

MSNA - Research Terms of Reference

Food Security Among Urban Populations

SOM 2604 Somalia

July 2026

REACH Informing
more effective
humanitarian action

1. Executive Summary

A. General information						
Country of intervention	Somalia					
Type of Emergency	☒	Natural disaster	☒	Conflict	☐ Other (<i>specify</i>)	
Type of Crisis	☐	Sudden onset	☐	Slow onset	☒ Protracted	
Mandating Body/ Agency	Food Security Cluster IPC Core Group					
IMPACT Project Code	27BMF					
Overall Research Timeframe)	25/06/2026 to 29/08/2026					
Research Timeframe	1. Data Analysis Plan (DAP) sent for validation: 25/06/2026		7. Preliminary presentation/joint analysis workshop: 17/08/2026 – 29/08/2026 (IPC post-Gu analysis workshop)			
	2. Pilot/training: Training of Trainers (ToT): 12/07/2026 Enumerator Training + Pilot: 15/06/2026 – 19/06/2026		8. Key Findings brief validation: 08/10/2026			
	3. Start data collection: 21/07/2026					
	4. Data collected: 01/08/2026					
	5. Clean dataset sent for validation: 06/08/2026					
	6. Data analysis sent for validation: 09/08/2026					
Humanitarian milestones <i>Specify what will the assessment inform and when e.g. The shelter cluster will use this data to calculate</i>	Milestone			Deadline		
	☐	Donor plan/strategy			_/_/_/_	
	☐	Inter-cluster plan/strategy			_/_/_/_	
	☐	PiN calculation / HNRP			_/_/_/_	
	☒	IPC (Integrated food security Phase Classification)			24/09/2026	
	☐	Cluster plan/strategy			_/_/_/_	
	☐	NGO platform plan/strategy			_/_/_/_	

PiN numbers for the HNO analysis	☐	Other (Specify):	_/_/_/_/_	
Audience Type & Dissemination Specify who will the assessment inform and how you will disseminate to inform the audience	Audience type		Dissemination	
	☒ Strategic ☐ Programmatic ☐ Operational ☐ [Other, Specify]		☒ General Product Mailing (e.g. mail to NGO consortium; HCT participants; Donors) ☐ Cluster Mailing (to all ICCG members) and presentation of findings at next cluster meeting ☒ Presentation of findings (e.g. at ICCG meeting; Cluster meeting) ☒ Website Dissemination (Relief Web & REACH Resource Centre) ☒ IPC post-Gu analysis workshop ☐ [Other, Specify]	
Detailed dissemination plan required	☒	Yes	☐	No
General Objective	To inform the Integrated Food Security Phase Classification (IPC) in Somalia, particularly the post-Gu analysis workshop in late August 2026, this assessment aims to provide updated multi-sectoral analysis, focussed on food security and contributing factors, for the IPC. This analysis will focus on the severity of needs among the crisis-affected population in the context of ongoing climatic shocks and the protracted displacement crisis. The goal is to contribute to a more targeted and evidence-based response.			
Specific Objective(s)	- Generate statistically representative, household-level evidence on food security outcomes among host communities and IDPs in selected urban areas to support the Acute Food Insecurity (AFI) workstream of the 2026 IPC post-Gu analysis. - Assess the severity and prevalence of factors contributing to food insecurity, including livelihood disruption, market access, water insecurity, and access to assistance. - Identify differences in food security outcomes and contributing factors between population groups and across Units of Analysis.			
Research Questions	- What is the severity and prevalence of acute food insecurity among assessed urban host community and urban IDP households, and how does this vary across assessed Units of Analysis? - To what extent are households experiencing livelihood disruption, income loss, asset depletion, reduced purchasing power, and barriers to accessing markets and essential goods, and how are these factors associated with food insecurity? - What are other key humanitarian needs affecting crisis-affected households that contribute to food insecurity, and how do these vary in nature and severity across population groups and Units of Analysis?			
Geographic Coverage	This assessment aims for geographic representation of selected urban areas in south and central Somalia. This will be achieved at the Unit of Analysis (UoA) level to inform the IPC: trained enumerators will conduct in-person surveys in 8 UoA in			

	Bay, Bakool, and Gedo regions selected by the IPC group to ensure representative results for urban host population and urban IDPs. These UoA were identified respecting several conditions, including: (1) priority areas IPC technical working group partners are not covering, (2) the ability to be accessed by REACH field team, (3) not covered with primary data collection for the April 2026 update analysis , and (4) the existence of urban livelihoods (e.g., pastoralist and rural areas would not count).			
Secondary data sources	See Section 3.3 Secondary Data Review			
B. Sampling				
Population groups <i>Select all population group which your assessment will collect data on</i>	☒	IDPs in camp	☐	IDPs in informal sites
	☐	IDPs in host communities	☐	IDPs [Other, Specify]
	☐	Refugees in camp	☐	Refugees in informal sites
	☐	Refugees in host communities	☐	Refugees [Other, Specify]
	☒	Host communities	☐	[Other, Specify]
Structured questionnaire (Quantitative) – <i>Select all the apply</i>	☒	Probability sampling	☐	Non - Probability sampling
	☐	Individual	☒	Household
Data collection level:	☐	Settlement	☐	Other (specify): _ _ _ _ _
If probability sampling				
Sampling method: ☐ Random sampling ☒ Cluster sampling (2 stages random) The sampling is stratified: ☒ Yes ☐ No If yes what are the stratifications: ° Geographic: Units of Analysis (according to livelihood zones) ° Population groups: Urban host community and urban IDPs What is the Primary sampling unit (PSU): 1 km ² hexagons in urban areas, covering settlements / IDP sites If cluster sampling, what is the minimum cluster size? 8 Sampling frame: Do you have the population number at PSU level for all population groups? ☒ Yes ☐ No Selection: Probability Proportional to Size (PPS) : ☒ Yes ☐ No Selection of PSUs with replacement? : ☒ Yes ☐ No Aimed precision at stratification level: 90% level of confidence 10+/- % margin of error				

Buffer: 10%													
Total sample size: 1,088													
Resampling: Do you have a reserve list of PSUs / households in case of inaccessible area ? ☒ Yes ☐ No													
Data collection method: ☒ Face to face ☐ Remote data collection													
C. Questionnaire													
Questionnaire design	For Food Security, all Tier 1 indicators from the Indicator Bank are included. Please refer to 'Annex 1: Modification to the Core Indicators' for any modifications, and to the DAP for a full list of all included indicators. XLSform for mandatory indicators The kobo questionnaire provided for the mandatory indicators was used without alteration: ☐ Yes ☒ No												
Data management platform(s)	☒ IMPACT ☐ UNHCR ☐ Other, Specify												
Expected output type(s)	☒ Presentation (Preliminary findings) #: 1 ☐ Interactive dashboard #: __ ☐ Report #: __ ☐ Profile #: __ ☐ Presentation (Final) #: __ ☐ Webmap #: __ ☒ Factsheet : 1 ☒ Map #: at least 1 (coverage map plus any maps per sector) ☐ Other, Specify] #: __												
Data publication plan	<table border="1"> <tr> <td>☒</td> <td>Final (anonymised) dataset public, available on REACH resource center</td> </tr> <tr> <td>☐</td> <td>Final (anonymised) dataset public, through HDX connect</td> </tr> <tr> <td>☒</td> <td>Analysis table public, available on REACH resource center</td> </tr> <tr> <td>☐</td> <td>Analysis table public, available on HDX</td> </tr> </table>	☒	Final (anonymised) dataset public, available on REACH resource center	☐	Final (anonymised) dataset public, through HDX connect	☒	Analysis table public, available on REACH resource center	☐	Analysis table public, available on HDX				
☒	Final (anonymised) dataset public, available on REACH resource center												
☐	Final (anonymised) dataset public, through HDX connect												
☒	Analysis table public, available on REACH resource center												
☐	Analysis table public, available on HDX												
Visibility Specify which logos should be on outputs	<table border="1"> <tr> <td colspan="2">REACH</td> </tr> <tr> <td colspan="2">Donor:</td> </tr> <tr> <td colspan="2" style="text-align: center;">  </td> </tr> <tr> <td colspan="2">Coordination Framework:</td> </tr> <tr> <td colspan="2" style="text-align: center;">  </td> </tr> <tr> <td colspan="2">Partners:</td> </tr> </table>	REACH		Donor:				Coordination Framework:				Partners:	
REACH													
Donor:													
													
Coordination Framework:													
													
Partners:													

	Acted, Food and Agriculture Organisation (FAO), FSNAU, International Rescue Committee (IRC), IOM - UN Migration, UNICEF, World Food Programme (WFP)
--	---

2. Rationale

2.1 Background

Somalia's protracted humanitarian crisis continues to be shaped by armed conflict and climate shocks. Instability and conflict remain [widespread](#) across much of the country, with [high levels](#) of conflict events and fatalities persisting over the past 12 months. This in turn affected humanitarian actors' ability to reach affected populations, with conflict being the [primary driver](#) of access constraints.

Concurrently, climate volatility continues to impact the situation. As Somalia emerges from a [drought](#) that began during the Deyr rainy season (October – December 2025) and [carried over](#) into the Jilaal dry season (January – March 2026), household recovery has been [hampered](#) by an [erratic](#) Gu rainy season (April – June 2026), while [FAO projects](#) heightened flood risk for the latter half of 2026.¹

The climate-conflict nexus is [fuelling displacement](#) and extreme vulnerability, with [6 million people](#) facing acute food insecurity (IPC Phase 3+) and 1.9 million children under five suffering acute malnutrition as of June 2026. These conditions [further exacerbate](#) active outbreaks of communicable diseases like acute watery diarrhoea (AWD)/cholera, measles, and diphtheria.

Furthermore, an international hydrocarbon supply shock has led to elevated prices across the country, putting pressure on [household purchasing power](#) and [inhibiting humanitarian operations](#). Meanwhile, as households contend with several overlapping shocks, relief funding remains [constrained](#), with the hyper-prioritised 2026 Humanitarian Needs and Response Plan only 20% funded as of June 2026.

The multifaceted nature of these challenges requires a comprehensive humanitarian response, though critical information gaps exist, with more recent data required to account for dynamic and changing conditions. This assessment addresses these gaps via multi-stage cluster sampling across selected areas to inform the IPC post-Gu analysis, collecting data on food security and contributing factors.

2.2 Intended impact

This assessment is focused on food security in urban areas, aiming to provide statistically representative data on food insecurity and contributing factors across two population groups (host communities and IDPs) in three regions (Bay, Bakool, and Gedo) for the Integrated Food Security Phase Classification (IPC) analysis in August 2026. By providing a common scale for the severity of food insecurity, this analysis guides critical interventions and informs the decision-making and response planning of several stakeholders, including government actors, humanitarian agencies, and donors.

The assessment will provide robust, evidence-based insights into the severity and drivers of need at the sub-national level, enabling stakeholders to identify priorities across key sectors and focus areas – including Food Security & Livelihoods, WASH, Cash & Markets, Nutrition, Health, and AAP – and to tailor responses to the specific vulnerabilities of affected population groups. In doing so, it contributes to more coordinated, efficient, and needs-based humanitarian action, supporting improved targeting

¹ There is high confidence in El Niño development ([Columbia University](#), [Copernicus](#), [WMO](#)), with some climate models predicting a particularly strong El Niño ([NOAA](#), [ECMWF](#), [The Climate Brink](#)). El Niño conditions are typically associated with above-average Deyr rainfall across Somalia ([Columbia University](#), [EU JRC](#)), which can impact the Deyr planting and growing season ([WFP](#)). The most recent El Niño spanned from spring 2023 to spring 2024 ([JMA](#)), during which Somalia experienced [severe flooding](#) that drove food insecurity ([FEWS NET](#), [IPC](#)) and caused 1.2 million displacements ([UNDP](#)). [Historical flood data](#) and [river breakage monitoring](#) from FAO SWALIM suggest that many areas are vulnerable to flood risks.

and optimised resource allocation for the most vulnerable households in crisis-affected areas of Somalia.

3. Methodology

3.1 Methodology overview

The assessment is designed to ensure geographical coverage across eight Units of Analysis (UoAs) in Somalia, with the objective of collecting representative findings for two distinct population groups – urban host communities and urban IDPs living in sites – to inform IPC food security analysis. These findings will be representative at the UoA level, with a 90% confidence level and 10% margin of error. To meet the required sample sizes for representativeness, a 10% buffer will be added to the minimum calculated sample. Data collection will take place in eight UoAs (see Table 1) that were selected in close coordination with IPC core group members to avoid duplication and target priority areas. Surveys will be conducted through in-person household interviews by trained REACH enumerators.

Units of Analysis (e.g., Gedo urban host communities) refer to a defined population group (e.g., urban host communities) within a specified geographic area (e.g., Gedo region). Geographic areas are delineated at the administrative level through a collection of districts selected to align with livelihood zones,² under the assumption that the population living there is exposed to similar contextual shocks and exhibits similar livelihood patterns. For the purposes of this assessment, only urban areas within these districts comprise the Unit of Analysis. These standardised geographic-demographic units serve as the primary stratum for data collection, enabling comparable, population-representative analysis of humanitarian conditions across distinct population groups within defined boundaries.

This assessment's Units of Analysis follow those IPC uses for analysis and reporting, enabling direct data harmonisation, though in some cases surveys may not be conducted in certain areas due to inaccessibility at the time of data collection. IPC classifies crisis severity at the UoA level to prioritise interventions, reporting severity levels (Phase 1-5) per UoA (e.g., 'Bay Urban IDPs face IPC Phase 4'). This approach ensures humanitarian action is precisely targeted to the most vulnerable groups within their specific geographic and demographic contexts. For IPC's methodological details, refer to the [IPC Technical Manual](#).

Table 1. Units of Analysis (to inform IPC food security analysis)

	Units of Analysis	Districts ³
1	Bay – Urban Host Community	Bur Hakaba, Diinsor, and Qansax Dheere

² Livelihood zones are 'geographical areas within which people share basically the same patterns of access to food and have the same access to markets' (FSNAU). Defining areas by livelihood patterns is an integral step to (1) understanding the systems that link people with food supplies, and (2) analysing how these baseline scenarios might be affected by shocks (Save the Children, FEWS NET). For the mapping of Somalia's livelihood zones, see FEWS NET; on the process of delineating these areas, see FSNAU. Capturing urban livelihoods has been a consistent facet of IPC analysis in Somalia (IPC), and baseline profiles exist for several of the country's urban areas, including Baidoa, Bosasso, Galkacyo, and Garowe.

³ This refers to the constituent districts of the Units of Analysis selected for data collection through consultations with IPC partners; surveys will only be carried out in accessible urban areas of each district.

2	Bay – Urban IDPs	Bur Hakaba, Diinsor, and Qansax Dheere
3	Gedo – Urban Host Community	Baardheere, Belet Xaawo, Ceel Waaq, Garbahaarey, and Luuq
4	Gedo – Urban IDPs	Baardheere, Belet Xaawo, Ceel Waaq, Garbahaarey, and Luuq
5	Bakool – Xudur Urban Host Community	Xudur
6	Bakool – Xudur Urban IDPs	Xudur
7	Bakool – Urban Host Community (without Xudur)	Ceel Barde, Waajid
8	Bakool – Urban IDPs (without Xudur)	Ceel Barde, Waajid

3.2 Population of interest

In alignment with the IPC, target populations for this assessment are: urban host communities and urban IDPs living in sites. A host community is the local population of the community in which displaced persons temporarily reside ([IOM](#), [UNHCR](#), [Somalia Shelter Cluster](#)).

According to the [Guiding Principles on Internal Displacement](#), IDPs are defined as ‘persons or groups of persons who have been forced or obliged to flee or to leave their homes or places of habitual residence, in particular as a result of or in order to avoid the effects of armed conflict, situations of generalised violence, violations of human rights or natural or human-made disasters, and who have not crossed an internationally recognised state border’.

Somalia’s [National Policy on Refugee-Returnees and Internally Displaced Persons \(IDPs\)](#) follows this definition, but adds two additional categories of IDPs: (1) people who were forcibly evicted from their settlement and who have not received adequate housing/land alternative/appropriate compensation, and (2) pastoralists who have lost access to their traditional nomadic living space through loss of livestock or loss of access to grazing land or water points or markets, and have therefore left their habitual living space.

3.3 Secondary data review

See secondary data review matrix in [Annex 2](#).

3.4 Primary Data Collection

Cleaning the IDP site Master List for sampling

To support the sampling process, the updated CCCM Q1 2026 IDP Site Master List served as the primary data source. The lack of reliable secondary sources for triangulating population data necessitated a two-step cleaning process to refine the master lists for sampling:

1. **Geospatial Cleaning:** The GIS team mapped all households’ locations and sites, considering the population groups expected to be addressed in each location. This step aimed to verify whether the coordinates listed in the master list accurately corresponded to actual settlement locations.

2. **Field Site Verification:** Field officers worked closely with enumerators to assess the accessibility of each site, classifying them as accessible, inaccessible, or of unknown accessibility. Given the protracted and dynamic nature of the conflict in Somalia, this step was essential to determine which sites could feasibly be reached by the field team. Accessibility was defined based on logistical feasibility (e.g., road and bridge access), absence of non-state armed actors, and absence of active clan conflicts.

This two-step cleaning process was designed to ensure accurate planning for data collection and to minimise the need for resampling during fieldwork.

Sampling methodology

The sampling approach focuses on targeted assessment areas: urban areas in eight selected Units of Analysis (UoA) for IPC food security analysis. For the IPC, the population is stratified by population group: host community or IDP. From this stratification, two distinct groups will be covered in this assessment: urban host communities and urban IDPs. The sampling is tailored to reflect the unique context and needs of each group.

This assessment uses a structured and statistically robust sampling methodology, aligned with the IPC framework. Given the diverse contexts and operational realities on the ground, the sampling design adopts a two-stage cluster sampling approach based on probability techniques, ensuring that findings are representative at the UoA level.

The sampling methodology relies on advanced tools such as the HQ sampling platform and geospatial population data provided by REACH. This ensures rigor in both sampling and field-level implementation. Further details are below.

Urban Host Community Household Sampling

Urban host areas are initially segmented into hexagonal clusters of 1 km². These clusters are overlaid with [population density data](#) from [WorldPop](#) to identify those that meet specific criteria: a minimum density of 300 persons per km² and a total urban population of at least 5,000.⁴ Only clusters that satisfy both conditions are retained for the sampling. To avoid overlapping with IDP data, clusters in which more than 30% of the population consists of IDPs in CCCM-listed IDP data are excluded.

Once cleaned, the remaining clusters undergo probability proportional to size (PPS) sampling via the HQ sampling multi-cluster tool; this is the first stage, ensuring that more populous areas have a greater likelihood of selection.⁵ After selection, any known IDP sites within the clusters are manually mapped

⁴ This definition of urban combines 'towns' (moderate density urban clusters) plus 'cities' (high density urban centres) according to the [Global Human Settlement Layer](#) (GHSL) framework, an initiative of the European Commission, FAO, UN-Habitat, ILO, OECD, and The World Bank to provide a harmonised global definition of the degree of urbanisation, endorsed by the [UN Statistical Commission](#) in 2020. For the methodological manual, see [here](#). While IPC does not use a fixed definition of urban for food security analysis ([IPC](#); see also [WFP](#)), the selected definition follows [past assessments](#) that have fed into IPC analyses in Somalia.

⁵ Each eligible hexagon's household total (HH_total) serves as the size measure for PPS selection. Hexagons with larger populations have a higher probability of being selected, which ensures the sample is representative of the population distribution. The HQ sampling multi-cluster tool takes the cleaned hexagon list with HH_total, hex_ID, cluster_id, and geographic coordinates as inputs, applies the PPS algorithm, and outputs the final selected clusters. This approach ensures that larger settlements are proportionally represented in the sample while still providing geographic coverage across the study area.

and excluded. In the second stage, eight households per selected cluster are randomly chosen for interviews using simple random technique.

Urban IDP Household Sampling

For urban IDPs, each verified IDP site from the CCCM dataset is treated as a discrete sampling cluster. CCCM geopoints that fall inside of hexagonal clusters with urban characteristics (defined above) are retained for the sampling frame. PPS sampling is applied at the site level, using the estimated number of households as a basis for selection. Within each sampled IDP site, at least eight households are selected at random using the HQ multi cluster tool. This ensures robust representation of the urban displaced population living in IDP sites.

Table 3. Sampling frame by UoA

	UoA	Total pop HH ⁶	# clusters	# clusters sampled	HH sampled size	UoA target population
1	Bay – Urban Host Community	60173	70	11	128	Urban Host Community
2	Bay – Urban IDPs	34842	43	12	120	Urban IDPs
3	Gedo – Urban Host Community	58981	73	12	120	Urban Host Community
4	Gedo – Urban IDPs	73086	48	10	136	Urban IDPs
5	Bakool – Xudur Urban Host Community	18831	17	10	136	Urban Host Community
6	Bakool – Xudur Urban IDPs	25204	33	11	120	Urban IDPs
7	Bakool – Urban Host Community (without Xudur)	27348	14	9	144	Urban Host Community
8	Bakool – Urban IDPs (without Xudur)	26008	14	8	184	Urban IDPs
	Total	324,473	312	83	1,088	

⁶ This refers to the total number of households living in the eligible clusters for a given Unit of Analysis (under ‘# clusters’ column), based on available population data from WorldPop.

GPS-Based Point Generation

The physical location of each interview is determined through GPS-based random point selection. These points are generated within the boundaries of the selected clusters or sites and are weighted by population density to ensure equitable geographic coverage. Enumerators navigate to the GPS coordinates using mobile applications such as Maps.me and conduct interviews with the nearest household. This methodology ensures spatial randomness and minimises selection bias.

To successfully implement this approach, the following are required:

- Accurate and current shapefiles for administrative boundaries.
- Reliable population distribution and density data.
- Well-trained data collection teams proficient in using digital navigation tools.

Key requirements for effective sampling:

1. **Detailed Area Mapping:** Relying on the updated CCCM Site Master list, the GIS team will generate a map of the accessible IDP sites in Somalia. This map will clearly mark the boundaries of IDP sites and districts, including any major landmarks near or within the targeted sites. This detailed understanding of the layout is crucial for defining a sampling frame and navigating the area efficiently.
2. **Divide the targeted areas into camp sites:** The targeted areas for urban IDPs will then be divided into sites to implement systematic sampling. If the target site is not of a manageable size, the site will be further divided into sub-sites by direction.

Addressing Probable Implementation Challenges:

1. **Absent Households:** If a selected household is found to be vacant during the initial visit, the neighbouring household will be surveyed.
2. **Unforeseen Obstacles:** In situations where the predetermined route is obstructed due to unforeseen circumstances (e.g., construction work), the field team will document the obstacle and deviate slightly to maintain the systematic selection by following the next available pathway within the target site.

The field team has experience implementing second-stage sampling in previous rounds of assessment, and all field officers and enumerators will be re-trained on the methodology to ensure there are no major deviations in the sampling protocol. In cases of required deviations due to unforeseen circumstances, changes will be documented in the cleaning log with proper justifications.

Using replacement list of PSU – To account for unexpected challenges that might prevent data collection from specific locations, the following re-sampling steps will be implemented. Before data collection begins, the same number of backup samples will be generated for each of the two population groups within each Unit of Analysis. This ensures that there are back-up samples available in case any settlements become inaccessible during data collection. On the day of data collection, if a settlement cannot be reached, the team will directly switch to collecting data from its designated backup location. To maintain transparency, a clear log will be kept by the SDO (Senior Data Officer) documenting which districts and settlements were replaced using the backup samples.

To streamline the re-sampling process, field officers will be asked to check the accessibility of target location on the Thursday before or Sunday of the working week, coordinating with the field manager. By proactively evaluating accessibility beforehand, we aim to pre-emptively address any potential obstacles and minimise delays that could arise during the re-sampling process.

Data collection – Following consultations with the IPC core group, data collection should be finalised in early August to be ready for the IPC analysis in mid-to-late August. REACH will streamline the assessment tool to meet IPC-specific requirements.

Planned data collection: From 21 July 2026 to 01 August 2026.

The assessment will be implemented through in-person interviews in targeted areas. The field implementation of the assessment will be supervised by REACH Field Officers (FOs), who report to the Field Manager (FM). The FOs and FM will receive assessment-specific training during 12-13 July 2026, covering the review of the tool (questionnaire, including technical definitions and concepts) and data cleaning.

Following the training of trainers, FOs will then be redeployed to their bases to provide training to enumerators at the regional level. As much as possible, REACH will hire enumerators (1) with prior experience with MSNA or other assessments conducted by REACH, and (2) locally, to ensure that we reduce operational challenges related to accessing hard-to-reach locations.

Two full days of training for enumerators will aim to present, explain, and test the assessment tool before data collection. Following the training, a pilot data collection will be conducted. This will allow the examination of the enumerators' understanding of the content of the questions and functionalities of the tool, through feedback sessions with the FOs and analysis of the data; the tool may be adjusted based on feedback (e.g., in case logical problems are identified).

Following this, data collection will commence. Enumerators will use the Maps.me app with pre-downloaded offline maps and pre-loaded household location data (CSV or KML) to navigate to assigned households via GPS coordinates. However, throughout data collection, REACH FOs will review the targeted locations within each UoA. In case of an inaccessible location, the FOs will immediately communicate this to the data officers to re-run the sampling for that district. Households will be randomly selected according to the final sampling framework, with the questionnaire being administered either to the head of the household or another adult able to communicate on behalf of the household.

The questionnaire will be coded on the Kobo tool and be accessible to all enumerators on REACH data collection smartphones. Enumerators will begin the interview by introducing themselves and requesting the respondents' informed consent to proceed. As FOs will not be able to accompany all enumerator teams on the ground, they will identify team leaders in each team to supervise the progress of data collection and ensure regular communication with their respective FO. Each day enumerators, under the supervision of their team leader and/or of the FO, will upload the survey forms and debrief the team leader and/or the FO of any issues encountered during data collection. The FOs, Data Officer, and Assessment Officers (AOs) are responsible for data checking and cleaning procedures at the end of each day and for communicating feedback to the enumerators and team leaders.

Incoming data will be monitored through a data collection tracking tool. Each day REACH FOs will communicate with the data team the total number of surveys completed per UoA. The coverage shall be monitored based on the targeted number of surveys per population group. In case one location becomes inaccessible, the REACH data team shall rerun the UoA sampling for the affected UoA only and replace the sample for the affected UoA.

3.5 Data Processing

At the end of each day, the team leaders will oversee the uploading of collected data from enumerators' smartphones to REACH's KoBo Collect server. The REACH data team will then retrieve all datasets for spatial verification, a crucial step involving cross-examination of GPS coordinates of the completed surveys, to ensure they fall within a predetermined radius of the target settlement. Any surveys failing to meet this criterion, in addition to the surveys completed in a short period of time, will be flagged for further attention by field officers.

Upon completion of the preliminary verification, the data team will remove all personally identifiable information (PII) and disaggregate datasets by districts. Any outliers in the data, including translation of any required 'other' responses from Somali to English is then reviewed and cross-checked by the field officers. Once cleaned, the data is checked by the assessment team for a secondary review based on logic checks, during which any additional recommendations or follow-ups are communicated to the field officers and enumerators for their daily briefings. Additionally, as part of the quality assurance process, all changes and cleaning made to the submitted surveys are tracked in a comprehensive data cleaning log.

To standardise this process, **Standard Operating Procedures (SOP) for data cleaning** are used: a step-by-step guide for key data cleaning issues, including checking the time stamp of each survey, issues with skip logic and outliers. The SOP will be developed based on the MSNA household survey tool and REACH's [Data Cleaning Minimum Standards Checklist](#). See Annex 3.

3.6 Analysis Plan

Data Disaggregation and Stratification:

Following a two-stage cluster sampling design, the analysis will disaggregate the cleaned dataset by key strata per UoA for urban host communities and urban IDPs. This stratification ensures a granular understanding of the data by considering factors like economic activity, particularly for UoA (aligned with Livelihood Zones). In addition, the analysis will maintain the disaggregation of the population groups both across the entire dataset and within each UoA, to ensure consistency and comparability across all strata.

Weighting and Aggregation:

Weighting was applied at the Unit of Analysis (UoA) level for each population group to ensure statistically valid inferences and correct potential sampling biases. Stratification was based on livelihood zone–population group combinations. Weights were calculated at the stratum level and aggregated across national, regional, district, and population group levels.

During data cleaning, all inaccessible clusters were excluded from the sampling frame, which included only accessible areas. If clusters became inaccessible during data collection, they were replaced with nearby clusters of similar population size and characteristics to minimise bias. These adjustments were incorporated into the final weights, applied after data collection based on the final sample size and adjusted sampling frame.

Further Analysis Considerations:

If the data quality allows, the analysis will be further disaggregated by demographic characteristics such as Age, Gender, and Disability (AGD). This will provide a more nuanced understanding of how the data varies across different demographic sub-groups. However, since the samples were not specifically generated considering the AGD component, any analysis based on these demographics will be indicative of household's situation.

Finally, the analysis results will be presented in a clear and organised manner, separated by sector. Each sector's findings will be further segregated based on the previously mentioned strata (UoA, population groups).

4. Key ethical considerations and related risks

For detailed guidance on how to complete this section, see also Step 5 of the [IMPACT Research Design Guidelines](#)

The proposed research design meets / does not meet the following criteria:

<i>The proposed research design...</i>	<i>Yes/ No</i>	<i>Details if no (including mitigation)</i>
... Has been coordinated with relevant stakeholders to avoid unnecessary duplication of data collection efforts?	Yes	
... Respects respondents, their rights and dignity (<i>specifically by: seeking informed consent, designing length of survey/ discussion while being considerate of participants' time, ensuring accurate reporting of information provided</i>)?	Yes	
... Does not expose data collectors to any risks as a direct result of participation in data collection?	Yes	
... Does not expose respondents / their communities to any risks as a direct result of participation in data collection?	Yes	
... Does not involve collecting information on specific topics which may be stressful and/ or re-traumatising for research participants (both respondents and data collectors)?	No	While the assessment does not collect highly sensitive personal information, some questions relating to food insecurity, livelihood loss, displacement, shocks, or household vulnerability may cause discomfort or emotional distress for some respondents. To mitigate this risk, the questionnaire uses standard indicators that have been developed and validated through established technical processes and have been widely used

		in humanitarian assessments. Enumerators will receive training on informed consent procedures, sensitive interviewing techniques, and the identification and management of signs of distress. Participation will be voluntary, and respondents will be informed that they may decline to answer any question or terminate the interview at any time without consequence. If a respondent becomes distressed, enumerators will discontinue the relevant line of questioning or end the interview if requested. Feedback and accountability mechanisms will also be made available to participants. Enumerators will also have access to supervisory support throughout data collection and will be encouraged to raise any concerns arising from repeated exposure to potentially distressing accounts.
... Does not involve data collection with minors i.e. anyone less than 18 years old?	Yes	
... Does not involve data collection with other vulnerable groups e.g. persons with disabilities, victims/ survivors of protection incidents, etc.?	Yes	Enumerator training prioritises ethical data collection (non-intrusive questions, sensitivity awareness and option to not respond to the question/not participate in the survey) to protect unidentified vulnerable groups.
... Follows IMPACT SOPs for management of personally identifiable information ?	Yes	As per the SOPs, all personally identifiable information will be deleted and is only used to contextualise analysis.

5. Roles and responsibilities

Table 3: Description of roles and responsibilities

Task Description	Responsible	Accountable	Consulted	Informed
<i>Research design</i>	Senior Assessment Officer (SAO), Owen Boed	SAO	Research Manager (RM) Angie Martin , Field Manager (FM) Hajir Hussien , Field Officers (FOs), Data and GIS Unit Sulaiman Anwari and	Country Coordinator (CC), Deputy Country Coordination (DCC), IPC,

			Charity Nduta , IMPACT HQ Focal Point, Cluster Coordinators Working Group Coordinator, IPC TWG	Clusters, WGs, Field Staff
Supervising data collection	FO	FM	SAO	CC, DCC
Data processing (checking, cleaning)	FOs, SDO, AO, SAO	SAO	FMs, IMPACT HQ Focal Point, TWG	CC, DCC
Data analysis	AO, SAO	SAO	, IMPACT HQ MSNA Focal Point, TWG	IPC
Output production	AO, SAO, SDO	SAO	, IMPACT HQ Focal Point, Cluster & WG Coordinators, IPC TWG	CC, DCC
Dissemination	AO, SAO	SAO	CC Ana Garay , FMs, FOs, IMPACT HQ Focal Point, TWG	IPC
Monitoring & Evaluation	AO, SAO	SAO	CC, DCC, FMs, FOs, SDO, IMPACT HQ MEAL Unit, IMPACT HQ Focal Point, TWG	IPC, ACTED PD, IMPACT HQ RDD
Lessons learned	AO, SAO, SDO, FMs, FOs	SAO	CC, DCC, RM, IMPACT HQ M&E Team, IMPACT HQ Focal Point	IPC, ACTED PD, IMPACT RDD

Responsible: the person(s) who executes the task

Accountable: the person who validates the completion of the task and is accountable of the final output or milestone

Consulted: the person(s) who must be consulted when the task is implemented

Informed: the person(s) who need to be informed when the task is completed

6. Data Analysis Plan

The DAP was developed in collaboration with global and in-country clusters/working groups. Questions were selected from the MSNA indicator bank that could generate data on Acute Food Insecurity (AFI) outcomes and contributing factors and that address the information needs of IPC AFI analysis. Indicators were shared with IPC TWG partners and clusters for feedback; over several rounds of IPC assessment, the tool has been refined and optimised to contribute the most useful information for IPC AFI analysis. The data analysis plan (DAP) for the assessment will be published online. Findings will be disaggregated by population group and geographic area where appropriate.

7. Data Management Plan

Administrative Data	
Research Cycle name	Food Security Among Urban Populations
Project Code	27BMF
Donor	SIDA
Project partners	Agency for Technical Cooperation & Development - ACTED
Research Contact	Angie Martin – RM (angie.martin@impact-initiatives.org) Owen Boed – SAO (owen.boed@impact-initiatives.org)

Data Management Plan Version	Date: 21/06/2026	Version: 01
Related Policies	IMPACT Data Protection Policy	
Documentation and Metadata		
What documentation and metadata will accompany the data? <i>Select all that apply</i>	☒ Data analysis plan	☒ Data Cleaning Log, including: X Deletion Log X Value Change Log
	☒ Code book	☐ Data Dictionary
	☒ Metadata based on HDX Standards	☐ [Other, Specify]
Ethics and Legal Compliance		
Which ethical and legal measures will be taken?	☒ Consent of participants to participate	☐ Consent of participants to share personal information with other agencies
	☐ No collection of personally identifiable data will take place	☒ Gender, child protection and other protection issues are taken into account
	☒ All participants reached age of majority	☐ [Other, Specify]
Who will own the copyright and Intellectual Property Rights for the data that is collected?	REACH	
Storage and Backup		
Where will data be stored and backed up during the research?	☒ IMPACT/REACH Kobo Server	☐ Other Kobo Server: <i>[specify]</i>
	☒ IMPACT Global Physical / Cloud Server	☒ Country/Internal Server
	☒ On devices held by REACH staff	☐ Physical location <i>[specifies]</i>
	☐ [Other, Specify]	
Which data access and security measures have been taken?	☒ Password protection on devices/servers	☒ Data access is limited to REACH staff members.
	☒ Form and data encryption on data collection server	☐ Partners signed an MoU if accessing raw data
	☒ Anonymised dataset with no PII shared with partners	
Kobo Access Rights		
Kobo Access	Person	Account Name
View Form	Enumerators	enumsom_cash
View and Edit Form	Sulaiman Anwari, Senior Data Officer	Sulaiman_anwari
View Form and Submit Data	Sulaiman Anwari, Senior Data Officer	Sulaiman_anwari, enumsom_cash
Download Data	Sulaiman Anwari, Senior Data Officer	Sulaiman_anwari,
Raw Data Access Rights		
Raw Data Access	Reason	Person

Accountable	Accountable	Sulaiman Anwari, Senior Data Officer			
Access	Spatial verification of collected data	Charity Nduta, GIS Officer			
Preservation					
Where will data be stored for long-term preservation?	X	IMPACT / REACH Global Cloud / Physical Server	X	OCHA HDX	
	X	REACH Country Server	☐	[Other, Specify]	
Data Sharing					
Will the data be shared publicly?	X	Yes	☐	No, only with mandating agency / body	
Will all data be shared?	☐	Yes	X	No, only anonymised and cleaned data will be shared	
	☐	No, [Other, Specify]			
Where will you share the data?	X	REACH Resource Centre	X	OCHA HDX	
	☐	Humanitarian Response	X	IPC GSU to upload onto the Analysis Platform for IPC analysis	
Data protection risk assessment					
Have you completed the Indicators Risk Assessment table below?	X	Yes	☐	No, no information that potentially allows identification of individuals is to be collected.	
	[Please complete the first 4 columns in the Indicators Risk Assessment table below]				
Risk indicator	Type of identification risk	Disclosure implications	Benefits	Class	Required mitigation
Respondent contact details (name and phone number)	Direct identification of the interviewee	Potential threat if interviewee's details are disclosed especially those in conflict areas or members of minorities	Data checks and cleaning	B1	To be permanently deleted after data collection
Respondent's location (region, district and settlement name)	Direct identification of the interviewee	Potential threat of being targeted by armed actors	Geographical disaggregation of results	B2	To be permanently deleted after data analysis and visualisation
Responsibilities					
Data collection	Sulaiman Anwari, SDO – sulaiman.anwari@impact-initiatives.org				
Data cleaning	Sulaiman Anwari, SDO – sulaiman.anwari@impact-initiatives.org				
Data analysis	Sulaiman Anwari, SDO – sulaiman.anwari@impact-initiatives.org				
Data sharing/uploading	Sulaiman Anwari, SDO – sulaiman.anwari@impact-initiatives.org				

8. Monitoring & Evaluation Plan

IMPACT Objective	External M&E Indicator	Internal M&E Indicator	Focal point	Tool	Will indicator be tracked?
Humanitarian stakeholders are accessing IMPACT products	Number of humanitarian organisations accessing IMPACT services/products Number of individuals accessing IMPACT services/products	# of downloads of x product from Resource Center	Country request to HQ	User_log	X Yes
		# of downloads of x product from Relief Web	Country request to HQ		☐ Yes
		# of downloads of x product from Country level platforms	Country team		☐ Yes
		# of page clicks on x product from REACH global newsletter	Country request to HQ		☐ Yes
		# of page clicks on x product from country newsletter, sendingBlue, bit.ly	Country team		☐ Yes
		# of visits to x webmap/x dashboard	Country request to HQ		☐ Yes
IMPACT activities contribute to better program implementation and coordination of the humanitarian response	Number of humanitarian organisations utilising IMPACT services/products	# references in HPC documents (HNO, SRP, Flash appeals, Cluster/sector strategies)	Country team	Reference_log	<i>IPC, all clusters' strategies (Protection, Shelter and Non-Food Items, Health, Education, WASH, Nutrition, AAP, CCCM).</i>
		# references in single agency documents			

Humanitarian stakeholders are using IMPACT products	Humanitarian actors use IMPACT evidence/products as a basis for decision making, aid planning and delivery	Perceived relevance of IMPACT country-programs	Country team	Usage_Feedback and Usage_Survey template	<i>Opened survey for REACH Somalia, throughout the year. The link is sent to partners when sharing any REACH product.</i> <i>REACH team will share the survey link after the release of MSNA factsheet, to at least 20 partners (clusters, IPC, data collection partners, etc.)</i> <i>Logging of any form of feedback in the internal M&E framework (e-mails, comments during meetings, etc.)</i>
		Perceived usefulness and influence of IMPACT outputs			
		Recommendations to strengthen IMPACT programs			
		Perceived capacity of IMPACT staff			
		Perceived quality of outputs/programs			
	Number of humanitarian documents (HNO, HRP, cluster/agency strategic plans, etc.) directly informed by IMPACT products	Recommendations to strengthen IMPACT programs			
Humanitarian stakeholders are engaged in IMPACT programs throughout the research cycle	Number and/or percentage of humanitarian organisations directly contributing to IMPACT programmes (<i>providing resources, participating to presentations, etc.</i>)	# of organisations providing resources (i.e. staff, vehicles, meeting space, budget, etc.) for activity implementation	Country team	Engagement_log	X Yes
		# of organisations/clusters inputting in research design and joint analysis			X Yes
		# of organisations/clusters attending briefings on findings;			X Yes

ANNEX 1: MODIFICATION TO THE CORE INDICATOR (IF RELEVANT)

Module	Indicator	Question	Please explain what modifications were made?	Justification for the change?	Change made in consultation with IMPACT HQ? If yes, who was consulted?
Household Food Consumption	HDDS	Did your household eat [food group] yesterday during the day and night at home?	Added 'at home' to the question.	IPC request	Reported to HQ
Household Food Consumption	FCS	How many days over the last 7 days, did most members of your household (50% +) eat, inside the home, [food group]?	Added 'inside the home' to the question.	IPC request	Reported to HQ
Respondent and head of household	% of respondents who are the head of household	Are you the head of household, the main person who makes decisions on behalf of the entire household?	Added second phrase after comma.	To ensure respondents understand what is being asked.	Reported to HQ
AAP	% of households having reported priority needs by most commonly reported preferred modalities of assistance	What type of humanitarian assistance would best support your household, as a first priority?	To the response option 'Cash - Cash that can be used for any purpose, without restriction', the assessment team added '(this includes physical cash, mobile money, bank transfers, and cash via prepaid cards)'.	To adapt the response option for the Somali context.	Reported to HQ

ANNEX 2: SECONDARY DATA MATRIX

Type of source	Topics	Purpose	Full reference (+ link if possible)	Comments - key findings/information gaps
Humanitarian	2026 HNRP	The Humanitarian Needs and Response Plan	OCHA – HNRP Accessible here	Somalia's 2026 humanitarian crisis remains severe, with 4.8 million people requiring assistance driven by a fourth consecutive failed rainy season. This is coupled by ongoing conflict and forced evictions that have displaced over 3.3 million people. The response plan for 2026 targets 2.4 million people across 64 districts, within a total requirement of \$852 million (a 40% reduction from 2025), prioritising 1.6 million in the 21 most critical districts. This reflects shrinking resources rather than reduced needs. Shelter gaps remain acute with nearly 200,000 IDPs at risk from land disputes and evictions while the withdrawal of AUSSOM forces and UNSOS logistical cuts further threaten humanitarian access in key areas like Bay, Bakool, and Hirshabelle. The plan highlights the severely underfunded response with only 50% of people in need being targeted, which represents the lowest coverage in several years.
Humanitarian	IPC Results	IPC findings on Acute Food Insecurity	IPC - January 2026 report Accessible here	According to the IPC Acute Food Insecurity and Acute Malnutrition Analysis (April–June 2025), nearly 4.6 million people (representing 24% of Somalia's

			IPC - April 2026 report Accessible here Population tracking tool Accessible here	population) are projected to face high levels of acute food insecurity (IPC Phase 3 or above). Among this, 784,000 people in Emergency (IPC Phase 4) and 3.8 million in Crisis (IPC Phase 3). This is a significant deterioration from the January 2025 projection of 4.4 million. The crisis is being driven by below-average Gu season rainfall, high food prices, ongoing conflict and insecurity, and localised flooding along the Juba and Shabelle Rivers. The humanitarian funding reduction directly affected food, nutrition, health, and WASH service delivery across the country. On the nutrition front, the total burden of acutely malnourished children aged 6–59 months has risen to 1.8 million. This also includes 479,000 children facing Severe Acute Malnutrition (SAM). Therefore, reflecting the life-threatening consequences of funding cuts on Somalia's most vulnerable children.
Humanitarian	Internally Displaced Persons (IDPs)	Inform assessment methodology, including sampling design	CCCM Cluster Somalia HDX Data Shared bilaterally	There are 4,119 recorded IDP sites across Somalia as of January 2026. Approximately, 85% of the sites are informal settlements on private land and 90% of site are in urban areas according to the Detailed Site Assessment (DSA) of the CCCM Cluster.
Humanitarian	Livelihood Zones	Inform assessment methodology, including sampling design	FSNAU – Somalia: Livelihood zones Accessible here . Somalia Livelihood Zones (Shape files)	The map disaggregates Somalia into 19 livelihood zones with clear distinctions of the admin layers.

			Accessible here	
Humanitarian	Internally Displaced Persons (IDPs)	Definition of IDPs used in the assessment	OHCHR – Guiding Principles on Internal Displacement Accessible here .	Provides the definition, resolutions, and international standards mandated by the Special Rapporteur on the human rights of internally displaced Persons
IOM	Internally Displaced Persons (IDPs)	Displacement tracking in Somalia	IOM DTM Somalia reporting Accessible here	IOM's information management system to track and monitor numbers, locations, movements, needs and vulnerabilities of displaced people. Outputs support responses in humanitarian crisis, evidence-based decision making, durable solutions programming and migration research.
Governmental	Internally Displaced Persons (IDPs)	Provides a contextual definition and understanding of IDPs in Somalia	Federal Government of Somalia – National Policy Accessible here .	Provides a clear contextual definition of IDPs in the context of Somalia. The policy also provides the context and background of displacement and plans for durable solutions for the population group – including refugees and returnee populations
Academic – Research Institute	Presence of NSAGs	Understanding of context and access to regions	CrisisWatch – conflict tracker and early warning tool Accessible here	Al-Shabaab and IS-Somalia frontlines remained relatively quiet, though joint federal, Jubaland, and U.S.-trained Danab forces continued a slow advance toward Jamaame in Lower Juba supported by U.S. airstrikes. Meanwhile, the humanitarian situation is rapidly deteriorating as worsening drought conditions following poor rains across both the north and south.
Academic – Research Institute	Presence of NSAGs	Understanding of context and access to regions	ACLED – Conflict events data Accessible here	Armed clashes in Mogadishu have increased political tensions over Somalia's proposed electoral reforms. The dispute between the

				federal government, opposition groups, and some regional states is increasingly taking on clan dimensions, raising the risk of broader instability and potentially undermining efforts to combat Al-Shabaab.
Humanitarian	Overall Humanitarian Access in 2026	Provides details on the overall 2026 humanitarian landscape of Somalia	Somalia: Humanitarian Access Snapshot (01 April to 30 June 2026) Accessible here Humanitarian Access Snapshot Accessible here	In Q1 2026, Somalia recorded 53 humanitarian access incidents, above the 2025 average, mainly driven by conflict, inter-clan violence, political tensions, and military operations. Insecurity, attacks on aid personnel, administrative obstacles, drought, and regional instability continued to hinder humanitarian operations.
Humanitarian	Overall Humanitarian Access in 2026	Provides details on the overall access severity 2025.	Somalia Access Working Group: Severity of Access Overview (As of 14 January 2025) Accessible here	As of January 2025, humanitarian access across Somalia varied considerably, with the most severe restrictions concentrated in parts of South West, Hirshabelle, Galmudug, and Jubaland. These areas faced significant operational challenges due to insecurity, conflict, and administrative barriers, limiting the delivery of humanitarian assistance. In contrast, most areas in Somaliland and Puntland experienced relatively fewer access constraints, allowing humanitarian actors to operate with greater ease.

Humanitarian	Update on the rainfall and flooding	FAO's monitoring tool for climatic conditions	FAO - SWALIM maps NDVI and Rainfall Maps Accessible here	The CDI monitoring tool explores the use of three climatic conditions that influences drought conditions. This includes rainfall, temperature, and the Normalized Vegetation Drought Index (NVDI) which is proxy of soil moisture.
Humanitarian	Forecast for 2026 Gu "long rains" season	Predictions of the Gu Rains in Somalia and its possible impact on accessibility of sites for data collection and exacerbated needs due to flooding and potential displacements	FAO – Somalia Climate Outlook for the 2026 Gu Season Accessible here	Somalia is expected to face continued drought stress during the Gu 2026 season following failed rains in 2025, with severe water shortages, livestock losses, and food insecurity persisting. Although near-normal rainfall is the most likely scenario, recovery is expected to be slow and uncertain due to high temperatures, delayed rainfall, and the risk of both drought and localized flooding.
Humanitarian	Situation Update	Update on the floods	OCHA – Situation Report Somalia: 2025 Drought Emergency - Situation Report No. 1 (As of 27 November 2025) Accessible here Somalia: 2025 Drought Emergency - Situation Report No. 2 (As of 21 December 2025) Accessible here	Somalia's drought emergency continued to worsen in early 2026, with severe water shortages, livestock losses, crop failure, displacement, and rising food insecurity affecting millions of people. Although forecasted rains may provide some relief, recovery is expected to be slow, while limited humanitarian funding and response capacity continue to constrain assistance to the most vulnerable populations.

			Somalia: 2025 Drought Emergency - Situation Report No. 3 Accessible here Somalia: 2025-2026 Drought Emergency - Situation Report No. 4 Accessible here Somalia: 2025-2026 Drought Emergency - Situation Report No. 5 (As of 6 April 2026) Accessible here	
--	--	--	---	--

ANNEX 3: DATA CLEANING STANDARD OPERATING PROCEDURES

2026 Somalia Multi-Sectorial Needs Assessment (MSNA)

Data Cleaning Standard Operating Procedures

Introduction

Throughout data collection, cleaning will regularly take place to maintain the high standard of the assessment. Data cleaning and verification will take place daily. Feedback will be provided directly to REACH Field Officers (FOs) by the Senior Database Officer. Data cleaning will focus on identifying outliers in the data, contradictory or unlikely response options (logical inconsistencies), and suspicious patterns from enumerators. A cleaning log of all changes will be kept and will be available upon request after the REACH publication of datasets. It is crucial for the MSNA team to have supporting tools that can help us monitor the quality of data we are receiving from the field. This document will outline the proposed data workflow and responsibilities of each team member for the assessment.

After downloading the data following steps will be implemented:

1. Run the R script developed to implement the required checks on the dataset. Outputs of the R scripts are as follows:

1. **Cleaning log** file that will outline all issues that need to be addressed.

The uuid, question name, description of issue, feedback and old value will be included in the cleaning log file. Field officers will follow up on the flagged issues by double-checking it with respective enumerators and team leaders. Once checked and confirmed, field officers will fill the new value column and inform the data team. Once the new value is cross-checked and confirmed by the data team, the old value can be replaced with the new value and clean data can be generated.

2. **Raw data** specific for each location will be exported daily and saved in the shared folder. Each field officer is expected to check the feedback on their bases and complete the cleaning log. All data cleaning files should be shared with the GIS/Data team in a timely manner, so the team can access and maintain consolidated clean data.

Daily Data Checks & Cleaning

Daily Responsibilities

Senior/ Assessment Officer

1. Responsible for reviewing cleaning results daily and providing feedback to either Senior Database Officer/Field coordinator/ Field Officers.
2. In coordination with Field coordinator and field officers, communicate regarding security or logistical concerns that change sampling framework.
3. Responsible for ensuring each field team is on track to complete required target samples with quality surveys.

Senior Data / Database Officer

1. The Senior Data Officers will be responsible for downloading, deleting, and anonymizing data from kobo daily.
2. Runs daily R data checking script with clean data, and raw data from most recent day to identify errors for Field Officers to follow up on.
3. Runs the data monitoring dashboard scripts daily, to make sure that the information in the tracker is the most updated and accurate.
4. makes final call on survey deletions.
5. Responsible for overseeing changes to sampling framework and adjusting sampling targets accordingly.
6. Responsible for ensuring daily backups of the cleaned data and that proper file naming protocol is followed for cleaned data and cleaning log.
7. Communicate all data issues, feedback, and any other data related issues to the responsible S/FO, who contacts field teams and individual enumerators to clarify any issues with the data.

Senior/Field Officers

1. In constant communication between the data officers and individual enumerators regarding issues with data collection and data quality issues.
2. Daily checking information updated in the tracker and debriefed enumerators on any update to ensure data quality is maintained at high standards and that the target is achieved in their locations.
3. Make sure that each enumerator has uploaded their survey to the KoBo server before 5PM, or as soon as it is possible daily.
4. Ensure phones are fully charged prior to next day of data collection,

5. Ensure phones are set to the correct time and date prior to data collection. Achieved with steps below: Settings > General management > Date and time > Automatic date and time AND Use 24-hour format ON At the end of each data collection day. Once submitted, data will be downloaded by the Senior Data Officer daily by 7 PM at the latest, removing personally identifiable data and adding unique IDs; the data team will run the R scripts that will automatically spot errors – see table 1 below.
6. Each field officer will then receive the raw data and the cleaning log file with the identified issues on their respective data, upload to the respective folders in SharePoint. The enumerators are expected to check and correct all uploaded data-related errors, under the supervision of the team leader and field officer; in addition, field teams will check for vertical and horizontal errors in the data, following up with the respondent, if needed. Once the field teams have made the changes, all files should be uploaded to the Sharepoint.

Data Checking Best Practices for the Field Teams

1. When you apply a filter **REMEMBER TO CLEAR IT!** Otherwise, you will go ahead not looking at ALL the questionnaire!
2. **Fill all** the columns under new.value column, in the cleaning log, even in cases where the new value does not change from the old value.
3. **Know the tool:** only if you are fully familiar with the tool and respective skip logics will you be able to clean the dataset properly and catch logic mistakes.
4. **Spot-check enumerators** and ensure you understand the structure, logic and how questions are understood by both enumerators and respondents.
5. **Know the local context:** please use your understanding of the local context to help you catch mistakes, such as knowing that in this area of data collection it's not possible that somebody is using a flush toilet.
6. **Whilst reviewing and cleaning the data**, both a **horizontal and vertical logic should be applied**.
7. **Horizontal logic (scroll to the right):** check whether reported responses of each indicator of each survey are logically consistent and make sense in relation to each other.
8. **Vertical logic (scroll down):** with the data sorted *by enumerator*, scroll dataset downwards to check whether there are any suspicious response patterns for specific enumerators that suggest the enumerator is performing poorly or misunderstood a question or response option. Check the overall distribution of responses, that way you can find out which enumerator's surveys seem suspicious.
9. During data cleaning you **should** have **pen and paper to take some notes** for the next day's **morning briefing**. Alternatively, you can also just **use the cleaning log as briefing notes for the enumerator feedback**.
10. If providing individual feedback to one enumerator, sort your cleaning log by "enumerator."

11. If providing feedback on a common issue appearing across enumerators, sort the cleaning log by "issue". Keep the teams motivated!
12. Do not only flag the issues, ensure you explain why something would not make sense and articulate in which areas they have improved.

Data Checks

This below table will guide you through the checks in the RScript and what should be taken for each one. In short, data checking columns mainly flag potential issues within the data, since it will be too difficult to ensure that everything aligns perfectly and logged among the different people involved in the process. This process will help us to approach issues with the same action so it will be easy for us to collaborate and speed up the process.

Cell #	Issue	Action
	Survey time taken	Sort the time taken from Lowest to Highest or A to Z to check which surveys are filled in a short time and which surveys are filled in more extended periods. Minimum survey time = 25 minutes Maximum survey time = 90minutes
	consent	Check if there are any declined consents in the data so you can decide to take additional surveys based on your sample and delete the no consent surveys
	hh_size < 2 / hh > 14	Please check if the total household size is less than 2 or greater than 14 and cross-check any figure that needs to be verified or double-checked with enumerators or respondents.
	uuid	Check for duplicated surveys in the data.
	ind_age! = ind_dob_approx	Check if the age completed in years entered by the respondent is the same with age in months calculated from selecting the DOB in the tool.
	time_productive	Check for outliers using the inter-quantile range calculations.

	time_unproductive	Check for outliers using the inter-quantile range calculations.
	times_displaced	Check for outliers using the inter-quantile range calculations.
	salary	outliers using the inter-quantile range calculations.
	casual	outliers using the inter-quantile range calculations.
	business	outliers using the inter-quantile range calculations.
	own_product	outliers using the inter-quantile range calculations.
	govt_benefits	outliers using the inter-quantile range calculations.
	rent	outliers using the inter-quantile range calculations.
	remittance	outliers using the inter-quantile range calculations.
	family_friends	outliers using the inter-quantile range calculations.
	charity_donations	outliers using the inter-quantile range calculations.
	hum_assistance	outliers using the inter-quantile range calculations.
	food	outliers using the inter-quantile range calculations.
	rent_expenditure	outliers using the inter-quantile range calculations.
	water_allsources	outliers using the inter-quantile range calculations.
	non_food_items	outliers using the inter-quantile range calculations.

	utilities_allconnections	outliers using the inter-quantile range calculations.
	fuel_allvehicles	outliers using the inter-quantile range calculations.
	transportation_allvehicle	outliers using the inter-quantile range calculations.
	communications_allcosts	outliers using the inter-quantile range calculations.
	other_expenditure	outliers using the inter-quantile range calculations.
	health_related	outliers using the inter-quantile range calculations.
	non_fooditems	outliers using the inter-quantile range calculations.
	shelter	outliers using the inter-quantile range calculations.
	education_related	outliers using the inter-quantile range calculations.
	debt_repayment	outliers using the inter-quantile range calculations.
	all_other	outliers using the inter-quantile range calculations.
	Xx_other	All the other checks that may need translations are listed in the left column. Please check the other options provided and recode the values that can be found from the lists and translate the others. If translation, please do translate to English. If the value looks invalid, ask the enumerators for clarification.

	age_decision_maker ==resp_age, and age_decision_maker ==resp_age	Logical checks where the respondent mentions not to the decision maker, but the age and gender of the decision maker matches that one for the respondent.
	Return to area of origin (for IDP HHs only) == HOst Community	Check for Host communities that have mentioned to return to their area of origin
	salary, casual, business, own_product, govt_benefits, rent, remmitance, family_friends, charity_donations, hum_assistance, domestic_expenditure, food, rent_expenditure, water_allsources,	Check for 999 or 999 from all the listed indicators, it could mean the respondent is not sure about the response or the correct value. If confirmed, the specific entries will be removed

	non_food_items, utilities_allconnections, fuel_allvehicles, transportation_allvehicle, communications_allcosts, other_expenditure, expenditure_domestic, shelter, non_fooditems, health_related, education_related, debt_repayment, all_other	
	gps 1 == gps 2	Check if the gps recorded at the start of the survey is the same with the one that is recorded at the end of the survey.
	total_non_residing_children	Check for 999 entries; it must be removed later.
	Infant formula [siha, nunalac, aptamil, nido] (times)	Check for 999 entries; it must be removed later.

	Milk (tinned, powdered, or fresh animal milk) (times)	Check for 999 entries; it must be removed later.
	Sour milk or yoghurt [i.e. caana fadhi] (times)	Check for 999 entries; it must be removed later.
	resp_hoh_yn == "yes" & resp_age < 20	respondant is HoH but younger than 20
	fsl_fcs_score < 10	FCS is too low, and the values will be cleared
	fsl_fcs_score > 70	FCS is too high in a humanitarian situation
	fsl_fcs_score == 0 & cm_expenditure_frequent_food > 10	respondent mentioned to have spent more than 10 USD on food this month on their income but they have consumed some food groups (FCS above 0), please confirm where they got the food
	fsl_fcs_score >= 28 & fsl_fcs_cereal < 5	The FCS score is higher than or equal to 28 while fsl_fcs_cereal element is less than 5
	fsl_fcs_score <= 28 & cm_expenditure_frequent_food > 200 & hh_size < 7	The FCS score is less than or equal to 28 and hh size is less than 7, while food_exp is more than 200 USD
	wash_drinking_water_source %in% c('tank_truck', 'cart_tank', 'bottled_water', 'sachet_water') & cm_expenditure_frequent_water < 5	The wash_drinking_water_source is selected 'bottled_water' OR 'tanker' OR 'cart_with_small_tank' but cm_expenditure_frequent_water is less than 5 USD

	ind_age_schooling_n == 0 & cm_expenditure_infrequent_education > 0	The ind_schooling_n == 0 while cm_expenditure_infrequent_education reported > 0
	fsl_fcs_cereal < 4	Cereals, grains, rice, maize, sorghum, and tubers is less than 4. Please correct your answer.
	fsl_fcs_meat >= 5 fsl_fcs_oil < 4	Meat or poultry consumption is more than 4, please cross-check Oil consumption is less than 4, please cross-check
	fsl_hdds_cereals == "yes" & fsl_fcs_cereal == 0	cereal was consumed within last 24 hrs but not reported in fcs
	fsl_hdds_dairy == "yes" & fsl_fcs_dairy == 0	diary was consumed within last 24 hrs but not reported in fcs
	fsl_hdds_fruit == "yes" & fsl_fcs_fruit == 0	fruit was consumed within last 24 hrs but not reported in fcs
	fsl_hdds_meat == "yes" & fsl_fcs_meat == 0	meat was consumed within last 24 hrs but not reported in fcs
	fsl_hdds_eggs == "yes" & fsl_fcs_meat == 0	egg was consumed within last 24 hrs but not reported in fcs
	fsl_hdds_fish == "yes" & fsl_fcs_meat == 0	fish was consumed within last 24 hrs but not reported in fcs
	fsl_hdds_legumes == "yes" & fsl_fcs_legumes == 0	legumes was consumed within last 24 hrs but not reported in fcs
	fsl_hdds_oil == "yes" & fsl_fcs_oil == 0	oil was consumed within last 24 hrs but not reported in fcs

	fsl_hdds_sugar == "yes" & fsl_fcs_sugar == 0	sugar was consumed within last 24 hrs but not reported in fcs
	fsl_hdds_condiments == "yes" & fsl_fcs_condiments == 0	condiments was consumed within last 24 hrs but not reported in fcs
	ind_aduld_above_18 == hh_size	fsl_rcsi_mealadult turn to zero
	wash_sanitation_facility_sharing_n > 15	Sanitation facility shared by >15 HHs

Data collection tracking

MSNA survey tracker dashboard is built to provide a live snapshot of the data collection against the targeted sample at the region, district, cluster and strata status level. The dashboard is linked to the kobo, and it provides the numbers of submissions in the server in a timely manner.

The tracker consists of 6 main coverage statistics.

1. Number of interviews received per pop group (urban host community and urban IDP) per region
2. Number of interviews received per pop group (urban host community and urban IDP) per district
3. Number of interviews received per pop group (urban host community and urban IDP) per strata.
4. Number of deleted interviews per district and region and per enumerators.
5. Overall number of surveys collected per district against the sample size.

The purpose of enumerator tracker would be to ensure enumerators' productivity. The Data Officer will generate this tracker based on number of surveys per enumerator per day, average number of data cleaning issues per enumerator, enumerators following suspiciously similar survey paths. This will be closely monitored and tracked. If the trend of the enumerator is consistently poor, putting the data quality at risk, an update will be provided daily. Each enumerator is expected to complete no more than 6 surveys per day.

Spatial Verification

Responsible(s): GIS/ team

The master settlement list will be used as the base map for checking data collection. The collected data from the field will be overlaid over the base map and the distance between the base map settlements point and the collected points will be calculated. The points which are far from the target settlement's locations will be flagged for action in the cleaning log.

Neither Accurate nor Precise:

This means points collected are captured far apart from each other and out of the targeted location. That's why it's neither precise nor accurate.

Accurate, not Precise:

This means points are accurate and fitting one location/polygon, but the precision is missing.

Precise, not Accurate:

This means the points are precise but according to the point that we're referencing it's not accurate. So, either the site has been relocated or the sites has the same name.

Accurate and Precise:

This means the points are accurate and precise on the targeted location.

No GPS Coordinates:

This means there are no GPS coordinates captured from the field. We must discourage this to happen in the field and make mandatory for each enumerator to take the GPS coordinates after each survey.

Re-sampling SOP

- **FOs log their issues in a daily location issues tracker.** Important to record are:
 - **District**
 - **Livelihood zone**

- **Location name** (settlement name / IDP site name)
- **Target** (number of HH surveys to conduct in that location)
- **Issue** (only 3 possibilities allowed: ***Inaccessible, Not found / unknown, Located in another district***)
- **Name of FO** logging the issue

It is KEY to stick to this format, to avoid messiness and a heavy workload. A location being “too far” is NOT a valid issue.

- **The designed data person checks this document daily to provide feedback to FOs:** he/she checks if an issue is valid and provides feedback in a “feedback column” to the FO. Feedback can be: “**Invalid – go to location anyway**” or “**Valid – will be resampled**”.
- **For the “valid issues”, designated data person resamples from the ‘improved population data sources’.** This should also happen on a daily / need’s basis. Resampling needs to happen taking the final targets into account. Resampled locations (with all their information included: district, name, coordinates, target number, etc.) can be entered in a 2nd tab of the location issues tracker called “**Resampled locations**”, but to be sure, should also be emailed to the relevant FO. You can also make a column “Acknowledged by FO” in this spreadsheet, which the relevant FO must mark (X) when seen & understood.
- Data team should **keep a record of how the sampling frames are evolving in real time:** what locations are going to be visited in the end, with how many interviews as a target. We have three separate sampling frames here: non-displaced, protracted IDPs, new IDPs. This is also important for the special verification checks during data cleaning: we need to know what **the eventual, real sampling frame is that got implemented!** For this, make a copy of the initial/original sampling frames, and create ‘**dynamic versions**’, where you mark **rows that were deleted (inaccessible / unknown / wrong district) in RED** and **newly added rows (that got resampled) in GREEN**.

RACI-structure of tasks:

Task description	Responsible	Accountable	Consulted	Informed
DAILY: FOs enter location issues in daily location issues tracker.	FOs	SDO (SAO if SDO is on leave)	Field Coordinator, GISM, SAO	RM, AO, SAO

DAILY: Check location issues tracker & provide feedback to FOs.	SDO (SAO if SDO on leave)	SDO & SAO	GISO, SAO, RM	RM, AO, SAO
Weekly/as required: Resampling.	SDO (SAO if SDO on leave)	SDO & SAO	GISO, SAO, RM	RM, AO, SAO
DAILY: Update dynamic sampling frames.	SDO (SAO if SDO on leave)	SDO & SAO	GISO, SAO, RM	RM, AO, SAO