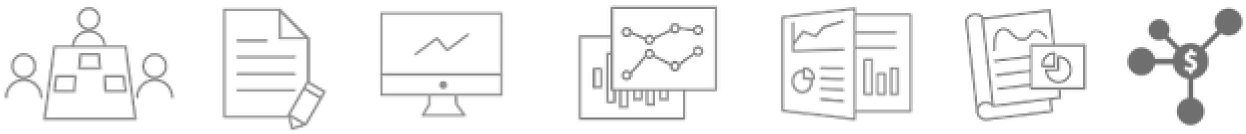
Guidelines and Templates

to facilitate assessment planning and design and to support the “technical” implementation of existing guidances

Last update : 28 April 2016

Why a “Protection Assessment Toolkit”?

The kit is a comprehensive set of tools to guide survey teams through every step of a protection assessment – from overall planning, design and data collection in the field to data processing, analysis, interpretation, documentation and dissemination. The kit is also organised to ensure that all findings and recommandation will allow to inform prioritisation between all potnetial protection activities and subsequent ressources allocation in terms of programme.



Assessment

This toolkit supports the technical implementation of already existing guidelines and guidances, namely the UNHCR Tool for Participatory Assessment in Operations, the Needs Assessment for Refugee Emergencies (NARE) Checklist for Refugees operations and the Rapid Protection Assessment Tools (RPAT) for IDPs operations, the Guidelines for Integrating Gender-Based Violence Interventions in Humanitarian Action, the Child Protection Rapid Assessment Toolkit, the Joint IDP Profiling Service Essential Toolkit- JET, the Heightened Risk Identification Tool, The Operational Guidance and Toolkit for Multipurpose Cash Grants etc. The proposed technical implementation approach was inspired by the UNHCR Standardised Expanded Nutrition Survey and the UNICEF Multiple Indicator Cluster Surveys.

Why “technical”?

On the top of the summary narrative guidelines below, the toolkit is organised around 4 keys technical componnents that facilitate the easy replication and customisation of an assessment from an operation to another. Re-using tested proofed platforms, questions & report format, technical staff in charge will save significant amount of time.

NOTE

- A pre-organised `XLSFORM Questions Library` in order to leverage good practices and enforce core questions during design of the assessment form.
- A `KoboToolBox` secured server to perform the data collection. This server uses the xlsform developed during the design phase.
- Multiple `R statistical analysis scripts` to clean & analyse data, and then to generate automatically standard report and to facilitate the creation of presentation slides and infographics.
- A series of `Github Repository` to exchange collaboratively analysis between operations and generate a knowledge base. Github is also used for the maintenance of the toolkit itself.

For whom?

The targeted audience of the toolkit are *technical staff tasked to work on a protection assessment*: assessment focal points, information management officers and data scientists… The toolkit is to be used once all emergency assessment and secondary data review have been implemented and developped. With a bit of configuration at the begining, the toolkit will allow for the quick generation of results and will ensure that minimum data quality standards are enforced. In addition, because of the standard data format that are enforced through the kit, it will be possible to obtain comparable results from different operations.

Disclaimer

The toolkit is a *collaborative effort*: if you have suggestions, please share them through this link.

The Toolkit is a *work in progress*: if you identify issues, please share them through this link.

Table of Content

Assessment methodology & form design

Planning

Assessment Project Document . . . 7
Memorandum of Understanding . . . 11
Terms of Reference for Assessment Focal Point . . . 15

Methodology

Sample Size & Weight Calculation . . . 19
Household, Individuals & Key Informant Selection . . . 25
Mapping and Household/keyInformant/Community Listing . . . 27

Form

Protection Topics . . . 29
Questions Library . . . 35
Guidelines for Customisation . . . 41

Data collection

Preparing for fieldwork

Fieldwork Training and Agenda . . . 45
Pre-test Phase . . . 47
Managing Devices . . . 51

Fieldwork manual

Instructions for Interviewers . . . 53
Instructions for Supervisors and Editors . . . 61
Instructions for Managers . . . 65

Using KoboToolBox

Data Protection Impact Assessment 67
Server Configuration 71
Data Entry 73

Analysis & dissemination

Analysis Steps

Discover 83
Predict 87
Advise 89

Report

Analysis Scripts 91
Model for Key Findings Report 95
Model for Final Report 97

Communication

Slides & Infographics 99
Sharing microdata for social scientist 101

Credits

Large parts of the toolkit are extracted from other existing guidelines referenced above.

Assessment Project Document

Background and Objective 7
Governance Structure 7
Questionnaires Content 8
Sample Design 8
Recruitment and Training of Fieldwork Staff 8
Fieldwork 9
Budget 9
Timeline for Tasks 9

NOTE

A protection Assessment needs to be carefully planned. The First step of this plan is to document all the elements of the Asesssement

Background and Objective

Protection Assessment are designed to collect statistically sound, internationally comparable estimates of key indicators and analysis that are used to assess the protection situation of displaced persons (Refugees & IDP’s) and to prioritize protection intervention.

Governance Structure

- Give the name and type (government agency or other agency) of implementing agency.
- Provide overview of Memorandum of Understanding (MoU) (Parties, critical components affecting survey planning, etc)

- Give the names and affiliations of those who will be responsible for the management, technical work, and coordination activities. Include the survey coordinator, the sampling expert, and data processing expert assigned from the implementing agency, as well as others, if applicable. If already identified, the UNICEF MICS Consultant and other key regional experts/consultants together with their respective responsibilities should also be included.
- Describe the roles and contributions of national and international stakeholders and funding agencies.
- Describe the status, composition and roles and responsibilities of the Steering and Technical committees.
- Provide other details on the governance structure and human resources as needed

Questionnaires Content

List the excluded modules and provide reasons for the exclusion of each module. If the module or questions are not applicable to the country/survey you are not expected to provide a detailed explanation. However, for other exclusions, please include specific details about the reasons the modules and/or the questions that are not planned to be included and provide references to the data needs assessment.

Provide information on the plans for the translation and back-translation of the questionnaires into local languages and plans for pre-testing the questionnaires. Indicate that the pre-test results will be compiled in a report, and that the results of the pre-test will be used to further modify, customize, and finalize the questionnaires.

Sample Design

In this section, under separate sub-headings, as appropriate, describe:

- The Type of sampling design shoudl be stated (Rationale for sampling design explained)
- State definition of Household used in the assessment
- Sample size, including the expected numbers of households, women, men and under-5s. Include information on sub-sampling of men, if applicable
- How the sample size was calculated, including the indicators used for the calculation of the sample size
- The level of disaggregation sought for reporting
- What sample frame will be used and if the sample frame needs to be updated, plans for mapping, listing and household selection

Recruitment and Training of Fieldwork Staff

In this section, under separate sub-headings as appropriate, describe:

- Plans for recruitment of fieldwork staff, including details of the type of personnel (interviewers, data entry, supervisors, measurers, data entry clerks), their education/background, sex, numbers etc.
- Timing of training
- Length of training
- Methodology and content of training
- Profiles of trainers
- How training will be organized – central location, in separate districts, including how standardization will be ensured if not central location

Fieldwork

In this section, under separate sub-headings as appropriate, describe:

- Timing of fieldwork, constraints on timing of fieldwork
- Team composition, including numbers
- Expected duration of fieldwork and how the duration was calculated
- Plans for monitoring data collection and fieldwork supervision as well as plans for handling questionnaires for data entry
- Fieldwork logistics

Budget

In this section, under separate sub-headings as appropriate, describe:

- Expected total cost of the survey
- Breakdown of total cost by budget line items
- Amount of funding secured and funding source(s)
- Amount of extra funding needed, including plans, if any, on how the funding shortfall will be secured

Timeline for Tasks

- Identify survey coordinator, survey personnel, and plan survey; establish steering and technical committees
- Adapt and pre-test questionnaires; translate questionnaires and manuals
- Carry out sampling and household listing; order scales, boards, salt test kits, and GPS equipment
- Complete logistical arrangements
- Select and train fieldwork personnel (interviewers, editors, measurers, and supervisors)

- Conduct pilot study and collect data
- Complete data processing, including secondary editing
- Prepare summary findings report and final report, and disseminate widely; prepare survey archive

Memorandum of Understanding

Implementation Model 11
Field Level Memorandum of Understanding 12

IMPORTANT

Whether it is for the data collection or the analysis, protection assessment are often done in partnership. To avoid confusion and misunderstnading within the proces, it is recommended to establish a clear Memorandum of Understanding

Implementation Model

According to the particular context, the assessment might be implemented using une of the following models :

- Assigning full data collection responsibility to different agencies in different geographic areas
- Pooling human and logistical resources centrally
- Delegating data collection responsibilities to one or more NGO partners, ideally local NGOs, through a project agreement.

Regardless of the model chosen, it is important to observe four principles:

- The objectives and methodology of the asessment need to be decided by consensus among particpant. Agencies who assume particular data collection responsibilities do not acquire a privileged say in choosing indicators, sites or data collection methods.
- Particpant contribute resources to the implementation of the asessment, to the measure of their capacities and possibilities.
- Common standards and understanding of assessment questions need to be agreed and maintained.

- Clear focal points need to be appointed per each particpating organisation with operational responsibility for data collection.

Field Level Memorandum of Understanding

The following is an outdated example that might provide some inspiration for the inital drafting of the memorandum. Any memorandum needs to be cleared by the Bureau before signature.

Scope of Works

The following is an example of MoU.

> - *The following Memorandum of Understanding between the United Nations High Commissioner for Refugees and `Partner Name` is not attempt to repeat the basic principles of already existing global MOU, including descriptions of the agency responsibilities towards various populations, but will instead highlight the specific areas where close cooperation will be taking place.*
> - *This MoU covers the geographic Coverage for `population group`. All areas will be covered based on `sampling methodology`*
> - *UNHCR and `Partner Name` will agree on a joint questionnaire subsequently used for all data collection related to the profiling exercise, through mobile data collection devices and a `methodlogy` (for instance: combination data collection methodologies: direct observation, key informant interviews and as means of triangulation in areas with a high concentration of `population group`, carry out Focus Group Discussions).*
> - *All data collectors will be selected jointly and need to sign the code of conduct. They will receive a ½ day training on the principles and objectives of the code of conduct. In addition, data collectors will be trained at a minimum in: interviewing techniques, protection principles, data protection (confidentiality, informed consent etc.), referral mechanism, usage of mobile devices, the questionnaire, and basic security principles. This introductory training will be conducted before any data collection takes place and will last 3 to 4 days. One day of field testing with the data collectors is foreseen at the beginning of the roll-out. Technical documentation of the profiling exercise will include:*

data collectors manual, team leader manual, data analysis plan, data entry manual and the respective Standard Operating Procedures for the project.

- The role of the Government from the beginning of this exercise is crucial to ensure a responsible handover at the end of the project and overall ownership by the Government. To this end, the Government will be included in the project through training and in particular Line Ministries in the localities to contribute to data collection.
- The questionnaire will be presented to and feedback sought from partners (UN, NGOs and Government) to achieve buy-in and ensure that the questionnaire meets the information needs of partners providing protection and assistance to `population group`; however, given the need for timely information delivery, contributions from partners are expected to be provided within a week from the day they receive the draft questionnaire.
- UNHCR and `Partner Name` will have joint data ownership. The data will be stored on a UNHCR server, while UNHCR and `Partner Name` will have both administrative rights to access the data base. UNHCR and `Partner Name` will disseminate and/or publish the data collected jointly. None of them will be producing the data alone to publish products /reports /websites etc. under its own logo or other branding.. The design and layout of all information products of this project will be agreed jointly. Data collection will be an ongoing exercise to ensure that information is updated, relevant and timely. At the end of the project the data will be handed over to the Government.
- UNHCR and `Partner Name` will establish a joint budget and enter into a cost-sharing arrangement to finance this project.
- In line with UNHCR's mandate for refugees, UNHCR will be collecting data on refugees residing outside camps which is not part of this MoU.
- Nothing in this MOU shall affect the relations of either signatory to its Governing Body, nor the contractual relationship and administrative supervision of UNHCR and `Partner Name` to their operational partners.
- The implementation of the MOU will be in compliance with the respective administrative and financial rules and procedures of UNHCR and `Partner Name` and be subject to the availability of funds.
- This MOU will enter into force upon signature and shall be of indefinite duration.

> - *This MOU may be terminated by either party upon 90 days written notice.*
> - *This MOU may be modified at any time by mutual consent of the parties.*
> - *The Representatives of both organizations will meet when necessary to discuss policy issues and will nominate officers to meet regularly to review strategic and implementation issues of particular interest to both organizations and to propose possible courses of action to address them.*

Terms of Reference for Assessment Focal Point

Background 15
Methodology 15
Objectives 16
Deliverables 16
Reporting 16
Time frame 16
Qualification & Experience required 16

Background

Insert

- background on current protection situation,
- details on previous assessment or what is known of the protection situation.

Explain why the current assessment is planned and if there is a specific trigger that would indicate a situation change.

Insert any other relevant detail.

Methodology

The survey will be developed using the Protection Assessment Toolkit. Survey methodology should be reviewed by UNHCR regional or HQ Protection Infornmatiuon Management Officer prior to data collection.

Objectives

The Assessment Focal Point will oversee the Protection Assessment for: - `Operation` - `Population Group` - `Geographic Coverage` - `Timeframe`

Deliverables

- A final assessment report including recommendations on actions to address the situation is to be submitted at the end of the mission. Results of standardisation tests, details of data cleaning and plausibility checks should be presented in the final report.
- Standardised tables as presented in the Protection Assessment Toolkit.
- The findings and major recommendations are to be presented to partners at the mission level (oral presentation and slides).
- The final analysis script.

Reporting

The consultant will report on regular basis to the UNHCR [insert title of UNHCR person responsible], who will have the overall responsibility of the survey.

Time frame

The consultancy will last approximately [insert number of months], starting from [Insert start date].

Qualification & Experience required

The successful candidate will:

- Have a university degree or the equivalent in social science, with a specific competency in humanitarian emergencies.
- Have significant experience in undertaking surveys (design and methodologies, staff recruitment and training, field supervision and data analysis/write up).
- Be familiar with the survey methodology and R statistical language.

- Be fluent in English with excellent writing and presentation skills [insert any other language requirements].

Sample size & Weight Calculation

Sampling strategy . . . 19
Non-probabilistic . . . 19
Probalistics . . . 21
Sample universe for non-proabilistic approach . . . 21
Defining sample universe and size for probabilistic approach . . . 21
In refugee Context . . . 22
In IDP context . . . 22
Sample Weight . . . 22
How are the oversampled/undersampled areas corrected in data analysis? . . . 23
What does it mean to normalize the weights? . . . 23

Sampling strategy

Sampling strategy can be either probabilistic or non-probabilistic

Non-probabilistic

Convenience sampling

It is a frequently used method in emergency situations. It relies on sampling those respondents who are easiest to access.

Practically speaking those coudl be either: * key Informant that are ready to report by themself * Individuals or household among those who have settled along roadsides, or who present themselves to administrative center of the returnee settlement or the assistance desk etc.

The danger with this type of monitoring is that it will often lead to biased results as the sample may not be representative of the majority, i.e. those with the most resources or power are often the ones who settle in the most easily accessible areas.

Purposive sampling

It is based on previous knowledge about who might be able to provide valuable or specific information. It uses the judgement of community representatives, project staff or assessors to select typical locations and/or informants. The sampling of children or women, for example, is a type of purposive sampling.

Purposive sampling can also be done through Key Informant.

- Advantages: Moderately rigorous if well and clear criteria for sampling are followed. Useful when targeting specific groups of affected population or specific affected areas. Less time consuming and less expensive than representative sampling
- Disadvantage: Generalisations are biased and not recommended. Samples are not representative of population due to subjectivity of respondents

Snowball & Respondent-driven sampling

Snowball sampling (or chain sampling, chain-referral sampling, referral sampling) is a non-probability sampling technique where existing study subjects recruit future subjects from among their acquaintances. This technique is subject to numerous biases. For example, people who have many friends are more likely to be recruited into the sample.

This approach might underweight the most vulnerable individuals.

A declination of snowball sampling is the Respondent-driven sampling -RDS approach. It combines “snowball sampling” with a mathematical model that weights the sample to compensate for the fact that the sample was collected in a non-random way. While data requirements for RDS analysis are minimal, there are three pieces of information which are essential for analysis (RDS analysis CANNOT BE PERFORMED without these fields for each respondent):

- Personal Network Size (Degree) - Number of people the respondent knows within the target population.
- Respondent’s Serial Number - Serial number of the coupon the respondent was recruited with.
- Respondent’s Recruiting Serial Numbers - Serial numbers from the coupons the respondent is given to recruit others.

Purposive sampling using strata

In this case, the risk of loosing certain componnent of the population is addressed by defining strata.

Time-Location Sampling

This Time-Location Samplin (TLS) approach can be used when the goal is to have a representation of population in movement. The idea is to sample persons at locations and at time at which they may be found. Time-location sampling is used to sample a population for which a sampling frame cannot be constructed but locations are known at which the population of interest can be found, or for which it is more efficient to sample at these locations. This approach is be appropriate for **border monitoring** for instance.

You can find a here practical guidelines for TLS.

Probalistics

Random sampling

Stratified sampling

Cluster sampling

Sample universe for non-proabilistic approach

In this case (Desk interview or key Informant), the more observation the better.

Some kind of credibility scoring can be obtained for each locations based on a review of the key informant.

Defining sample universe and size for probabilistic approach

If you need a purely random sample, the size of the sample is a calculation that takes 3 variables:

- Size of the full population
- Confidence level: for what proportion of the population you want to get the right estimation (usually either 90%, 95% or 99%)

- Error Margin (or confidence interval): How much error are you willing to tolerate for each questions? i.e. + or – your estimated ratio for each questions on the top of the confidence interval (usually either 5%, 2% or 1%)

There are online calculator for this.

For 400,000 Syrians	5% error margin	2% error margin	1% error margin
90% Confidence level	272	1694	6692
95% Confidence level	384	2387	9379
99% Confidence level	662	4105	15929

For 150,000 Afghans	5% error margin	2% error margin	1% error margin
90% Confidence level	272	1682	6511
95% Confidence level	383	2363	9026
99% Confidence level	661	4036	14937

Usually the decision on the right confidence level and error margin to be selected is also influenced by cost implication and the final usage of the figures that is looked for.

In refugee Context

Data is coming from proGres

In IDP context

Data is coming from a *Displacement Tracking Matrix*

Sample Weight

Over-sampling in regions with small populations ensures that they have a large enough sample to be representative. Under-sampling is done in regions with large populations to save costs. Sample weights are mathematical adjustments applied to the data to correct for over-sampling, under-sampling, and different response rates to the survey in different regions.

How are the oversampled/undersampled areas corrected in data analysis?

The samples are designed to permit data analysis of regional subsets within the sample population. When the expected number of cases for some of these regions is too small for analysis, it is necessary to oversample those areas. When the expected number of cases for some of these regions is unnecessarily large, those areas may be undersampled to accommodate logistical or budgetary constraints.

During analysis, it is then necessary to “weight down” the oversampled areas and “weight up” the undersampled areas. The developing of the sampling weights has taken this factor into account. Always use the weight variable found in the DHS data set. Even in surveys that come from a self-weighting sample, it is still necessary to use the sampling weights in analysis because the response behavior may differ by response groups.

What does it mean to normalize the weights?

After the weights are initially calculated, they are normalized, or standardized, by dividing each weight by the average of the initial weights (equal to the sum of the initial weight divided by the sum of the number of cases) so that the sum of the normalized/standardized weights equals the sum of the cases over the entire sample. The standardization is done separately for each weight for the entire sample.

The entire set of household sample weights is multiplied by a constant, thus, the total weighted number of households equals the total unweighted number of households at the national level.

Individual sample weights are normalized separately for women and men. Thus, the total weighted number of women equals the total unweighted number of women, and the total weighted number of men equals the total unweighted number of men. Women and men are normalized separately because all non-HIV calculations are performed on women and men separately. We do not provide survey estimates on the joint population of women and men combined for anything other than HIV prevalence.

Household & Key Informant Selection

IMPORTANT

This chapter is not written yet.

Mapping Household Listing

> **IMPORTANT**
>
> This chapter is not written yet.

It is recommended that, before actual data collection in the field, a preparatory visit is done to each site of the assessment. This site preparation visit is the first contact between the assessment team and affected communities, and responds to the objectives of:

- Informing key community stakeholders about the assessment,
- Gathering critical site information required for effective primary data collection,
- Confirming (or refuting) assumptions made in the planning stage, including logistics issues, travel time, limits of accessibility, and time available for fieldwork as well as highlighting implications of security issues (curfews, movements of armed actors, and risks of mines or explosive remnants of war).

During site preparation (which can be undertaken by the field team leader, together with a second team member), meetings are held with key stakeholders in which team members will:

- Introduce themselves,
- Introduce the assessment objectives,
- Agree the assessment schedule and confirm that is suitable to the community and that will include representation of necessary population groups as planned, including those most vulnerable,
- Work with authorities, community leaders or members of the local community to identify KI interview and FGD participants,
- Confirm that the site has the characteristics of the strata it has been selected for,
- Identify locations for interviews and FGDs and confirm the suitability of these structures,
- Provide information about the work plan , confirm KI interview and FGD start times,
- Inform whether the participants should expect food or refreshments,
- Take the details of community leaders for any unforeseen communication.

A master list of sites which have been visited should be kept by the assessment team leader and noted on a map of the affected area to ensure that sites are not visited twice and that an appropriate geographic selection of sites has been used.

Protection Topics

The Right based approach . . . 30
Vulnerability Profile . . . 31
Analysis Topics . . . 32
Favourable Protection Environment . . . 32
Fair Protection Processes and Documentation . . . 32
Community Empowerment and Self Reliance . . . 32
Basic Needs and Essential Services . . . 33
Security from Violence and Exploitation . . . 33

IMPORTANT

Translate assessment results into programmatic response recommendations is challenging when it comes to protection activities.

The analysis of vulnerabilities allows to generate protection activities recommandation.

The analysis of vulnerabilities is complex as it should allow to understand:

- How **multiple vulnerabilities** interact between each other (prioritisation through criticality)?
- What is, for **each vulnerability type**, the specific profile of a population group in terms of magnitude and severity?

Translate assessment results into programmatic response recommendations is challenging when it comes to protection activities.

The Right based approach

Protection programme are designed according to the Right Based approach. As such, more than the needs, it is the vulnerability of specific population group toward right violation that informs ressources allocation.

Vulnerability level for each specific risk = function(risk **occurrence**) + function(**exposure** to risk of basic right violation) + function(**coping capacity** before or after violation)

Protection activities for UNHCR:

- Prevent occurrence of basic right violation event (**Environment Building actions**) &&
- Limit the effects of violation events consequences (**Remedial actions**) &&
- Respond to basic right violation event (**Responsive actions**)

Protection analysis is therefore linked to specific events linked to specific Right groups defined by UNHCR Result Based management.

Rights Group	Basic right type	Violation event
Favourable Protection Environment	Non-refoulement	Non-admission, Refoulement
-	No discrimination	Detention
-	Freedom of movement	Eviction, deportation
Fair Protection Processes and Documentation	Documentation	No issuance/renewal of residency
-	Right to a nationality	Birth Registration
Community Empowerment and Self Reliance	Livelihood	Negative coping mechanism (Prostitution, child labour, early marriage, begging, stealing,.)
-	Peaceful co-existence	Intercommunity violence
Basic Needs and Essential Services	Education	Children out of school
-	Food	Malnutrition
-	Social & economic rights	Non access to services (healh, schools, MPSS)
-	Housing, Wash & Shelter	Sub standard living conditions
Security from Violence and Exploitation	Age, gender and diversity mainstreaming	Sex & Gender based violence, child abuses

Vulnerability Profile

On each topic, the protection assessment should allow to identify the affected population's vulnerability profile:

- Threat in terms of **criticality** , i.e how important is the threat compared to other. This is measured through priority ranking.
- Threat in terms of occurence in order to define **magnitude**
- Capacities and capabilities in order to define **severity** – "Severity" (or intensity) expresses the degree of unmet needs (it is thus related to shortages and deficits, as opposed to fulfillment and wellbeing) or the egree of something harmful, harsh, stern, irreversible or not desirable. As such it is expressed through a form of rating.

The vulnerability profile (combination of criticality/magnitude/severity) for each topic will allow to prioritise the relevant activities. Severity and Priority are therefore appropriate way to measure vulnerability.

The risk profile will help designing the best intervention approach for each group:

Activity	Criticality	Magnitude	Severity
Environment Building actions	++	+	-
Remedial actions	+	++	+
Responsive actions	+	-	++

Analysis Topics

Favourable Protection Environment

List of analysis modules

- Protection-sensitive border mechanism
- Counselling & Legal aid
- Freedom of movement & Detention
- Multiple Displacement & Movement

Fair Protection Processes and Documentation

List of analysis modules

- Birth Registration
- Civil status documentation
- Family re-unification
- Refugee registration
- Risk of eviction

Community Empowerment and Self Reliance

List of analysis modules:

- Social cohesion
- Housing, shelter & wash conditions

- Employment & livelihood
- Child Labour
- Assets & Budget

Basic Needs and Essential Services

List of analysis modules:

- Refugees with specific needs
- Assistance received
- Negative coping mechanism
- Food security
- Access to health services
- Education & Out of School Children

Security from Violence and Exploitation

List of analysis modules:

- Sexual Harassment & Violence (Rape)
- Harassment between youth
- Gender inequality
- Domestic Violence (Parental stress, Isolation at home)
- Forced and early marriage
- Violence among children
- Armed recruitment
- Harassment & Violence towards LGBTI
- Security & Civil violence

Question Library

Intro: from the plan to the form . . . 35
Questions componnent . . . 36
Questionnaire Componnent for Household Representative . . . 36
Questionnaire Componnent for Individual Women . . . 36
Questionnaire Componnent for Individual Men . . . 37
Questionnaire Componnent for Children & Youth . . . 37
Question design . . . 37
Checklist . . . 37
Context information . . . 37
Locations & Geography . . . 38
Likert questions . . . 38
Using the library . . . 40
The XLSFORM format . . . 40
Filtering the questions . . . 40
Questionnaire anaysis report . . . 40

IMPORTANT

Objective of the question Library:

- Avoid replicating **bad practices**
- Ensure that all **topics of interest** are well covered in the analysis framework
- Check that for each selected topcis, all **required questions** to calculate the indicators and vulnerability profile (severity, magnitude, criticality) are included
- Deliver quickly **final reports** through reproducible analysis workflow

Intro: from the plan to the form

During the design stage of the assessment, the team must consider how each data element collected will be compiled, aggregated, analysed and disseminated to satisfy the information needs. The data analysis plans allow the assessment team to identify data elements on data collection forms that are not analysable as well as information management requirements that have not been met. Once missing data elements have been identified, forms can be amended accordingly by modifying, removing or adding data elements.

A common mistake for assessment teams is to collect too much data that is neither analysable nor will be used in decision making. The temptation to ask too many questions means that teams gather poorer quality data which obstruct useful analysis. If a data analysis plan shows that too much data will be collected, then data collection forms should be revised and shortened accordingly.

Each data element collected should be linked to:

- An information need linked to a Protection Topic
- Contextual information in relation with the Population Group (Refugees, IDPs)
- Data source - i.e. from aggregated Household Information, Key Informant or Focus Group discussions
- Type of Analysis: I.e. Correlation, Dispersion, Average,
- Specific Protection indicator

Questions componnent

In the Protection Assessment at the Household level, some questions should be linked to specific members within the household:

Questionnaire Componnent for Household Representative

- Profile of Household Members
- Education
- Child Labour
- Civil Documentation
- Property
- accomodation Conditions

Questionnaire Componnent for Individual Women

- Access to Mass Media and Use of Information and Communication Technology
- Attitudes toward Domestic Violence
- Marriage/Union
- Intention

Questionnaire Componnent for Individual Men

- Access to Mass Media and Use of Information and Communication Technology
- Attitudes toward Domestic Violence Marriage/Union
- Intention

Questionnaire Componnent for Children & Youth

- Birth Registration
- School attendance
- Early Marriage

Question design

Checklist

- Closing questions
- Skip patterns
- Ranges for selected questions
- Numbering questions
- Two questions in one
- Double negatives
- Clarity and simplicity
- Consultation
- Parcimony

Context information

- Metadata
- Enumerator
- Address
- Consent
- Household information
- Self-expressed Needs
- Subjective Evaluation
- Referral

- Notes

Locations & Geography

Pcoding location within a form is an important taks

it is possible to add pcode through a reference file, *itemsets.csv*, is available for each country. This reference file can be used when creating Cascading selects questions using the select_one_external function within XLSFORM questionnaires. The itemsets.csv file can be uploaded to UNHCR Kobo Server as a media file. It will be downloaded to any Kobo Collect like any other media file and saved to the [form-filename]-media folder. Clients like Kobo Collect load media files from the SD card and so a field data collection form with all locations within the country can load very quickly.

itemsets.csv for the MENA region countries can be downloaded from UNHCR MENA pcode

Likert questions

Likert scale also called "opininion scale" desribe a way to formulate questions related to opinion. A Likert item is simply a statement that the respondent is asked to evaluate by giving it a quantitative value on any kind of subjective or objective dimension, with level of agreement/disagreement being the dimension most commonly used.

Rating questions are often much easier to understand by individual than **ranking questions** where people may be forced to make one item worse or better than another, when they actually find them equal. Psychological research has proven that it's a very tough thing for people to rank more than three options. When people are given a cognitively difficult task like ranking question choices of for instance six things, often they will end up getting frustrated and will start to rank choices randomly. This will lead to higher dropout rates and even worse, messy data.

Likert questions allows for each individual parts to have the same value, if that is how people feel about them. Last the analysis of rating is more intuitve than the one of a rank that in addition might not have much statistical significance. Also from likert questions, it's possible to compute an average score from each separate ratings in order to get an overall ranking through a Borda Count.

When designing likert questions, attention need to be given to specific design issues specifically in terms of wording.

1. Length of the scale: limit the number of points along the scale.
2. Odd or even point scale: There should not be preferred or better choice.
3. Label the points in the scale: Avoid using just numbers to indicate the points on the scale.
4. Balanced scale: Make sure that the scale is balanced with an equal number of positive and negative categories. Center point on bipolar scale: A common mistake when creating a rating scale is including "no opinion" or "uncertain" as a middle response on a bipolar scale. These options are not actually a part of the scale order. A middle category in a scale between "agree"

and “disagree” would be “neither agree nor disagree”. Options such as: “no opinion”, “not sure”, “undecided”, “don’t know”, or “not applicable” are placed off the scale, in a separate space.
5. Match response to question: Be as direct and specific as possible, focusing the response options on what you want to measure.
6. Keep labels consistent: Finally, the labels used in the scale need to refer to the same thing.

Some examples are given below:

Agreement

- Strongly Agree
- Agree
- Undecided
- Disagree
- Strongly Disagree

Frequency

- Very Frequently
- Frequently
- Occasionally
- Rarely
- Never

Importance

- Very Important
- Important
- Moderately Important
- Of Little Importance
- Unimportant

Likelihood

- Almost Always True
- Usually True
- Occasionally True
- Usually Not True
- Almost Never True

Quality

- Very poor
- Poor
- Fair
- Good
- Very Good

Interest

- Not at all interesting
- Slightly interesting
- Moderately interesting
- Very interesting
- Extremely interesting

Using the library

The XLSFORM format

XLSFROM is a standard way to describe a form. It includes information not only on the type of questions but also on:

- Hints in order to add comments on methodology
- Contraints to avoid entering uncorrect information
- Relevant to allow for skip logic
- Support for multiple language
- Possibility to perfom calculation for data quality checking
- Possibility to repeat the same questions in a loop.

Filtering the questions

The question library is an online google spreasheet where a sries of questions are organised around multiple categories

To use it, simply copy or donwload the library, filter out the question you are not interested in and delete them.

The next step will be then to customise the form, page 41

Questionnaire anaysis report

The libary can generate a short report that sumerrise:

- What topcis are covered by the form
- Check that for each topics, questions are selected to allow for the estimation of magnitude, criticality & severity
- Summarise questions that can be raised in key Informant Interview, samples household interview, on site household interview or Focus group Discussions.

Guidelines for questionnaire customisation

What is "customisation"? . . . 41
What kind of "Customisation"? . . . 42
Modification . . . 42
Deletion . . . 42
Addition . . . 43
Rules and Useful Tips for Customisation . . . 43

NOTE

What is "customisation"?

Customisation (or adaptation) refers to the process during which the proposed Protection Assessment questionnaires are tailored to the population/context where a the Assessment is being conducted (that is, a national assessment, or a assessment conducted for a population group or for a selected area within a country), using standard principles and approaches, while maintaining global comparability of the indicators that will be derived from the collected data.

The customisation process is by no means an easy and straightforward one. Without a detailed understanding of all the standard tools and of the general principles and recommendations, customisation of questionnaires should not be attempted at country level without the assistance of an expert. During the customisation process, it is also absolutely critical that lessons learned from previous data collection activities are used effectively, and wherever necessary, tools are tested before final decisions are made. Testing may include organized pre-testing, field testing, piloting, and in some cases, cognitive testing. Analysis of raw data from previous assessment and data collection activities, as well as results from these efforts should also be undertaken for successful customisation of standard questionnaires.

What kind of "Customisation"?

Customisation of the "Questionnaire Modules", "Questions" and "Response Categories" are necessary for at least two basic reasons: * No single country/survey is expected or recommended to use all of the modules in proposed questionnaires * No single standard questionnaire can accurately represent all human experience around the globe

Customisation covers the following types of changes to the standard protection questionnaires: * Country-/assessment-specific modifications to already existing standard questions and response codes, * Deletions from the standard questionnaires, and * Additions to the proposed questionnaires.

Modification

Certain parts of the proposed questionnaires must be modified. Indeed, in several instances, the proposed questionnaires include clear directives that a change or modification needs to be made. These cases are indicated using text such as "insert local name". Similarly, response categories that require customisation are also indicated.

Deletion

No assessment is recommended to retain all of the modules and questions of the proposed protection questionnaires. First, there will always be some topics that will not be relevant in certain countries or regions. Second, decisions on the content of any assessment will ideally be made as a result of a thorough data gap assessment, generally based on the required analysis, and, for example, when information is available from other recent data sources, certain modules or sets of questions will be dropped. The process and analysis involved in a comprehensive data needs assessment will vary, but is a crucial step in determining the content of the assessment.

Determining what to exclude from the assessment is a balancing act that should take data needs into account, but also learn from countless experiences of data quality issues as a result of overloaded questionnaires. Country priorities will guide decisions, but may also work against achieving an optimum questionnaire size if negotiations turn more political than technical.

A final consideration will also rest with the ability to implement an adequate sample size, as this is often constrained by budget on one hand and on the other the known data quality issues associated with large sample sizes. For instance, some indicators are difficult to measure in low fertility settings, demanding higher sample sizes or complicated sample designs. Unless such issues can be technically addressed, the exclusion of such indicators may be necessary.

Addition

Some Protection Asessment may also add topics, modules and questions which are not already in the proposed questionnaires. These could include additions that the proposed questionnaires already point to (for example, adding household assets to the list already in the questionnaires), or additions of modules or sets of questions that are not covered in the proposed questionnaires.

From the onset of considerations of what could be added in, you should know that this will affect the technical support available as well as require changes and considerations throughout the package of tools available, from sampling, training, instructions, and data entry application to tabulations and reporting.

As with the above exercise of deleting from the questionnaires, your entry point should be the indicator list or, alternatively, the tabulation plan. Questionnaire design is secondary to the need for precise information on what such proposed additions would be measuring and how such would be presented.

Only questions that are previously well-tested and validated should be included. Questions are often imported from other household surveys that have been conducted in the country. This does not necessarily mean that they are validated nor does it mean that such questions can work within the frame of a Protection Assessment.

If additions are made, please ensure that formatting and coding follow the rules in place for the proposed questionnaires. For entirely new topics it may be useful to build a new module and in other cases you will need to append to an existing module or insert within the existing flow. If you create new questions, submit them for addition in the library here.

Rules and Useful Tips for Customisation

Customise but do not compromise global comparability: * Assess the implications of changes; * Check that all required questions to calculate the indicators are included; * Check previous surveys to see how the customisation was done; * Consider translations for all major languages spoken among the survey population – Arabic, Kurdish,etc..

Pre-test rigorously to make sure that : * the questions are understood and the response categories are meaningful; * the language style that can be understood by everyone; * the skip-logic functions within the form are working well;

Fieldwork Training Agenda

Training 45
Day 1 45
Morning 45
Afternoon 46
Day 2 46
Morning 46
Afternoon 46

Training

Training of survey teams is fundamental for the quality of the survey information collected and preparation for the training and survey will take time, especially the first year the full assessment is implemented.

The training for the full assessment is recommended to last at least 2 days, depending on context and team experience. Extra staff should be trained in case someone is unable to perform the field work.

The main topics to cover in training of data collectors (note that team leader may be provided with a more in-depth training than some of the data collectors) are as follows:

Day 1

Morning

- Reason / objectives for Protection Assessment
- Composition of survey teams : roles and responsibility
- Sampling procedure: why sample?, explained in a way that surveyors can later on explain to community members when asked; and rationale and importance of representativeness
- Questionnaire and sheets: household-level information, child-level information, woman-level information, observations

Afternoon

- Introduction to the household and informed verbal consent
- Interview questions and interviewing techniques: go through each question for clarity, answer options, cultural appropriateness, gender sensitivities, avoid suggestive questioning but probe where necessary
- Age recording and use of local events calendar
- Practicing with real children and/or adults

Day 2

Morning

- Survey logistics
- Equipment
- Communication
- Travel
- Incentives / salary/allowances
- Food and drinks
- Accommodation, etc.

Afternoon

- Pilot test

Pre Test Phase

Why "pre-testing"? 47
Objectives 47
Organisation of the Pre-test 48
Clusters selected for pre-test 48
Personnel 48
Training 48
Fieldwork 49
Conclusions 49
Pre-test Results and Recommendations 49
Questionnaires 49
Average duration of interviews 49
Manuals 49
Interview process considerations 50
Assessment process considerations 50

Why "pre-testing"?

Use pre-test to see:

- if changes have affected the flow of the questionnaires;
- if response categories are comprehensive; any answer falling into the "other (specify)" category of a multiple choice question and that constitutes about 5 percent or more of all answers to that question should be considered as a serious candidate for a separate answer category of its own; and
- if translated questionnaires are working.

Objectives

Briefly describe the overall objectives and the focus of the pre-test. Example text is provided below: These are the areas you must address under the Pre-test Results and Recommendations section below.

The objective of the pre-test was to verify that the questionnaires are functioning well in the [country] context and specifically to make sure that:

- Translations are accurate;
- Standard questions are clear and response categories are adequate for the assessment population;
- Country-specific response categories are adequately included;
- Country-specific questions and modules are adequately tested;
- Difficult or sensitive questions/modules are identified so that extra training can focus on these questions during the fieldworker training;
- Respondents do not misinterpret the questions (questions are not ambiguous or difficult to understand);
- Changes in wording or improved translation have been incorporated when required;
- The questionnaires flow smoothly;
- Interviewer instructions in questionnaire and in the Instructions for Interviewers are clear and sufficient;
- There is adequate space on the questionnaires and all answers can be clearly coded;
- New codes for common answers that were not included in the original questionnaires are created;
- Average duration of interviews is calculated in order to plan the fieldwork;
- Daily workload per interviewer/team can be planned.

Organisation of the Pre-test

Clusters selected for pre-test

Describe the pre-test locations: Where were the households located for the pre-test, how and why were these locations selected, etc.

Personnel

Present the trainers and interviewers (trainees) of the pre-test. Include information on the future involvement of the participants in the rest of the assessment process

Training

List the dates and content of the pre-test training, as well as how it was organised. Some detail is useful on agenda (include as appendix) and training methodology, as it can serve as lessons for the main training. Include other details as relevant: Venue, recommendations for main training, etc.

Fieldwork

Provide the dates of actual pre-test fieldwork. Also detail on organisation (logistics, teams, areas, etc.) is very useful.

Conclusions

Describe how the findings and observations from the pre-test were collected and discussed and what process for making changes to the final questionnaires was used. It is useful to provide detail on the methodology used to obtain qualitative findings (meetings, observation forms, etc.) as well as how data analysis was organised (test of CS Pro application, tabulation programs, etc.)

Pre-test Results and Recommendations

Relating to the objectives listed above, this section should include findings from the actual data gathered as well as the qualitative findings from the pre-test, including those obtained from discussions with interviewers after the pre-test fieldwork concluded.

Questionnaires

This section is the main output of the Pre-test Report. The use of the table below is recommended. Please add all modules in the questionnaire. Make sure that all suggested changes are listed and that evidence is provided for final decisions. Please include observations on all country-specific modules and questions.

Average duration of interviews

Calculate the average duration of interview for each questionnaire using the data collected in the pre-test. Typically, as interviewers become more familiar with the tools, this time will decrease and therefore a realistic duration should be proposed and included in the introductory sentences on the cover pages of the questionnaires.

Manuals

Describe and list any changes or additions required in the Instructions for Interviewers as well as those introduced in the Instructions for Supervisors and Editors. Such changes typically involve translation issues, instructions for country-specific questions, but also for country-specific response categories. Appropriate corrections are incredibly helpful and will especially inform the main field work training.

Interview process considerations

Describe and address the observations from the pre-test that relate to interviewing that will be relevant for training and monitoring in the main field work (for example issues in approaching households, dealing with sensitive module and questions, flow of field work, roles and responsibilities, etc.)

Assessment process considerations

Describe here the observations, suggestions, and decisions related to the assessment planning and next steps for finalising the questionnaire (training contents/agenda, logistics, staff, UNICEF support, etc.)

Managing Devices

Android Devices 51
Computer 52
STEADY POWER SUPPLY 52
USB BATTERY PACKS 52

Android Devices

The software platform used for the data consolidation has specific clients developped on the Android Mobile Operating System. It's also possible, although not first recommended to use any browser (including any browser on a smartphone) to record information.

A certain number of smartphones or tablets will be needed for each survey team to have one or two phones, as well as a few backups.

For ease of use, smartphones and tablets with a large touchscreen and slide-out keyboard are preferable. Equipement might be borrowed from UNHCR HQ, or country operations could have their own set of phones.

Minimal Requirements for the devices:

- Android Version 2.3 (recent phones are generally platform 4 or higher)
- Screen size of 4 inches is recommended
- GPS Chipset
- Wifi connection

Required Applications:

- Latest recommended version of KoboCollect
- GPS Test, if you need GPS coordinates for the survey
- Andexplorer (or equivalent file explorer)

Computer

We strongly recommend to have a dedicated survey computer. This computer will be set up specifically for the survey. Note that using the computer for other purposes (e.g. internet) may cause compromise the security of the data you will collect.

Minimal Requirements:

- 4G RAM
- Installed with Windows 7 (alternatively XP works as well)

Software Required:

- Microsoft Excel (English Version)
- R
- Rstudio
- Github

STEADY POWER SUPPLY

Access to a good electrical supply is essential for the router and computer and for over-night charging capacities of the phones. Where electrical supply is unstable with unexpected power-cuts or planned power savings, alternative charging options must be considered before the survey.

USB BATTERY PACKS

Even though they cannot replace steady electrical power, battery packs are a useful secondary source of energy for the phones. We recommend having at least 2 for a survey where the phones are daily charged. This way, if a phone lack battery during the survey, the enumerator can charge it while continuing the survey.

Instructions for Interviewer

Eligible Respondents for the Household Questionnaire 53
Eligible Respondents for the Individual Questionnaires 54
Finding and Re-Visiting Households 55
How to handle the interview 56
General Points 57
Role of the interviewer 59

Eligible Respondents for the Household Questionnaire

In each sampled household you visit, you should begin by interviewing a knowledgeable adult member of the household to fill in the Household Questionnaire. All modules of the Household Questionnaire will be administered to this person, referred to as the Household Respondent, including the modules in the questionnaire where the information collected is about other household members. The Education module is one such example.

For the purposes of the Household Questionnaire, an adult is defined as someone age 15 years and over. However, young adults (below age 18) may not be the most ideal members to interview. Therefore, in cases when there is another older household member (for instance, the parent of the 15 year-old) available to interview, you should prefer to interview this person who is likely to be more knowledgeable about the household. Whenever possible, you should use your preferences to interview the household member who is likely to be more knowledgeable.

On the other hand, interviewing the household head is not a requirement and you are not expected to ask for the household head to do the interview.

You should also keep in mind that for practical reasons, it may be an advantage to begin the Household Questionnaire with a mother or primary caretaker (of a child under five years of age), since many of the questions/modules are about children, and mothers/caretakers provide more accurate responses to such questions better than anybody else. While you should not make a special effort to ensure this, you will indeed start the interview with such persons in many cases, since, in practice, these persons are more likely to be at home than, say, male household heads.

There should only be one respondent to the Household Questionnaire and the other members of the household should not respond to any part of the questionnaire. Multiple respondents to the questionnaire will undoubtedly lead to an uncontrolled, low quality interview, and may lead to errors in recording responses. Ideally, the household respondent is not expected to consult other members that may be available in the household. However, you may allow the household respondent to ask other members in order to get more correct information, especially on information such as age, which may affect the eligibility of some members for individual questionnaires, or modules where age checks are important, such as the education, child labour, or the child discipline modules.

Eligible Respondents for the Individual Questionnaires

When you have completed the Household Questionnaire, you will have identified women (age 15-49), men (age 15-49) and 'mothers or primary caretakers' (age 15 or above) of children under five to whom you or other interviewers in your team will administer the individual questionnaires. * You should interview separately all women age 15 through 49 who reside in the household to fill in the Questionnaire for Individual Women. * You should interview separately all men age 15 through 49 who reside in the household to fill in the Questionnaire for Individual Men. * You should administer the Questionnaire for Children Under Five to mothers of children under 5 years of age who are residing in the household. If the mother is not recorded in the List of Household Members (if the mother is not a member of this household), then the person who is acknowledged by the household respondent as the primary caretaker (in HL15 in the List of Household Members of the Household Questionnaire) should be the respondent to the Questionnaire for Children Under Five.

You will identify these individuals by completing the List of Household Members in the Household Questionnaire.

If you visit a household where there are no members eligible for the individual questionnaires, you must still ask questions about the household to a knowledgeable adult household member and complete the Household Questionnaire.

As a general rule, the respondent to any of the questionnaires must be at least15 years old. This also applies to the mother or primary caretaker of a child under age 5; in the rare event that a mother or primary caretaker is less than age 15 you should record 'Other' as result of the interview in UF9 and specify that the mother/caretaker is less than age 15 and therefore cannot be interviewed. No other respondent is permitted than the mother/caretaker identified in the List of Household members.

Finding and Re-Visiting Households

Your supervisor will give you a list or tell you how to find the households to visit. You must visit all these households and should not replace these households with other households that are not selected for interviews.

If no one is at home when you go to interview the household, ask the neighbours whether anyone lives at this location. If it is occupied, ask the neighbours when the household members will return. Arrange with your supervisor to go back to the location when the household members are expected to be back; for example, at the end of the day. Note such plans on your cluster control sheet and note the time you are to return on the first page of the questionnaire (Household Information Panel).

If no adult household member is at home, arrange to come back at another time. Do not interview a household member younger than age 15, a temporary caretaker of the children, such as a daytime babysitter, and do not interview anyone who does not usually live in the household. The rule to interview a knowledgeable adult household member cannot be relaxed or violated under any circumstances.

Each household in the sample has to be visited at least three times (two re-visits) before you can mark HH9 (Result of household interview) as 'No household member or no competent respondent at home at time of visit', unless otherwise instructed by your supervisor. There may be cases when you learn that the household will be away for an extended period, and will definitely not return within the fieldwork period, in which case HH9 would be marked as 'Entire household absent for extended period of time'. In such cases, three visits to the household may not be necessary. However, even in such cases, the ultimate decision will have to be taken by your supervisor.

If an eligible woman or man, or a mother or primary caretaker is not available for the individual interview or is not at home, ask a household member or neighbour to find out when she/he will return. Note this on the Woman's, Man's or Under-5's Information Panel, follow your supervisor's instructions, and return to interview her/him at that time. Do not take responses for these questionnaires from anyone other than the eligible person her/himself.

The person to be interviewed for the Questionnaire for Children Under Five should be the mother. A person other than the mother of the child under five can be interviewed only if the mother is living elsewhere or is deceased, and therefore does not appear in the List of Household Members in the Household Questionnaire. In these cases, the person who is acknowledged by the household respondent as the primary caretaker of the child in that household should be interviewed. If the mother/primary caretaker is not available for interview or not at home, try to find out when she/he will be available and return to the household later. If the person will not be available or will not return home at a time later that day when it is feasible to interview her/him, follow the instructions of your supervisor about the number of times you should attempt the interview.

If a child under five is not available, but the mother/primary caretaker is available, you can complete the Questionnaire for Children Under Five, with the exception of the Anthropometry module, since you need the child to perform measurements. In such a case, complete the questionnaire with the mother/primary caretaker, but leave the Anthropometry module blank to be completed during the next visit. Note this and discuss with your supervisor. If the child is still not available after the re-visit(s), record the result in question AN2 as 'Child not present'. Re-visits should be planned by supervisors, if possible, to measure the heights and weights of children, when children are not present at the time of first visit to the household.

Ask your supervisor if you are in doubt about what to do when you cannot locate a household, or you cannot complete an interview. Always keep a record on the cluster control sheet of the households you visited where nobody was at home. If it is not possible to interview an eligible woman or man, record this on the Woman's or Man's Information Panel of the respective questionnaires. If it is not possible to interview a mother or primary caretaker, record this on the Under Five Child Information Panel of the Questionnaire for Children Under Five.

How to handle the interview

The interviewer and the respondent are strangers to each other; therefore, one of the main tasks of the interviewer is to establish rapport with the respondent. The respondent's first impression of you will influence her/his willingness to participate in the survey. Make sure that your appearance is neat and you also appear friendly as you introduce yourself.

On meeting the respondent, the first thing you should do is to introduce yourself, stating your name, the organization you are working for, the objectives of the survey, and what you want the respondent to do for you. You are advised to avoid long discussions on issues which are not related to the survey and which may consume a lot of your time.

After building rapport with the respondent, ask questions slowly and clearly to ensure the respondent understands what he/she is being asked. After you have asked a question, pause and give the respondent time to think. If the respondent feels hurried or is not allowed to form his/her opinion, he/she may respond with "I don't know" or give an inaccurate answer.

Specifically, the following guidelines will help you handle interviews: * Ensure that you understand the exact purpose of the survey and each question. This will help you to know if the responses you are receiving are adequate or relevant. * Remember the survey schedule, and remember that you are part of a team. Do not stay and talk for too long, but do not rush the interview either. * Ask the questions exactly as they are written. Even small changes in wording can alter the meaning of a question. * Ask the questions in the same order as they are given on the questionnaires. Do not change the sequence of the questions. * Ask all the questions, even if the respondent answers two questions at once. You can explain that you must ask each question individually, or say "Just so that I am sure…" or "Just to refresh my memory…", and then ask the question. * Help your respondents feel comfortable, but make sure you do not suggest answers to your questions. For example, do not 'help' a woman remember various contraceptive methods. Those cases when you are expected to 'help' the respondent, such as probing for answers or using information to remind the respondent of dates, ages, and durations are clearly indicated on the questionnaires, and are topics that are covered during your training. * Do not leave a question unanswered unless you have been instructed to skip it. Questions left blank are difficult to deal with later. When questionnaires arrive at the central office for editing and data entry, it may look as though you forgot to ask the question. Always write in '0' when a zero answer is given. For some questions, the code 'DK' will already be provided, and after you are sure that the respondent is unable to provide you with an answer, you will be able to circle this response. In questions where a 'DK' response is not printed on the questionnaire, you must make sure that the respondent provides an answer. In exceptional cases where this may not be possible, indicate this on the questionnaire with a note. * Record answers immediately when the respondent gives you the responses. Never rely on writing answers in a notebook for transfer to the questionnaire later. * Check the whole questionnaire before you leave the household to be sure it is completed correctly. * Thank the respondent for her/his cooperation and giving you time to interview her/him. Leave the way open to future interviews (for re-visits). Avoid over-staying in the respondent's household even if he/she is very friendly and welcoming.

General Points

Make a good first impression The first impression a respondent has of you is formed through your appearance. The way you dress may affect whether your interview is successful or not. Dress neatly and simply.

When first approaching the respondent, do your best to make her/him feel at ease. With a few well-chosen words, you can put the respondent in the right frame of mind for the interview. Open the interview with a smile and greetings and then proceed with your introduction as specified on your questionnaire.

If and when necessary, tell the respondent that the survey will help the development of plans for children and women and that his/her cooperation will be highly appreciated.

Gain rapport with the respondent Try not to arrive at the selected household at an inconvenient time of day, such as mealtimes, or too late or early during the day. Try to arrive when the respondents will not be too busy to answer questions.

Introduce yourself by name and show your identification. Explain the survey and why you want to do interview in the household, exactly as your introduction tells you to.

Be prepared to explain what is meant by confidentiality and to convince respondents to participate if they are reluctant.

Make sure that the respondents do not confuse you with others who might be visiting households for other reasons; for instance, for selling goods.

If the respondent refuses to be interviewed, note the reasons on the questionnaire, if possible.

Remain calm and polite at all times.

Always have a positive approach Never adopt an apologetic manner, and never approach with such words as "Are you too busy?". Such questions will invite refusal before you start. Rather, tell the respondent "I would like to ask you a few questions".

Stress confidentiality of information collected Always stress confidentiality of the information you obtain from the respondent. Explain to the respondent that the information you collect will remain strictly confidential and that no individual names will be used for any purposes, and that all information will be grouped together and depersonalized when writing the report. Use a language understandable by the respondent to get this message across. Never mention other interviews or read the questionnaire with other interviewers, the editor or the supervisor in front of a respondent or any other person. This will automatically erode the confidence the respondent has in you.

Probe for adequate responses You should phrase the question as it is in the questionnaire. If you realize that an answer is not consistent with other responses, then you should seek clarification through asking indirect questions or some additional questions so as to obtain a complete answer to the original question. This process is called probing. Questions, while probing, should be worded so that they are neutral and do not lead the respondent to answer in a particular direction. Ensure the meaning of the original question is not changed.

Pause and wait if the respondent is trying to remember difficult items.

Ask the respondent to clarify her/his answer if necessary. You may have misunderstood the response.

Check for consistency between the answers a respondent gives. Treat the questionnaires as tools that you are using to converse with the respondent. Try to understand and remember the responses, and if there is an inconsistency, ask the questions again. However, never point out to the respondents inconsistencies that you may have identified in a manner that may be understood as if you are testing the respondent's honesty or integrity.

Answering questions from respondent The respondent may ask you some questions about the survey or how he/she was selected to be interviewed or how the survey is going to help her/him, before agreeing to be interviewed. Be direct and pleasant when you answer. The respondent may also be concerned about the length of the interview. Please be frank to tell him/her how long you are likely to take to administer the questionnaire.

Interview the respondent alone The presence of a third person during the interview can prevent you from getting frank and honest answers from the respondent. It is, therefore, very important that the interviews are conducted privately and that all the questions are answered by the respondent only. This is especially important in the case of the Woman's and Men's Questionnaires, which include several topics that the respondents will consider to be "personal" or "private". If other people are present, explain to the respondent that some of the questions are private and request to talk to her/him while alone.

Handling hesitant respondents There may be situations where the respondent simply says "I don't know", or gives an irrelevant answer or acts in a manner suggesting he/she is bored or contradicts earlier answers. In all these cases, try your best to make him/her get interested in the question. Spending a few moments to talk about things unrelated to the interview (e.g. his/her town or village, the weather, his/her daily activities etc.) may be useful.

Adopt a non-judgemental attitude "Social desirability response bias" is a potential problem in surveys and refers to the tendency for respondents to present a favourable image of themselves to the interviewers. Sensitive questions may lead respondents to adjust their answers so as to appear politically correct or socially acceptable. Questionnaire items with strong social norms (such as adherence to religious or cultural expectations), or adopting attitudes/activities/objects that are widely considered desirable or undesirable tend to elicit "socially acceptable answers" rather than correct and honest answers. To minimise social desirability response bias it is very important to adopt a non-judgemental attitude and to not display any of your own attitudes, such as cultural or religious values, political preferences, and the like.

Role of the interviewer

Interviewers play a central role in the collection of data and the ultimate outcome of the exercise depends on how they conduct the interviews. Success, therefore, depends on the quality of the interviewers' work. It is, therefore, important that you are consistent in the way you put the questions to the respondent.

In case a response is not clear, you should probe further.

In general, the responsibilities of interviewers will include: * Locating the structure and households in the sample that are assigned to them, and administering the questionnaires * Identifying all the eligible respondents * Interviewing all the eligible respondents in the households assigned to them * Checking completed interviews to be sure that all questions were asked * Making re-visits to interview respondents who could not be interviewed during the first or second visit due to various reasons * Ensuring that the information given is correct by keeping the respondent focused to the questions * Including their specific observations or notes on the last page of each questionnaire * Preparing additional notes for the field editor and supervisor on other problems or observations

Instructions for Supervisors & Editors

Responsibilities . . . 61
Task . . . 62
Monetary Advances for Field Expenses . . . 62
Arranging transportations & accomodations . . . 62
Contacting local authorities . . . 63
Using maps to locate clusters . . . 63

IMPORTANT

This chapter is not written yet.

Responsibilities

The field supervisor is the senior member of the field team. He/she is responsible for the well-being and safety of team members, as well as the completion of the assigned work and the maintenance of data quality.

The specific responsibilities of the field supervisor are to make the necessary preparations for the fieldwork, to organize and direct the data collection in his/her assigned clusters, and to spot check the data collected in especially the Household Questionnaire.

Preparing for fieldwork requires that the field supervisor: (1) Obtains sample household lists and maps for each area in which his/her team will be working, discuss any special issues, such as potential security conditions in certain areas. (2) Becomes familiar with the area where the team will be working and determine the best arrangements for travel and accommodations. (3) Contacts local authorities to inform them about the survey and to gain their support and cooperation. (4) Obtains all monetary advances, supplies and equipment necessary for the team to complete its assigned interviews.

Task

Careful preparation by the field supervisor is important for facilitating the work of the team in the field, for maintaining interviewer morale and for ensuring contact with the central office throughout the fieldwork.

Organizing fieldwork requires that the field supervisor: (1) Assigns work to interviewers, taking into account the linguistic competence of individual interviewers, and assures that there is an equitable distribution of the workload. (2) Coordinate the work of the measurer by making sure he/she knows where to find the households that interviewers are conducting interviews in and approximately how many children and at what time a visit to the household should happen. (3) Maintains Cluster Control Sheets, and makes sure that assignments are carried out. (4) Makes spot checks of the Household Questionnaire (and individual questionnaires when appropriate) by conducting interviews according to the procedure described below. (5) Regularly sends completed questionnaires and progress reports to the fieldwork director and keeps headquarters informed of the team's location. (6) Communicates any problems to the fieldwork director. (7) Takes charge of the team vehicle, ensuring that it is kept in good repair and that it is used only for project work. (8) Makes an effort to develop a positive team spirit. A congenial work atmosphere, along with careful planning of field activities, contributes to the overall quality of a survey.

Monetary Advances for Field Expenses

The field supervisor should have sufficient funds to cover expenses for the team. Funds for team members should be distributed according to the procedures established by the survey director, if these have not been included in the per diem that is given directly to the interviewers.

The field supervisor should arrange for a system to maintain regular contact with the central office staff before leaving for the field. Regular contact is needed for supervision of the team by central office staff, payment of team members, and the return of completed questionnaires for timely data processing.

Arranging transportations & accomodations

It is the field supervisor's responsibility to make all necessary travel arrangements for his/her team, whenever possible, in consultation with the central office. The field supervisor is responsible for the maintenance and security of the team vehicle. The vehicle should be used exclusively for survey-related travel, and when not in use, should be parked in a safe place. The driver of the vehicle takes instructions from the field supervisor.

> VEHICLES ARE GENERALLY PROVIDED TO TRANSPORT THE TEAM TO ASSIGNED WORK AREAS. HOWEVER, IN SOME CASES, IT MAY BE NECESSARY TO ARRANGE FOR OTHER MEANS OF TRANSPORTATION, SUCH AS BOATS, HORSES, MULES, ETC. CUSTOMIZE THE PARAGRAPH ABOVE ACCORDINGLY.

In addition to arranging transportation, the field supervisor is in charge of arranging for food and lodging for the team. If they wish, interviewers may make their own arrangements, as long as these do not interfere with fieldwork activities. The lodging should be reasonably comfortable, located as close as possible to the interview area, and should provide a secure space to store survey materials. Since travel to rural clusters is often long and difficult, the field supervisor may have to arrange for the team to stay in a central location.

Contacting local authorities

It is the field supervisor's responsibility to contact the regional, district, local, and village officials before starting work in an area. Letters of introduction will be provided, but tact and sensitivity in explaining the purpose of the survey will help win the cooperation needed to carry out the interviews.

Using maps to locate clusters

> THIS SECTION PRESUMES THAT A FRESH HOUSEHOLD LISTING HAS BEEN CONDUCTED AND THAT UPDATED CLUSTER LOCATION AND SKETCH MAPS THEREFORE ARE AVAILABLE TO TEAMS.

A major responsibility of the field supervisor and the field editor is to assist interviewers in locating households in the sample. The fieldwork director will provide the supervisor with a copy of the Household Listing for the sample as well as base, location, and sketch maps of the clusters in which his/her team will be working. These documents will enable the team to identify the cluster boundaries and to locate the households selected for the sample. The representativeness of the sample depends on finding and visiting every sampled household.

Maps are generally needed during all stages of a survey, since they provide a picture of the areas in which interviews are to be carried out and help to eliminate errors, such as duplication or omission of areas. Moreover, maps help the team determine the location of sample areas, the distance to them, and how to reach selected households or dwellings.

Each team will be given general base and location maps, Household Listing Forms, and sketch maps, and written descriptions of the boundaries of selected areas. A cluster (i.e., PSU or EA) is the smallest working unit in any census or survey operation that can easily be covered by one enumerator. It has identifiable boundaries and lies wholly within an administrative or statistical area. The general base maps will show more than one cluster. Each cluster is identified by a number (for example, EA-010400105). Symbols are used to indicate certain features on the map such as roads, footpaths, rivers, localities, boundaries, etc. If symbols are shown on the map, the field supervisor and field editor should know how to interpret them by using the legend.

In most clusters, the boundaries follow easily recognizable land features such as rivers, roads, railroads, swamps, etc. However, at times, boundaries are invisible lines. The location and determination of invisible boundaries calls for some ingenuity, particularly in rural areas. If the location and sketch maps and descriptions do not provide enough detail, the following procedure is suggested:

In rural areas:

1. Identify on the map the road used to reach the cluster. When you reach what appears to be the cluster boundary, verify this by checking the location of actual terrain features and landmarks against their location on the map. Do not depend on one single feature; rather, use as many as possible.
2. It is usually possible to locate unnamed roads or imaginary lines by asking people living in the vicinity. In most cases, these people will know where the villages are and, by locating the villages, you can usually determine where the boundaries run. Local authorities may be helpful, as well as residents.
3. While there are cases in which boundaries shown on the map no longer exist (for example, they have been demolished), or have changed location (for example, a road has been relocated or a river has changed course), do not be hasty in jumping to conclusions. If you cannot locate a cluster, go on to the next one and discuss the matter later with the fieldwork director.

In urban areas:

1. There should be no problem with invisible lines, as urban areas generally have plenty of boundaries for use.
2. Street names in urban areas will often help you to locate the general area of clusters. Boundaries can be streets, alleys, streams, city limits, power cables, walls, rows of trees, etc.
3. Check the general shape of the cluster. This will help you find out if you are in the right place.
4. Read the written description.
5. You should locate all the cluster boundaries before you begin interviewing. For example, if the cluster is a rectangular block, the names of three boundary streets is not enough to unequivocally identify the cluster; check all four boundary streets.

Instructions for Managers

Data Protection Impact Assessment

UNHCR Data Protection Policy . . . 67
The 12 privacy principles . . . 68
Threat & vulnerability matrix . . . 69
Deleting dataset . . . 69
Technical measures necessary to comply with the policy . . . 70
Standard Operating Procedures necessary to comply with the policy . . . 70

IMPORTANT

This chapter is not written yet.

UNHCR Data Protection Policy

The Policy on the Protection of Personal Data of Persons of Concern to UNHCR has been issued in May 2015.

Data Protection Policy

In addition of the clarification of the basic principles to be applied when collecting information that includes **Personally Identifiable Information (PII)**, the policy recall the needs to carry out **Data Protection Impact Assessment (DPIA)**. A DPIA is required where the collection and processing or transfer of personal data is likely to be large, repeated or structural.

Bceause most of "Protection Assessment" fall under that policy as they are linked with the need to be able to process referals, personally identifiable information are often required. In addition, information such a precise coordinates of individuals with specific profile or needs might also fall unde the policy.

The 12 privacy principles

Data Protection Impact Assessments (PIAs) are an integral part of taking the "privacy by design" approach. This is done by enforcing the following 12 principles:

1. **Consent and choice**: presenting to the data subject the choice whether to allow the processing of their personally identifiable information (PII)
2. **Purpose legitimacy and specification**: ensuring that the purpose of data collection is specified and lawful
3. **Collection limitation**: limiting the collection of PII to that which is within applicable law and strictly necessary for the specified purpose(s)
4. **Data minimisation**: minimising the PII processed and the number of privacy stakeholders to whom PII is disclosed or who have access to it
5. **Retention and deletion**: ensuring that data is not kept for longer than is necessary for the purpose specified
6. **Accuracy and quality**: ensuring that the PII processed is accurate, complete, up to date (unless there is a legitimate basis for keeping outdated data), adequate and relevant for the purpose of use
7. **Openness, transparency and notice**: providing data subjects with clear and easily accessible information about the PII controller's policies, procedures and practices with respect to processing of PII
8. **Individual participation and access**: giving data subjects the ability to access and review their PII, provided their identity is first authenticated (Access and correction)
9. **Accountability**: assigning to a specified individual within the organisation the task of implementing the privacy-related policies, procedures and practices
10. **Information security**: protecting PII under an organisation's control with appropriate controls at the operational, functional and strategic level to ensure the integrity, confidentiality and

availability of the PII, and to protect it against risks such as unauthorised access, destruction, use, modification, disclosure or loss.

11. **Privacy compliance**: verifying and demonstrating that the processing of data meets data protection and privacy legislation by periodically conducting audits using internal or trusted third-parties
12. **Data transfers**: do not store or transfer personal data to third parties without adequate assurances that they will safeguard it to a standard comparable to that of the UNHCR.

Threat & vulnerability matrix

The template from the UK ICO can be used a starting point to developd the document.

An important point is to build the threat & vulnerability matrix:

Potential Threats to look at	Vulnerability	Risk	Mitigation
Cyber espionage			
Physical loss of data			
Technical failure			
Unauthorised acquisition			
DDOS attack / malware			
Insider privilege abuse			
Partner abuse			
Partner negligence			
Refugee complaints litigation			
Reputational damage			

Deleting dataset

Personal data that is not recorded in individual case files is not to be retained longer than necessary for the purpose(s) for which it was collected.

Though this should not minimise the need to save anonymised copies of the dataset in order to ensure potential new analysis of the data or longtidinal analysis. See the last chapter for more information on this, page 101.

Technical measures necessary to comply with the policy

- Maintaining physical security of premises, portable equipment, individual case files and records;
- Maintaining computer and information technology (IT) security, for example, access control (e.g. passwords, tiered access), user control, storage control, input control, communication and transport control (e.g., encryption).

Standard Operating Procedures necessary to comply with the policy

A DPIA would contain a general description of:

- the envisaged project,
- data sharing agreement in place
- arrangement involving processing of personal data (for instance clarififcation on protection referal),
- analysis of the risks to the rights of data subjects by virtue of the circumstances
- nature of the personal data processed,
- safeguards and security measures in place or proposed.

KoboToolBox Server Configuration

What is KoboToolBox? 71
Sign up and first login 72
Upload your form 72
Create users to collect data 72

IMPORTANT

This chapter is not written yet.

What is KoboToolBox?

KoBo Toolbox is an open-source tool for mobile data collection. It allows you to collect data in the field using mobile devices such as mobile phones or tablets, as well as with paper or computers.

The project has been orginally develoipped by the Harvard Humanitarian Initiative and was then adapted to humanitarian usage in colleaboration with OCHA.

KoBoToolbox was first created in 2009 and has grown over the years to include several projects and initiatives. It include: * FORMBUILDER: Easily create survey forms through an online user interface. * MOBILE DATA COLLECTION: Quickly and reliably collect data on Android, iOS, and many other devices, online or offline, in any language and with complex skip logic. * ANALYZING DATA: Inspect data moments after it was collected - and download it for advanced analysis in other software in Excel, CSV, KML, and other formats.

Sign up and first login

Visit https://kobo.unhcr.org and create a new account. After your account activation, through the link that was sent to you, you can log in to access your account.

Upload your form

You can now go to project and upload the form designed throught the question library.

Create users to collect data

More support on KoboToolBox is avaialble form this page

Data Entry on KoboToolBox

Using the client on Mobile Devices: koboCollect . . . 73
Server URL set up . . . 73
Collect data . . . 77
Uploading finalized data . . . 79
Using the web client: Enketo . . . 81
Collecting data offline . . . 82
Differences between the two options . . . 82

IMPORTANT

This chapter is not written yet.

KoBoToolbox allows data collection in multiple ways. Because KoBoToolbox is built on the Xform/ODK technology, our forms are compatible with a number of different tools that can be used for data collection. We recommend using KoBoCollect as the tool of choice for Android devices, and Web Forms (Enketo) for collecting data on all other devices (including iPhones, iPads, or any laptop or computer).

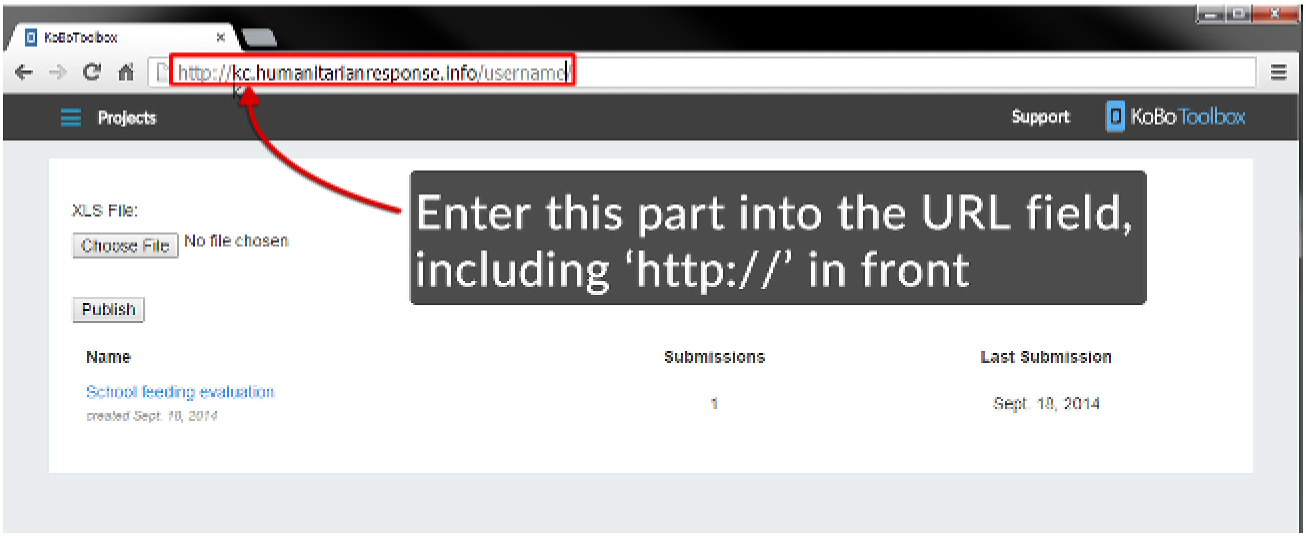
Using the client on Mobile Devices: koboCollect

KoBoCollect is an Android app that can be installed on any standard Android phone or tablet. To download the app to your Android device, click here.

Server URL set up

After installing KoBoCollect, you need to configure it so that it can be used together with your KoBoToolbox account for data collection. To do so, follow the steps:

- In KoBoToolbox, click on Projects within the menu sidebar on the left. Note the URL that is inside your browser at the top of your screen (for example, http://kobo.unhcr.org/username/)

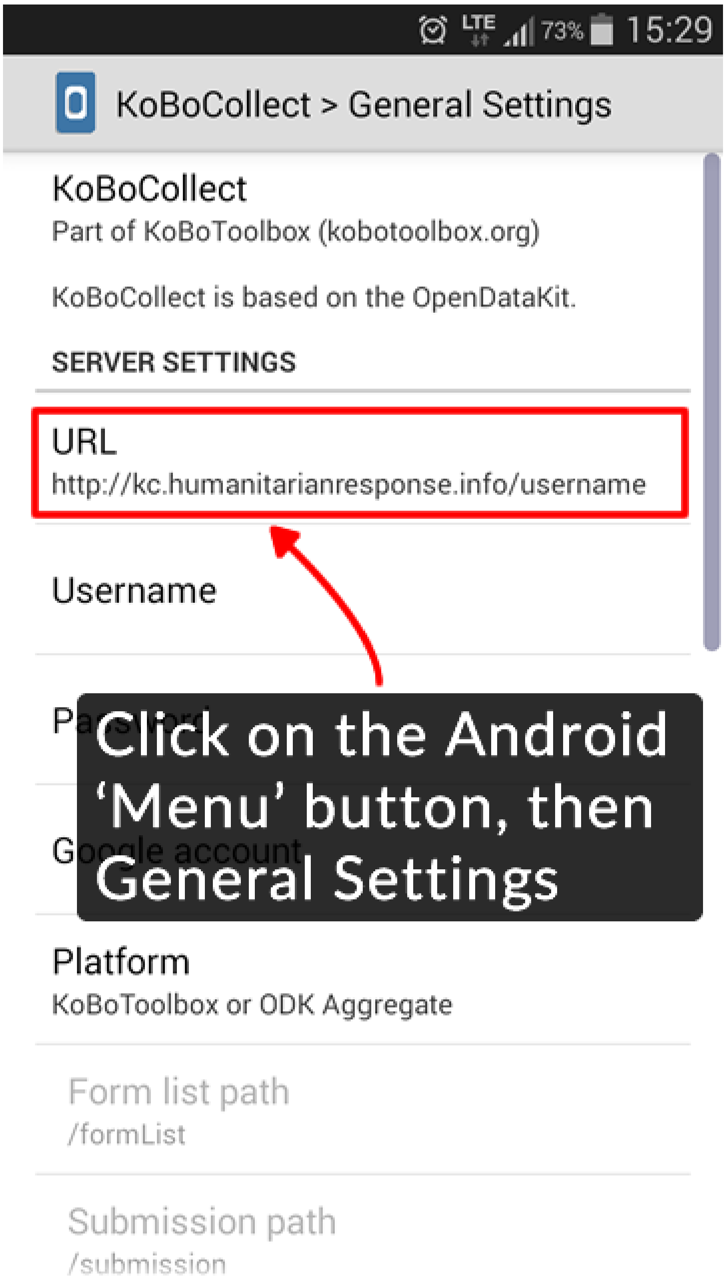


Offline

- On your Android device, open KoBoCollect and open the General Settings (click on the settings button of your device to access the settings).

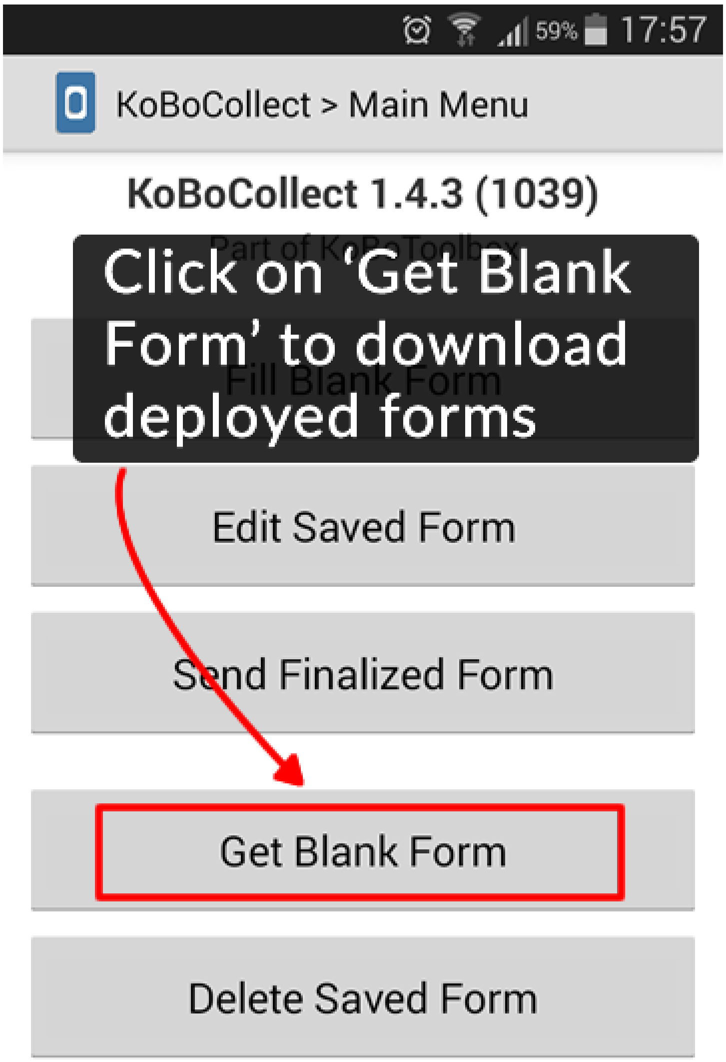
Offline

- In General Settings, under URL, enter the exact URL from step (2). Make sure you include the correct 'https://'



Offline

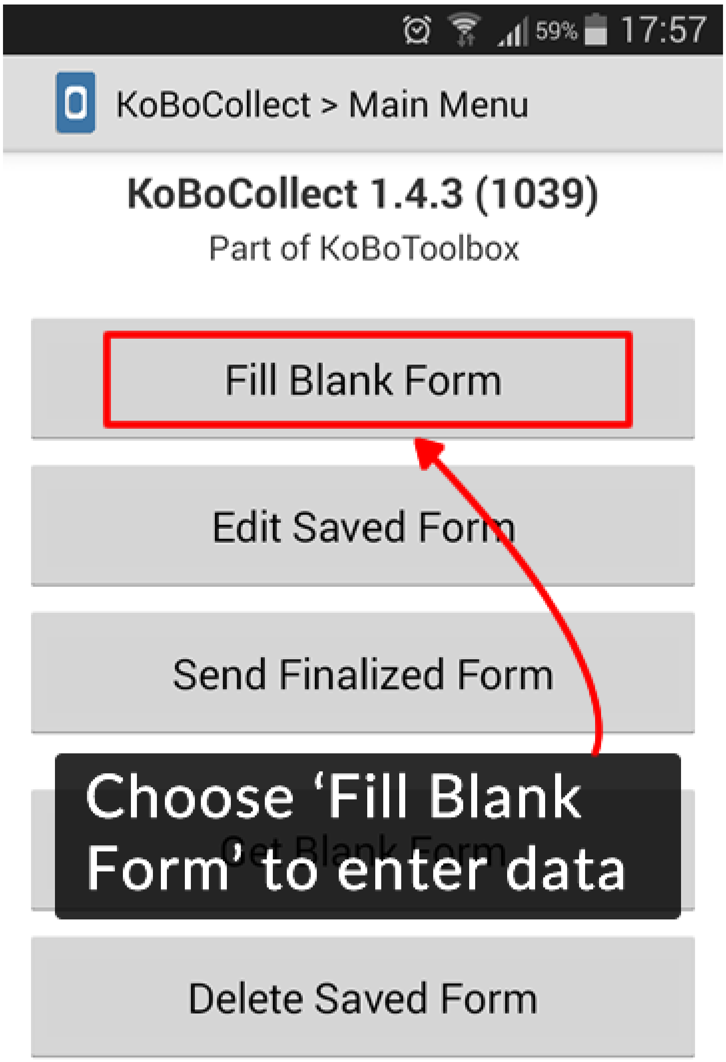
- Download forms from your account. Make sure you are connected to the Internet on your device. Also, you need to have deployed at least one project in KoBoToolbox. On the home menu of KoBoCollect, click Get Blank Form. A list of all your forms from your different projects will be shown. Click Toggle All (or select the ones you wish to download), then click Get Selected.



Offline

Collect data

- Click on Fill blank form

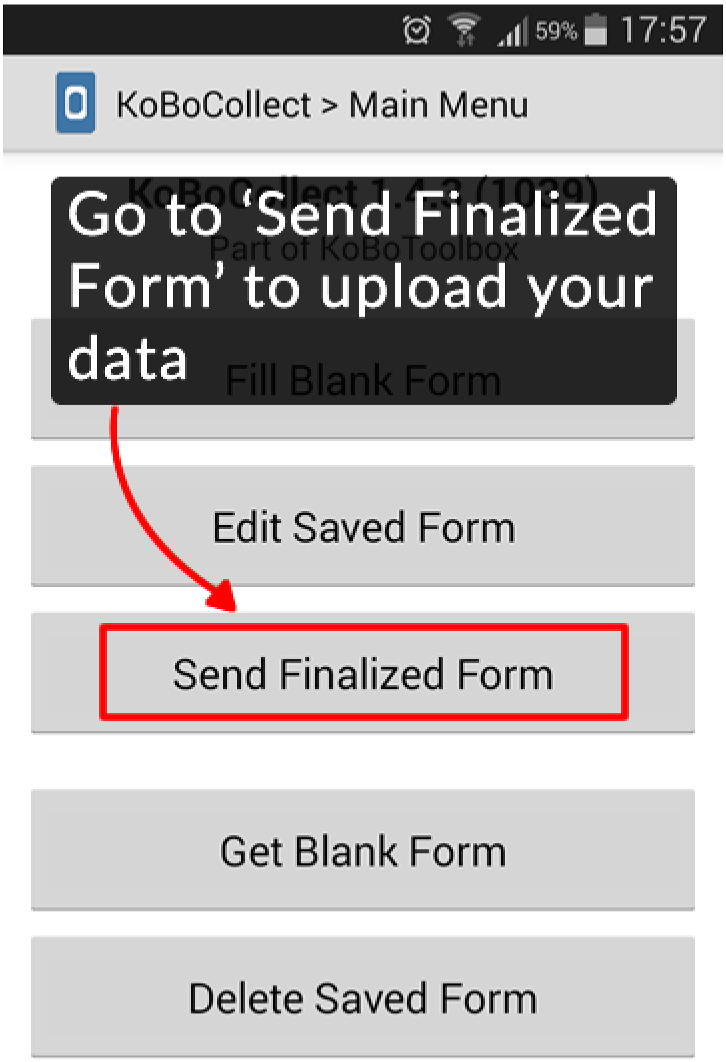


Offline

- Select the form to which you would like to enter data
- Go through all the questions (swiping your finger from right to left)
- At the end click on Save Form and Exit (making sure the form is marked as 'finalized')

Uploading finalized data

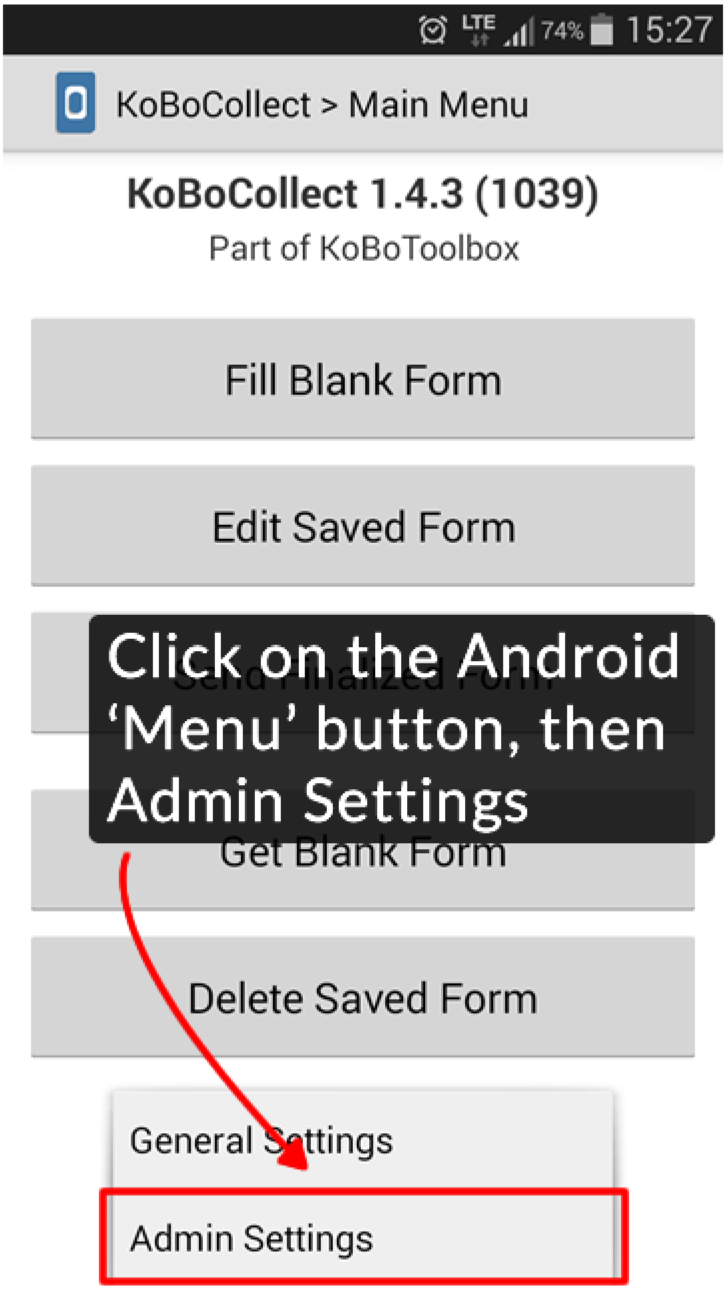
- From the home screen, click on Send Finalized Form



Offline

- A list of your most recently collected forms appears.
- Click Toggle all (or select the ones you wish to send), then click Send Selected.

It is possible to hide buttons and options within KoBoCollect. On the home screen click the Android menu button, then Choose Admin Settings and select the buttons you would like to hide from the different screens. If you set an admin password, interviewers won't be able to access the Admin settings to ever get access to these buttons.



Offline

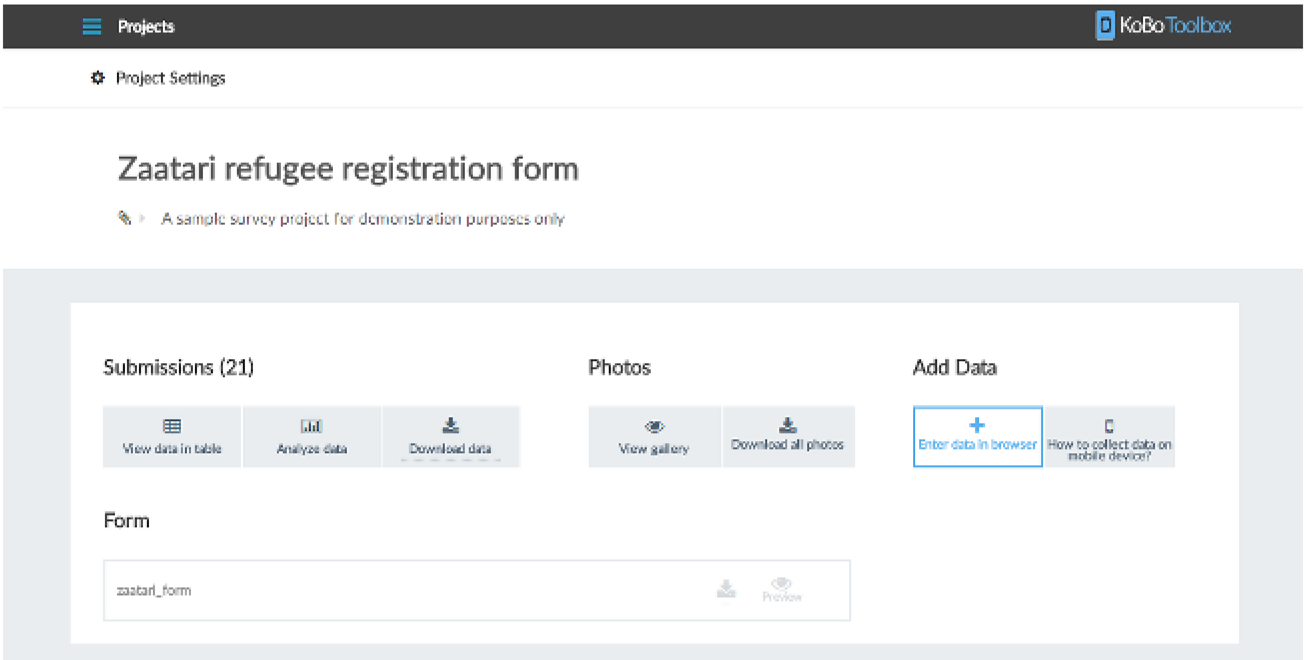
Using the web client: Enketo

Web Forms, also known as Enketo, are used by KoBoToolbox to preview forms and to enter data directly on a computer. Web forms also for collecting data on any mobile devices - even when offline at the time of data collection. It works on virtually any device, including iPhones, iPads, or any other smartphone, tablet, or computer. Some features are still being actively developed for Enketo, so some special questions may not be fully supported yet on every device. ### Start Collection

To start collecting data using web forms in your computer, simply click on the button 'Enter data in browser' in your project view. This is especially useful for testing purposes, but also when transcribing data from using paper forms.

Offline

To collect data using a mobile device, you need to copy the URL of your webform to your mobile device. You can simply send it by email or text message to any device. To obtain the URL of the webform, click either the 'Enter data in browser' or 'How to collect data on mobile device' buttons.

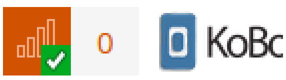


Offline

Collecting data offline

Enketo is also able to collect data while offline. However, it is essential to visit the URL once before going offline, and then saving it on the device (for example with a bookmark on the mobile browser).

Once your form has been fully loaded and cached, an offline availability icon (empty "signal bars" and a check mark) in the top-left corner will appear and indicate that the form can be accessed offline unless *browser's data are cleared*.



Offline

Enketo will display the form within that URL even without any Internet connectivity, and data will be store and queued until the next internet connection.

Differences between the two options

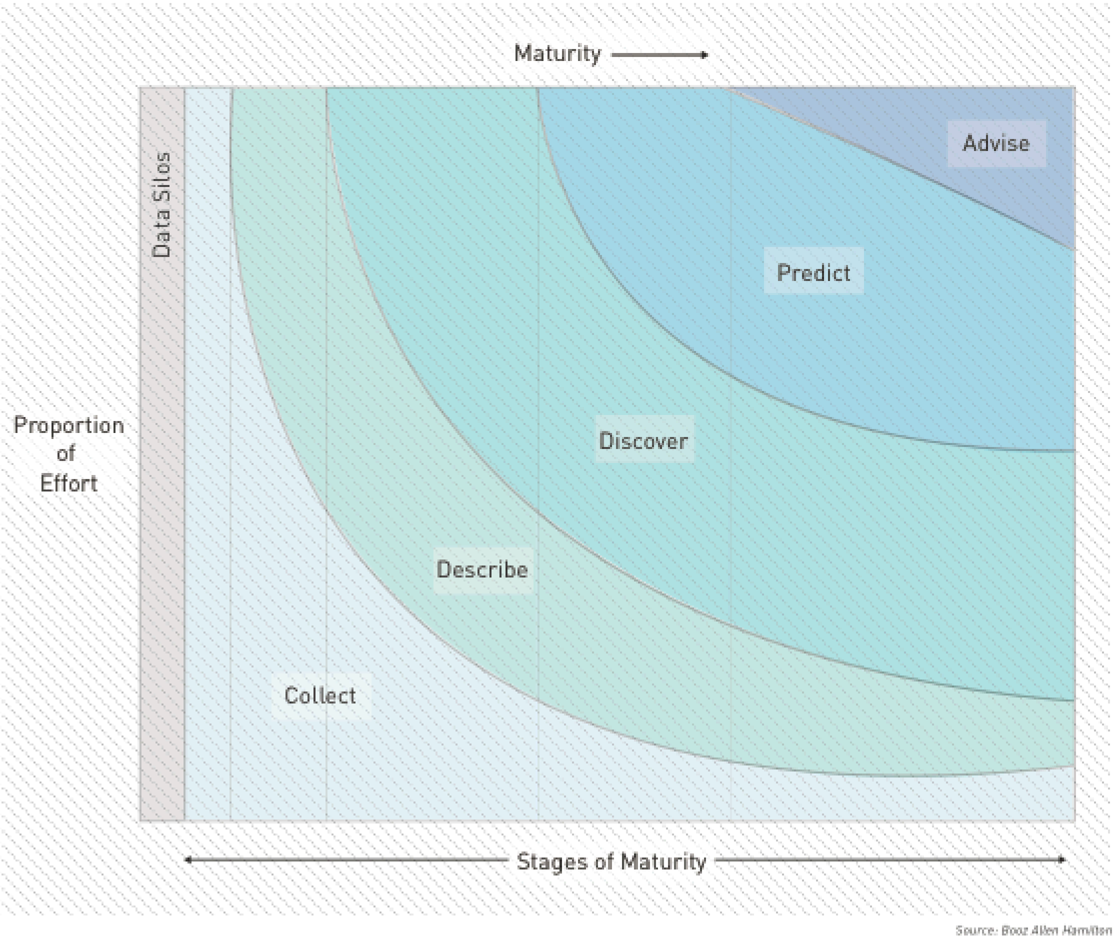
A detailled list of differences between the 2 options is here but in the differences are minimal and mostly impact the ammount of hidden metdata that can be collected as well the configuration of the look'n'feel of the form for end users.

Discover

Graphical review 85
Barchart 85
Histogramm 85
Maps 85
Statistical Test 85
Multivariate Analysis 85
Dimensionnality reduction 85
Clustering 86
Hierarchical Classification on Principle Components 86
Description of statistical clusters 86

IMPORTANT

This chapter is not written yet.



The Data Science Maturity Model

The first phase of the analysis consist at discovering the variable links.

Graphical review

Barchart

Histogramm

Maps

Statistical Test

Two random variables are called independent if the probability distribution of one variable is not affected by the presence of another. This is tested through the *Chi-squared Test of Independence* that allows to know if the independance hypothesis has a higher value than the .05 significance level.

The Chi-squared Test provide a *p-value*: if the p-value is greater than the .05, then the two tested variable are independent

Multivariate Analysis

Refugees profile are defined by multiple categories. However it is very difficult for the human brain to process more than 7 categories toggether. An important challenge to understand the profile of the population is to discover how categories interact together. Fortunately, since the 70's, Social scientist have developed technique that allow to discover statistical clusters among a specific population.

Multiple Correspondence Analysis (MCA) is a data analysis technique for nominal categorical data, used to detect and represent underlying structures in a data set.

Dimensionnality reduction

The first step of the analysis is to reduce the numbers of dimansion in order to represent each observation in a 2D space.

Clustering

Clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters). Refugee data are mostly categorical so clustering is done on the result of the Multiple Correspondence Analysis.

Hierarchical Classification on Principle Components

Description of statistical clusters

Predict

Regression Analysis 87

IMPORTANT

This chapter is not written yet.

Regression Analysis

Regression analysis is a statistical process for estimating the relationships among variables. Regression (and more specifically in the case of categorical data: logistic regression) can be used to predict certain characteristics or events linked to one household.

Advise

IMPORTANT

This chapter is not written yet.

Analysis Scripts

Defining Script for Analysis automation . . . 91
Categorical variables having several code options: . . . 92
Confidence intervals: . . . 93
Rounding decimal points: . . . 93
Decimal points in the results: . . . 93
Missing data or consent not provided: . . . 93
Always clean the data first before going into analysis: . . . 93
Age variable: . . . 93
Reproducible Analysis . . . 94

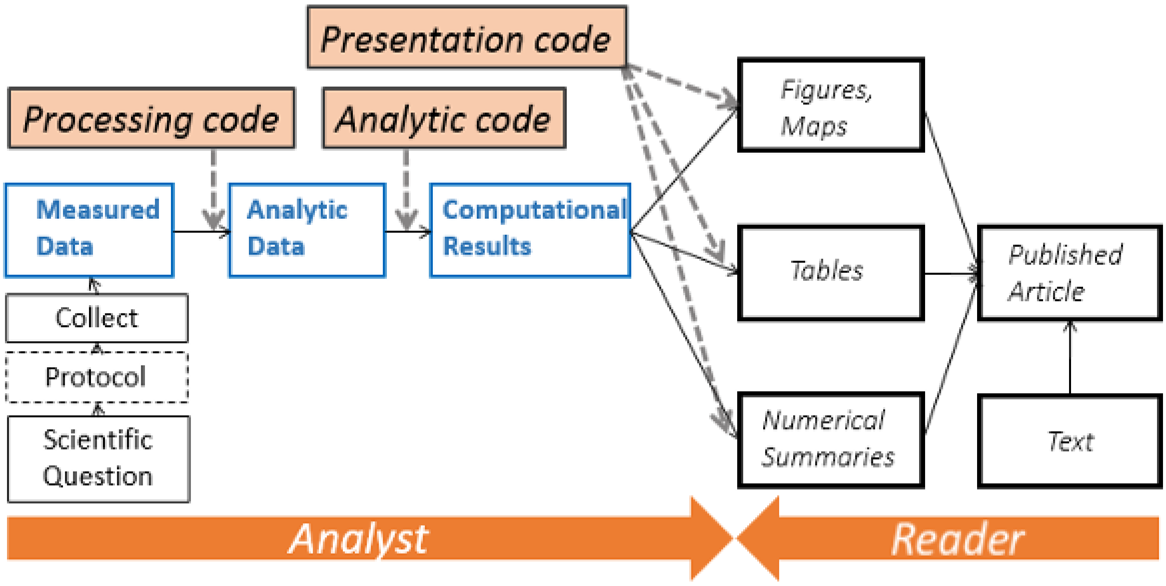
NOTE

Key Take Away :

Defining Script for Analysis automation

Using the right combination of packages fromt he R statistical language, it is possible to integrate all necessary data analysis steps into **scripts**:

- Data management (clean, recode, merge, reshape)
- Data analysis (test, regression, multivariate analysis, etc…)
- Data visualisation (plot, map, graph…)
- Writing up results (report and presentation generation)



This allow for reproducible data analysis workflow

Categorical variables having several code options:

- In the data cleaning process, make sure that the codes are the same as those on the standard questionnaire. E.g. vitamin A should be coded 1-3 in the questionnaire but code '4' may appear in the database. This code '4' needs to be excluded from analysis and not included in the denominator as it was a mistake in data recording on the field.

Code '8 or '98' for 'don't know':

- Make sure to exclude codes '8', '98' or '998' for 'don't know' from the analysis. This should not be part of the denominator in the calculation of indicators (e.g. diarrhoea, IYCF, ANC enrolment).

Code '6' or '96' for 'other':

- Make sure to include codes '6' or '96' for 'other' in the analysis. This should be part of the denominator in the calculation of indicators as it represents one response option (e.g. safe disposal of U3 stool, reason for not having a ration card, source of water).

Recoding with the 'if' command in Epi Info software:

- Do not forget to take into account missing values when using the 'if' command. This applies to many different variables. Whenever feasible, it is much better to use the 'recode' command instead of the 'if' command for this specific reason.

Confidence intervals:

- Different software often calculate confidence intervals as being negative and hence below zero or above 100. Negative confidence intervals or CI above 100 are meaningless. In the report, always round negative confidence intervals to ‘0’ and round those above 100 to ‘100’.

Rounding decimal points:

- Make sure to round properly decimal points according to basic rules:
- When decimal is between 1-4, round down.
- When decimal is between 5-9, round up.

Decimal points in the results:

- When the results is a whole number e.g. 30%, make sure to always write 30.0% with the ‘.0’ in the decimal place in the report. This ensures that the decimal point was not forgotten and is actually equal to zero.

Missing data or consent not provided:

- Data should be excluded from all analysis and should not be accounted for in the denominator.

Always clean the data first before going into analysis:

- Frequencies and means should be run on categorical and continuous variables, respectively.
- Missing data should be looked at and a record of them should be kept.

Age variable:

- When selecting age or creating an age variable category in Epi info software from the ‘months’ variable generated by ENA, don’t forget the ‘.99’ otherwise some children with an exact birth date may be excluded from the analysis. E.g. 6-23.99 (and not 6-23 or 6-23.9).

Reproducible Analysis

Always save newly generated variables into a new data file named following a naming convention to be respected by all involved in the survey data analysis.

Model for key Findings Report

IMPORTANT

This chapter is not written yet.

Model for Final Report

IMPORTANT

This chapter is not written yet.

Slides & Infographics

IMPORTANT

This chapter is not written yet.

Experience shows that it is easier for decision makers to see trends in data when it is visually presented in charts, maps and graphs. While valuable time should not be wasted making a visually flawless final report, attracting the readers' eyes to key messages and actionable information is strategic.

Organise Microdata for Social Scientist

Data sharing for research . . . 101
Anonymization techniques & Statistical disclosure control (SDC) . . . 102
Sharing via a safe mechanism: File encryption . . . 102
Restricting access to data . . . 103
Engaging in "Research Agreement" . . . 103
Research Confidentiality . . . 103
Reproducible research . . . 104
The Internation Household Survey Network & the DDI format . . . 105

IMPORTANT

This chapter is not written yet.

Data sharing for research

> The UNHCR and `Partner Name` will identify the staff to be part of the joint research team. Any data shared under this agreement will not be provided to any third party. For its part, UNHCR agrees to share defined and agreed upon data with the `Partner Name` for the purposes of the `Partner Name` and UNHCR collaboration on this project herein-defined as "`Project Name`". All information that

> would allow for identification of individuals will be excluded from these datasets, e.g. refugee ID number. UNHCR will share this information via a safe mechanism to reduce the likelihood of a third party accessing the data unlawfully. `Partner Name` will specify by name and title who will receive the information, who will have access to the information, and where the information will be kept, e.g. individual personal computer or server, all with the intent to avoid unlawful access and use of the information. Once the information is used for its defined purpose, the data will be disposed of at a date determined and in agreement by the two parties.

Anonymization techniques & Statistical disclosure control (SDC)

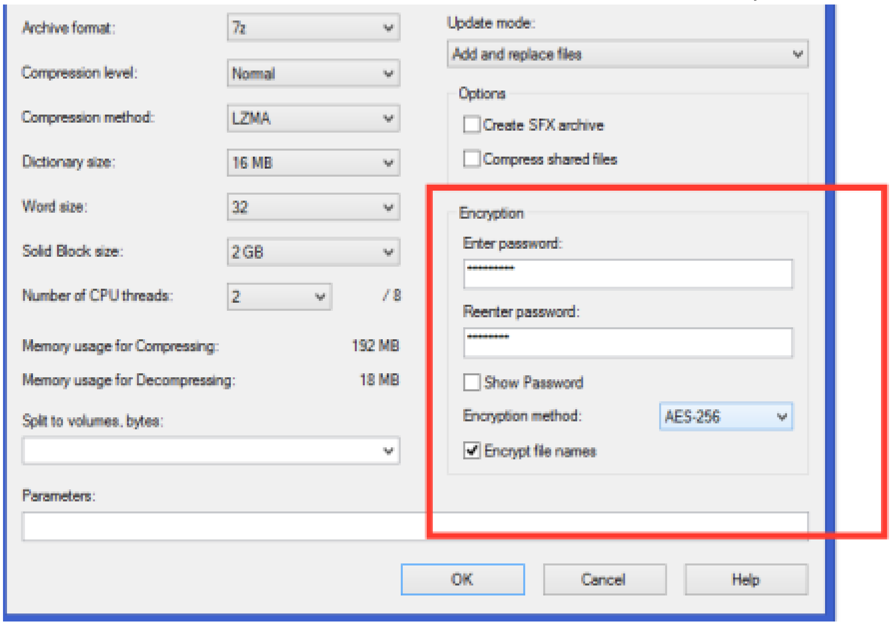
Once anonymised, a dataset does not fall anymore under the Policy on the Protection of Personal Data. Though there's a few articles about the failure of anonymization that shows how removing names & ID is not always sufficient to prevent "data re-identification". Many techniques can be used for "statistical disclosure control": suppression, inference control, banardisation, rounding or sampling. Other approaches includes rules like for instance "do not share figures for a spatial unit if it does not reach the 1000 refugees threshold"…

A dedicated R module is availalble to perform anonymisation.

Sharing via a safe mechanism: File encryption

What is a safe mechanism to share information: for instance which software to use for encryption, how to share password, etc. Potential requirements could include: - Use a well know encryption approach – The common standard si AES -Advanced Encryption Standard (AES) - Rely on open source software – so both parties can easily encrypt & decrypt without being tied to software procurement obstacle. - Combine encryption and file compression: so files are easier & lighter to share - The password used for the encryption should be at least 10 character long with a mixture of lowercase and uppercase alphabetic character, numbers and symbols. This should allow to build what is commonly called a strong password and should always be transmitted independently form the file (for instance on a separate paper sheet with no reference to the file it allows to open).

In terms of software, it is possible to use 7zip.



A sumarry of the principle above woud be: *Data files should be encrypted with AES-256 method using a strong password (at least 10 character long with a mixture of lowercase and uppercase alphabetic character, numbers and symbols) and compressed using the 7zip format with the 7zip software. Password will be transmitted printed on a paper that will need to be secured by the receiving agency*.

Restricting access to data

Need to set up a standard registry of person who work on UNHCR datasets

Engaging in "Research Agreement"

Research Confidentiality

A written and legally-binding Confidentiality Agreement must be signed by the lead researcher, all members of the research team that will have access to individually identifiable information from the records. The agreement coudl include the folowwing points:

> `Analysis Project Title` Principal Investigator: `UNHCR`
>
> *I,* `Resesarcher Name`*, from* `Resesarch Organisation Name`*, as a member of this research team, understand that I may have access to confidential information about study sites and participants. By signing this statement, I am indicating my understanding of my responsibilities to maintain confidentiality and agree to the following:*
>
> 1. *keep all the research information shared with me confidential by not discussing or sharing the research information in any form or format (e.g., disks, tapes, transcripts) with anyone other than the Researcher(s).*
> 2. *keep all research information in any form or format (e.g., disks, tapes, transcripts) secure while it is in my possession.*
> 3. *return all research information in any form or format (e.g., disks, tapes, transcripts) to the Researcher(s) when I have completed the research tasks.*
> 4. *after consulting with the Researcher(s), erase or destroy all research information in any form or format regarding this research project that is not returnable to the Researcher(s) (e.g., information stored on computer hard drive).*
> 5. *notify the local principal investigator immediately should I become aware of an actual breach of confidentiality or a situation which could potentially result in a breach, whether this be on my part or on the part of another person.*

Reproducible research

To ensure that research done on the dataset can be reproduced afterwards by internal staff both to check them and to refresh the analysis when we have new data a series of good practices shoudl be implemented:

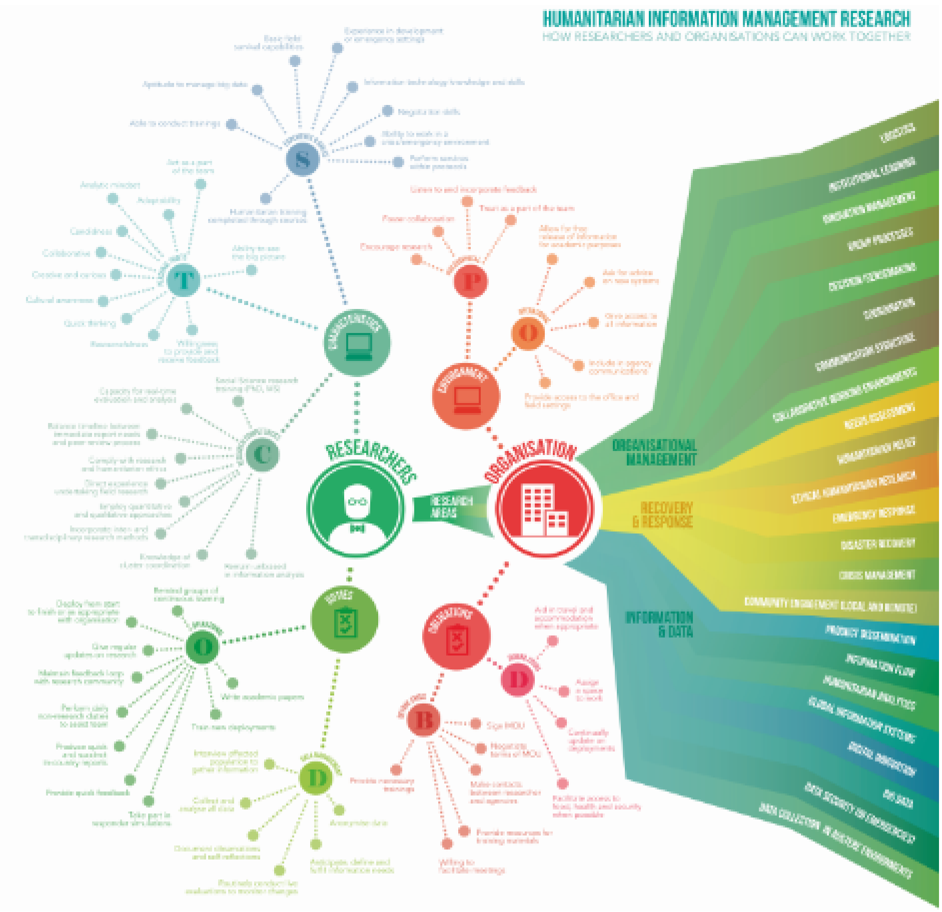
1. For every result, **keep track** of how it was produced
2. **Avoid manual data manipulation** steps
3. **Archive** the exact versions of all external programs used
4. **Version control** all custom scripts
5. **Record all intermediate results**, when possible in standardized formats
6. For analyses that include randomness, **note underlying random seeds**

7. Always **store raw data** behind plots
8. Generate hierarchical analysis output, allowing layers of increasing detail to be inspected
9. Connect **textual statements** to underlying results
10. Provide **public access** to scripts, runs, and results

The Internation Household Survey Network & the DDI format

Humanitarian Research in the context of social science and data analysis is still new but can benefit the organisedtion for instance to:

- Co-development and co-design of tools, protocols, products, processes, and innovations
- Facilitate organisational learning, keeping track of lessons learned, and providing a neutral stance for moderating innovation and change processes
- Access to wider body of knowledge, from academia or other organisations, and research in other fields.



7zip

To facilitate this process, the first approach woudl be to document the dataset according to the Data Documentation Initiative (DDI) metadata standard developped by the International Household Survey Network (IHSN).

Once the metadata are generated in the right format, it becomes possible to publish them within the ISHN Microdata catalog or the World Bank Microdata Library