Research Terms of Reference

Assessment of Hard-to-Reach (H2R) Areas in Northwest Nigeria

Research Cycle ID: NGA2202

Nigeria

September 2022

REACH Informing more effective humanitarian action

1. Executive Summary

Country of intervention	Nigeria						
Type of Emergency	□	Natural disaster	X	Conflict			
Type of Crisis	□	Sudden onset	□	Slow onset	X	Protracted	
Mandating Body/ Agency	European Civil Protection and Humanitarian Aid Operations (ECHO)						
Project Code	*35ANW*						
Overall Research Timeframe *(from research design to final outputs / M&E)*	01 /10/2022 to 30/04/2023						
Research Timeframe *Quarterly outputs (factsheets)*	1. Start collect data: start of the 1st October 2022		5. Preliminary presentation: N/A				
	2. Data collected: throughout until 31st March 2023		6. Outputs sent for validation: end of 3rd week of 4th month				
	3. Data analysed: end of 1st week of January and April 2023		7. Outputs published: end of 4th week of January and April 2023				
	4. Data sent for validation: beginning of 2nd week of January and April 2023		8. Final presentation: N/A				
Research Timeframe *Quarterly outputs (situation overviews)*	1. Start collect data: start of the 1st month		5. Preliminary presentation: 1st week of January and April 2023				
	2. Data collected: end of the 3rd month		6. Outputs sent for validation: 1st week of February and May 2023				
	3. Data analysed: end of 1st week of January and April 2023		7. Outputs published: 2nd week of February and May 2023				
	4. Data sent for validation: beginning of 2nd week of January and April 2023		8. Final presentation: 3rd week of May 2023				
Number of assessments	□	Single assessment (one cycle)					
	X	Multi assessment (more than one cycle) *Continuous data collection with quarterly output*					
Humanitarian milestones *Specify **what** will the assessment inform and **when*** *e.g. The shelter cluster will use this data to draft its Revised Flash Appeal;*	**Milestone**		**Deadline**				
	X	Donor plan/strategy	ECHO strategy, end of the year				
	X	Inter-cluster plan/strategy Northwest Coordination Forum (NWCF)	Quarterly findings to be presented at the NWCF.				
	□	Cluster plan/strategy	_ _/_ _/_ _ _ _				
	□	NGO platform plan/strategy	regular presentations to INGO Forum				
	X	CH Analysis Workshop	REACH through participation in bi-annual CH analysis workshops and FSL WASH/Nutrition/health analysis working group will provide timely updates and				

			participation in analysis to identify areas with highest levels / highest risk levels of severe/extreme food insecurity	
Audience Type & Dissemination *Specify **who** will the assessment inform and **how** you will disseminate to inform the audience*	**Audience type** X Strategic X Programmatic X Operational □ [Other, Specify]		**Dissemination** X General Product Mailing (e.g. REACH Nigeria mailing list, which includes sectoral and inter-sectorial coordination mail to NGO consortium; donors) X Cluster Mailing X Presentation of findings at Cluster meetings X Website Dissemination (Relief Web, REACH Resource Centre, & HDX) X Following the 3th month of data collection, quarterly sectoral factsheets and data sets reporting on conditions at the LGA level	
Detailed dissemination plan required	□	Yes	X	No
General Objective	To assist humanitarian actors in making more informed decisions about the scale, scope, and locations of response, through providing detailed information and analysis on humanitarian needs, displacement dynamics, and service access in hard-to-reach areas in Northwest Nigeria.			
Specific Objective(s)	1. To identify the cross-sectoral needs and vulnerabilities of populations in hard-to-reach areas, whether they are internally displaced persons (IDPs), returnees or host community members. 2. To provide up-to-date information on service provision and access in hard-to-reach areas. 3. To map main displacement patterns to and from hard-to-reach areas.			
Research Questions	• What are the needs and vulnerabilities of IDP, returnee and host community populations in hard-to-reach areas with regards to Food Security and Livelihoods (FSL), Health, Nutrition, Shelter & Non-Food Items (NFIs), WASH, Education and Protection and how do these change over time?[1] • Which services and types of humanitarian assistance do IDP, returnee and host community populations in hard-to-reach areas have access to and what access constraints exist? • What are the key demographic characteristics (e.g. type of populations present such as IDPs, returnees, etc) of populations living in H2R areas/settlements? • What are the key displacement trends in Northwest Nigeria (push and pull factors for new arrivals, month, and area of origin of IDP arrivals, intentions to move, etc.)?			
Geographic Coverage	H2R settlements in Local Government Areas (LGAs) in Northwest Nigeria, with a focus on Katsina, Sokoto and Zamfara. The pilot will take place in Katsina State[2].			
Secondary data sources	• DTM IOM reports, UNHCR, WFP essential needs analysis – Northwest (Zamfara, Katsina and Sokoto) Nigeria • ACLEDCrisisGroup • Sector-led assessments as available in reference to H2R and surrounding areas (FS, WASH, Education, Health, Nutrition, etc.)			

[1] Longitudinal analysis is only possible and will only be reported upon if the geographic coverage remains the consistent over time, i.e. data is being compared from the same H2R wards month to month.

[2] The pilot will take place in Batsari, Faskari, Kankara, Sabuwa and Safana LGAs of Katsina State, with a view to extending to other LGAs within Katsina, Sokoto and Zamfara in the future.

<table>
<tr><td></td><td colspan="6">• Partner-led assessments as available in reference to H2R and surrounding areas (Search for Common Ground, Solidarités International, Save the Children, UNICEF, Plan, International Crisis Group, MSF etc)
• Academic papers
• Online media sources</td></tr>
<tr><td rowspan="5">Population(s)
Select all that apply</td><td>□</td><td>IDPs in camp</td><td>X</td><td>IDPs in informal sites</td><td colspan="2"></td></tr>
<tr><td>X</td><td>IDPs in host communities</td><td>□</td><td>IDPs [Other, Specify]</td><td colspan="2"></td></tr>
<tr><td>□</td><td>Refugees in camp</td><td>□</td><td>Refugees in informal sites</td><td colspan="2"></td></tr>
<tr><td>□</td><td>Refugees in host communities</td><td>□</td><td>Refugees [Other, Specify]</td><td colspan="2"></td></tr>
<tr><td>X</td><td>Host communities</td><td>X</td><td>Returnees</td><td colspan="2"></td></tr>
<tr><td>Stratification
Select type(s) and enter number of strata</td><td>X</td><td>Geographical # Hard to Reach LGAs across Katsina State
Population size per strata is known. □ Yes X No
Threshold for reporting at LGA level is at 5% settlement coverage</td><td>□</td><td>Group #: _ _ _
Population size per strata is known?
□ Yes □ No</td><td>□</td><td>[Other Specify] #: _ _
Population size per strata is known?
□ Yes □ No</td></tr>
<tr><td>Data collection tool(s)</td><td>X</td><td>Structured (Quantitative)</td><td>X</td><td>Semi-structured (Qualitative)</td><td colspan="2"></td></tr>
<tr><td></td><td colspan="3">Sampling method</td><td colspan="3">Data collection method</td></tr>
<tr><td>Structured data collection tool # 1
Select sampling and data collection method and specify target # interviews</td><td colspan="3">X Purposive</td><td colspan="3">X Key informant interview (KII) (Target #): Participant are new arrivals or have spoken to someone living in the hard-to-reach location from 0- 3 months. Varies by LGA, aiming to cover at least 5% of settlements per month Given more than one quantitative survey may be collected on a given settlement, data from KIs reporting on the same settlement is aggregated to the settlement level. To present information at the LGA level, data from settlements is then aggregated to the LGA level.</td></tr>
<tr><td>Semi-structured data collection tool (s) # 1
Multi-Sectoral Service Access Gap FGD</td><td colspan="3">X Purposive</td><td colspan="3">X Focus group discussions (Target #): minimum of 3 FGD per LGA per month, to be led by saturation.

Participants newly arrive internally displaced persons (IDPs) who have left a hard-to-reach settlement in the last three months. Number of participants is between 5-8 and are usually separated by gender and age group.</td></tr>
</table>

Semi-structured data collection tool (s) # 2 *Multi-Sectoral Service Access Gap IDI*	X Purposive			In-Depth Interviews (Target #). A minimum of 4 IDIs per LGA per month. IDIs are conducted when FGDs are not feasible due to the accessibility of the location. IDIs are one on one phone interviews conducted with IDPs who newly arrive from hard-to-reach settlements in the last three months.		
Semi-structured data collection tool (s) # 3 Participatory Mapping FGD	X Purposive			X Focus group discussion (Target #): minimum of 3 FGD per LGA per month, to be led by saturation. Participants newly arrive internally displaced persons (IDPs) who have left a hard-to-reach settlement in the last three months.		
Target level of precision if probability sampling	N/A			N/A		
Data management platform(s)	X	IMPACT		□		
	□	[Other, Specify]				
Expected output type(s)	X	Situation overview #: One (1) quarterly	□	Report #: _ _	□	Profile #: _ _
	□	Presentation (Preliminary findings) #: One (1) quarterly	□	Presentation (Final) #: _ _	X	Factsheet #: Quarterly output, 6 sectors
	□	Interactive dashboard #:_	□	Webmap #: _ _	X	Map #: as needed
	□	[Other, Specify] #: _ _				
Access	X	Public (available on REACH resource centre and other humanitarian platforms)				
	X	Restricted (bilateral dissemination only upon agreed dissemination list, no publication on REACH or other platforms) Sensitive data will be shared via Northwest coordination group and INGO Forum on a need-to-know basis; all other data is publicly available.				
Visibility *Specify which **logos** should be on outputs*	All products should be REACH branded, with visible ECHO logos included, where other partners support or fund data collection, their logos should be included, or references provided on each document.					

2. Rationale

2.1. Background

Over the last decade, the seven states that comprise the Northwest region of Nigeria – Jigawa, Kaduna, Kano, Katsina, Kebbi, Sokoto, and Zamfara – have experienced deadly inter-communal conflict and organized crime, often referred to as banditry, resulting in the deaths of thousands and the displacement of hundreds of thousands of people across the region. The crisis is underpinned by inter-communal and farmer-herder violence, further exacerbated by the detrimental effects of climate change and desertification on agriculture and livelihoods, as well as high rates of illiteracy and the highest poverty rates in the country. Attacks, including shootings, cattle rustling, kidnapping, and looting, have resulted in increasing causalities in the region; between January and July 2022, there

were 2,229 reported fatalities in the Northwest[3], nearly reaching the total reported fatalities that occurred in the entire year in 2021 in the Northwest.[4]

With the rise of criminality, kidnapping and banditry in the Northwest region, which has rendered some locations inaccessible for service-provision and livelihood activities, REACH Nigeria is planning an assessment using the Area of Knowledge (AoK) methodology in Katsina State, to monitor the humanitarian needs and displacement of populations living in hard-to-reach areas. The AoK approach provides regular indicative tracking of humanitarian needs over time, to support in the identification and prioritisation of "hot-spot" areas experiencing alarming levels of humanitarian needs. To ensure this data directly informs humanitarian response planning, the findings will be shared through the Northwest Coordination Forum, and feedback will be sought from partners to inform research design and geographical targeting, as well as to triangulate the data and analysis.

2.2. Intended impact

The assessment will provide the humanitarian organisations with more in-depth information on the basic needs and availability of services in hard-to-reach areas in Northwest Nigeria. Although as of yet there is no humanitarian coordination system or humanitarian project cycle (HPC) framework in the Northwest in the way it exists in the Northeast, the assessment will feed information Northwest Coordination Forum, chaired by the Nigeria INGO Forum (NIF) and UNICEF, for better planning and coordination. This research in the Northwest will improve the understanding of the current situation and feed into a much-needed evidence base to advocate for more resources to address humanitarian needs in the region.

3. Methodology

3.1. Methodology overview

In recognition of the urgent need for data to improve the understanding of the humanitarian context in Northwest Nigeria, this assessment focuses on IDP, returnee, and host community populations in hard-to-reach or inaccessible areas. For the purpose of this assessment, hard-to-reach areas are defined as those areas that are not regularly accessible, either due to active conflict or due to humanitarian access restrictions, or a combination of these.

Key informant interviews (KIIs) will be conducted in accessible locations, with key informants (KIs) who are either IDPs arriving from the H2R areas within the last three months or those who have had contact with someone living in the H2R area in the last three months. Similarly, FGDs on service access and participatory mapping will be conducted in accessible locations, with KIs who are IDPs arriving from the H2R areas within the last three month. Quantitative data will be aggregated to the settlement level and analysed at the LGA level, and findings will only be reported if at least 5% of settlements in an LGA were assessed. FGD data will be analysed to provide further explanation of the results from the quantitative data. The data from the KIIs and FGDs is indicative of broad trends only and is not statistically generalizable.

Should the team face accessibility issues in Northwest, REACH Nigeria will transition to remote data collection for both qualitative and quantitative. Were FGDs will be supplemented with remote in-depth interviews (IDIs) with key informants who arrived from an H2R area within the course of the previous month. Local stakeholders and partners will help in identifying KIs for staff. Therefore, prior to remote data collection, a comprehensive stakeholder engagement exercise will be conducted, and contacts of relevant local stakeholders and guides as well as partners will be noted down for each garrison town/IDP camp. The KIs will be identified and recruited through engagement with local stakeholders and local guides (i.e. people working closely with community leaders in informal IDP sites) and shall serve as phone connectors for the Enumerator/FOs. During remote data collection, the designated enumerator/FO will conduct interviews with KIs from the target LGAs via phone, through the local guide/phone connector, who will be responsible for connecting the enumerator/FO and KI(s).

3.2. Population of interest

In recognition of the lack of information on populations remaining in H2R areas, this assessment focuses on IDP and host community populations in H2R or inaccessible areas. Through engagement with the government, as well as through internal security and access assessments, REACH identified numerous LGAs across Katsina State that are considered hard-to-reach, of which Batsari, Faskari, Kankara, Sabuwa, and Safana LGAs were considered particularly inaccessible, and were thus selected for the pilot. Data will be

[3] International Centre for Investigative Reporting, "Insecurity: 7,222 killed, 3,823 abducted in seven months – Report," ICIR (2 August 2022)
[4] ACLED, "10 Conflicts to Worry About in 2022: Nigeria," ACLED (2022).

collected at the lowest possible administrative unit – individual settlements – as derived from the most recent version of the Grid3 dataset (Nigerian Government Project, The Geo-Referenced Infrastructure and Demographic Data for Development) released in February 2021 (grid3.gov.ng). The level of coverage (proportion of settlements assessed in a given LGA) will be declared for each product when results are presented. LGA-level reporting in which less than 5% of settlements have been assessed will not be included in published REACH products.

3.3. Secondary data review.

Secondary data will be used throughout all stages of the research cycle to identify locations that are considered particularly inaccessible or where information gaps are greatest, as well as to support in the design of tools and to triangulate findings.

- Selecting geographic coverage: Areas about which very little information or secondary data is available will be targeted for H2R data collection, especially when this is a result of limited humanitarian access.
- Triangulation of H2R data in analysis and product drafting:
    - IOM Weekly Flash Reports on NW Displacement: IOM DTM releases weekly flash reports that provide figures on displacement in the NW, which were also used to inform geographic coverage.
    - Mobile Network Coverage Mapping: Mobile network coverage maps can help determine which areas are potentially cut off from phone coverage and communication lines, and are also potentially more likely to be considered inaccessible.
    - Any key sectorial reports on WASH, Health, Food Security, Livelihood and Cash Early Recovery.
    - IPC data and analysis, and health structural/programme data available.

3.4. Primary Data Collection

Quantitative data collection tools

Quantitative data will be collected pending access, the data collection teams may expand to other accessible areas in the future. Data will be collected through a structured multi-sector survey tool that captures settlement-level information on displacement, FSL, WASH, Shelter, Nutrition, Protection, Education, and Health. The same tool can be used in both in-person and remote data collection. In all REACH AoK field locations, the same multi-sector tool and questions are used and the tool was designed in collaboration with input from cluster partners. Feedback is encouraged from partners after the first quarter presentation of the report.

The tool is available in English and Hausa with translations provided by Translators Without Borders. Data is collected on mobile phones through Kobo Toolbox. At the end of each data collection day, the forms will be uploaded to the Kobo server, after which the datasets are merged, cleaned and uploaded to the REACH/IMPACT server on a daily basis. Data is collected by enumerators on a daily basis for the period of one month and who will be supervised by Field Officers (FOs) who in turn are managed by an Operations Manager and Assessment Officer (AO). Before the start of data collection, the FO has already identified the Kis that will be interviewed by enumerators a day before. In the morning before the team deploys to the field, enumerators are briefed by FO based on the targeted KIs and the settlement they're from and the data cleaning process of the previous day (described in the next section) to ensure appropriate coverage (due to the nature of the security context, it is difficult to implement a random sampling for AoK), debrief on any potential data collection errors and to achieve the highest quality in data collection. Analysis and report drafts will be done quarterly by the AO.

Qualitative data collection tools

In order to provide a more in-depth understanding of the dynamic context as well as to complement data collected through the quantitative tool, the monthly data collection cycle also includes FGDs. FGDs are conducted by Field Officers, with Enumerators translating when necessary. FGDs are separated by gender and age when there are enough participants.

- *Displacement mapping tool:* A participatory mapping tool for use within FGDs to capture displacement flows and push and pull factors to secure locations and from H2R areas. This is conducted each month to provide an analysis of displacement trends per geographical area. FGD participants are purposively sampled new arrivals (IDPs who arrived within the last three months).
- *Service access gap tool*: The service access gap FGD tool is used to capture the level of access to sectoral services (FS, ERL, Health, Nutrition, Shelter & NFI, WASH, Education and Protection) and service access constraints. In the

majority of cases this tool is used to look at wider settlement-level service access trends, with participants purposively selected according to their origin in a specific ward. Participants are purposively sampled new arrivals (IDPs who arrived within the last three months).

If accessibility becomes an issue, FGDs will be replace with in-depth interviews (IDIs) which will be complemented by the quantitative data. The population of interest for this assessment includes KIs who have arrived from a H2R settlement within the last three months and KIs who have direct knowledge of the H2R settlement. This is to ensure that KIs provide timely and detailed information.

The KIs will be identified and recruited through engagement with local stakeholders and local guides (i.e. people working closely with community leaders in the IDP camps). Prior to the remote data collection, a comprehensive stakeholder engagement exercise is conducted, and contacts of relevant local stakeholder and guides are noted down for each garrison town/IDP camp.

When recruiting KIs for IDIs, the FOs will reach out to the identified stakeholders/guides who will identify the KIs in their respective locations. After the local stakeholder/guide has identified a KI, they will call the designated FO who will conduct the IDI with the KI by using the stakeholder/guide's phone in the preferred language of the KI (English or Hausa). The FO will seek the consent of the KI to record the call, and if the KI consents, the FO will put his/her phone on speaker throughout the interview and record the call using a digital recorder. The audio file is then transferred to the Senior Field Officer (SFO) alongside a consent form. The FO immediately deletes the audio file after transcription from both the recorder and laptop.

After each IDI, the FO will plan for the next IDI in collaboration with the local stakeholder/guide, under the supervision of the SFO.

During the IDIs, the FOs will be based in either the office in Katsina or Yola, or other appropriate locations in relevant garrison towns. The KIs will be located in a safe and quiet place identified by the local stakeholder/guide. When possible, this location should be the locations normally used for FGDs. If that place is not available, the local stakeholder/guide will identify another quiet and private place.

The data collection tool will be semi-structured and be available in English and Hausa. Based on bi-weekly feedback meetings with FOs and the assessment officers (AO), minor changes might be added to the tool to adapt to the context and understanding of the questions.

Due to this sampling approach and the hard-to-reach methodology in general, data is only indicative and not representative.

Where possible, only respondents that have arrived accessible garrison towns very recently (0-3 months prior to data collection) will be interviewed.

KIs report on the settlement level. A minimum of one (1) key informant interview (KII) per settlement is required, and teams will seek to avoid more than 5 KIIs per settlement in order to avoid inefficiencies.

For focus group discussions (FGDs) and in-depth interviews (IDIs), a purposive approach will also be employed. Field Officers/local guides will seek out people with knowledge of the ward in question.

3.5. Data Processing & Analysis

Quantitative data
Every day, at the end of data collection, the surveys are uploaded on the REACH/IMPACT Kobo-server and downloaded in csv format as one dataset for a specific site. This dataset is cleaned during the evening, logging deleted entries and value changes, whilst the raw dataset is also stored. Given more than one quantitative survey may be collected on a given settlement, data from key informants reporting on the same settlement is aggregated to the settlement level using an R script. To reconcile divergent responses when responses are aggregated to the level of the settlement, the most common response provided by the greatest number of KIs is reported for that settlement. Questions in which the KIs from the same hard-to-reach settlement did not provide a most-common, or consensus response, are reported as "no consensus***".*** Data is analysed by the Assessment Officer feeding the R script output on Excel and/or Tableau.

Aggregation of KI responses to settlement level:

During the aggregation stage, each settlement is assigned one value for each question. If there is only one KI for a settlement, their answers automatically become the value for the settlement. If there is more than one KI, the value of the settlement is assigned accordingly:

- **Single response questions:** The majority of survey questions only allow a KI to select a single response. For most of these questions data presented at the settlement level is the most prevalent (modal) response of KIs survey responses from each settlement. In the event there is not a clear majority, for example if 3 KIs report "yes" and 1 reports "no", the settlement is assigned as "yes". However, if 2 KIs report "yes" and 2 KIs report "no", the settlement response is coded as No consensus (NC).
- For certain questions, noted in the analysis plan, the modal response is not considered and presence of some responses will super-cede others, this is to ensure that one KI's lack of knowledge about specific issues, for instance, protection concerns, do not cancel out the information that other KIs might know. In this category we have the following sets that prioritize certain responses over others:
    - "Yes" responses over others
    - "No" responses over other
    - Highest Response over all other responses
    - For example, if there are three KIs, and one notes that there are unaccompanied children in the settlement, but two say that there are not, the answer will be coded as "Yes" even though more KIs reported "No" because not all KIs might be aware of unaccompanied children within the settlement.
- **Multiple response questions:** Some questions present the KI with multiple responses from which to choose, such as the food regularly eaten by most people in settlement. We disaggregate each possible response option as a single-response and code the logic according to single-response questions. For these set of questions all the reported possible response options are reflected while aggregating the settlement data.

All questions will be analysed according to the % of assessed settlements responding for each answer. NC will be included as an option in the total responses for the aggregated data. In the analysis, it will not count towards the existing options, as will 'don't know' or 'don't want to answer'. For instance, for the question "In the last three months, were MOST people able to access enough food in [info settlement]?" The final possible values are 'yes', 'no', 'no consensus', 'don't know', or 'don't want to answer'. When producing a map of adequate access to food, the analysis will only count 'yes' answers as 'adequate access'. Correspondingly, when producing a map of inadequate access to food, the analysis will only count 'no' answers as inadequate access.

Aggregation of settlement responses to LGA level:

Data from settlements is aggregated to the LGA level using a R script which employs the following logic to calculate county-level responses:

- **Single response questions:** The majority of survey questions only allow a KI to select a single response. Data presented at the LGA level is the most prevalent response across all assessed settlements within the LGA.
- **Multiple response questions:** Some questions present the KI with multiple responses from which to choose, such as the main source of food in the settlement. For each possible response, REACH presents the percent of settlements in each LGA reporting that selected the response.
- **Yes/no questions:** Yes/no questions are presented as the percent of settlements in the LGA with a response of "yes".

For data in a LGA to be considered as having met the reporting threshold, at least 5% of settlements in each LGA must be assessed. If not, then the settlement data can be used for larger, aggregated outcomes, but cannot be used to make broader claims of the LGA as a whole.

Qualitative data

The FGD/IDI notes are transcribed by FO and sent to the SFO who store the documents in a clear labelled folder. A saturation grid and analysis is created to summarize the transcription, providing a clear overview of key discussion points per discussion topic and a better understanding FGDs/IDIs conducted, identify trends and themes at the LGA level. The data saturation grid and analysis is reviewed in-country and subsequently send to HQ for review and validation to ensure transparency, neutrality and consistencies. As both KI and FGD/IDI participants may have left the settlement at any point in the month prior to data collection, the findings may be indicative of the situation in the H2R area at different periods of time.

3.6 Reporting

Reporting will be done on a quarterly basis. It is expected that there will be two reporting cycles for this assessment. After analysis, the AO will be drafting quarterly sectoral factsheets, Situation Overviews (SO) and a preliminary presentation. The sectoral factsheet will be five, reporting on Food Security and Livelihood, Health and WASH, Protection, Population movement and communication, and Education and Shelter while the SO will be a more comprehensive in-depth report on the findings for the quarter. The report will be disseminated with the Northwest Coordination Forum, Partners, Sector Leads, State Emergency Management Agency (SEMA), and relevant stakeholders, as well as published on the REACH resource centre.

4. Roles and responsibilities

Table 2: Description of roles and responsibilities

Task Description	*Responsible*	*Accountable*	*Consulted*	*Informed*
Research design	Senior Assessment Officer (SAO)/Assessment Officer (AO)	(S)AO	Research Manager (RM), GIS Team, HQ Research Unit, Country Representative (CR)	Sectors, UN OCHA, relevant partners
Supervising data collection	Field Manager (FM)	(S)AO	RM, Operations Manager (OM)	Clusters, UN OCHA, relevant partners
Data processing (checking, cleaning)	Data Officer	Senior Data Officer	(S)AO, GIS Team	(S)AO, FM
Data analysis	(S)AO	(S)AO	GIS Team, HQ Research Unit	RM, CR
Mapping	GIS Team	GIS Team lead	(S)AO, HQ Research Unit	RM, CR
Output production	(S)AO	(S)AO	RM, GIS Team, HQ Research Unit	CR
Factsheets	(S)AO	(S)AO	RM, HQ Research Unit,GIS Team lead	CR
Situation Overview	(S)AO	SAO/AO	RM, CR, HQ Research Unit	CR
Dissemination	(S)AO	(S)AO	RM, CR	CR
Monitoring & Evaluation	(S)AO	(S)AO	RM, HQ Research Unit	CR
Lessons learned	(S)AO	(S)AO	RM, HQ Research Unit	CR

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*

5. Data Analysis Plan

RESEARCH QUESTIONS ADDRESSED WITH SEMI-STRUCTURED TOOL(S)

The Tool is in 2 parts:

- Service access
- Participatory mapping.

RESEARCH QUESTIONS ADDRESSED WITH STRUCTURED TOOL(S)

This tool is subject to review on a quarterly basis and may undergo slight alterations over time based on internal review and partner feedback.

6. 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		X 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		X Yes
		# of page clicks on x product from country newsletter, sendingBlue, bit.ly	Country team		X Yes
		# of visits to x webmap/x dashboard	Country request to HQ		X Yes
IMPACT activities contribute to better program implementation and coordination of the humanitarian response	Number of humanitarian organisations utilizing IMPACT services/products	# references in HPC documents (HNO, SRP, Flash appeals, Cluster/sector strategies)	Country team	Reference_log	Humanitarian Needs Overview Humanitarian Response Plan Sector Response Strategies UNOCHA Country Strategy
		# 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 Number of humanitarian documents (HNO, HRP, cluster/agency strategic plans, etc.) directly informed by IMPACT products	Perceived relevance of IMPACT country-programs	Country team	Usage_Feedback *and* Usage_Survey template	Survey monkey: As part of regular dissemination email, survey monkey sent every six months to assess usage of REACH products. Qualitative feedback: Each REACH staff responsible for reporting back to communications manager each time agency requests REACH information or provides feedback on how REACH information has been used
		Perceived usefulness and influence of IMPACT outputs			
		Recommendations to strengthen IMPACT programs			
		Perceived capacity of IMPACT staff			
		Perceived quality of outputs/programs			

		Recommendations to strengthen IMPACT programs			
Humanitarian stakeholders are engaged in IMPACT programs throughout the research cycle	Number and/or percentage of humanitarian organizations directly contributing to IMPACT programs *(providing resources, participating to presentations, etc.)*	# 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: REACH ACRONYMS

REACH Internal team:

- AO – Assessment Officer
- CC – Country Coordinator
- DBS – Database Specialist
- FA – Field Assistant
- FO – Field Officer
- FM – Field Manager
- GVA – Geneva (HQ)
- GISO – GIS Officer
- GISS – Geographical Information Systems (GIS) Specialist
- RM- Research Manager
- SAO – Senior Assessment Officer
- SFO – Senior Field Officer
- SGISO – Senior GIS Officer

Research terms:

- FGD – Focus group discussion
- FS – Factsheet
- IDI – In-depth interview
- H2R – Hard to Reach
- NW -- NorthWest
- KI(I) – Key informant (interview)
- NC – No consensus
- SDR – Secondary data review
- SO – Situation overview

Clusters/coordination bodies:

- AWG – Access Working Group
- CCCM – Camp Coordination and Camp Management
- FS – Food security
- ERL – Early recovery and livelihoods
- IMWG – Information Management Working Group
- ISCG –Inter-Sector Coordination Group
- WASH – Water, sanitation, and hygiene

Partners:

- ICRC – International Committee of the Red Cross
- IOM DTM – International Organization for Migration (IOM) Displacement Tracking Matrix
- IOM ETT – IOM Emergency Tracking Tool
- MSF – Médecins Sans Frontières
- OHCT – Operational Humanitarian Country Team
- UNOCHA – United Nations Office for the Coordination of Humanitarian Affairs
- WFP – World Food Programme

Data:

- ACLED – Armed Conflict Location and Event Data (project)
- INSO – The International NGO Safety Organisation

Other:

- LGA – Local Government Area
- IDP – Internally displaced person(s)

ANNEX 2: STEPS TO DAILY DATA CLEANING PROCESS

1. Daily data quality checks (HFCs)

a. DO downloads raw data from Kobo, cleans column headers, and stores one excel sheet with daily raw data and amends raw data master file for data collection round, using R
b. DO manually downloads all audit files from Kobo and stores them in the R "input" folder
c. DO checks logical consistencies of daily submissions and produces list for each enumerator of all logical issues detected and all "other" responses to be checked for possible recoding, using R. This list will be used as the cleaning log once filled out by the enumerators. The logical checks used by R for evaluating the surveys are contained in a separate excel sheet in the "input" folder of the R project. Any changes in the checks need to be done in that excel sheet.
d. DO sends lists with detected issues and "other" responses ("verification reports") via email to each enumerator, with FOs, FM, AO, and DO in CC, via R
e. Enumerator checks each detected issue and "other" response in the verification report and provide feedback in the "enumerator_feedback column on (1) why the issue occurred and (2) on the recommended action that should be done. Then, enumerator sends the report to the FO.
f. FO collects all filled-out verification reports from enumerators. FO checks – and if necessary, *modifies* – enumerator feedback on each issue and validates it by putting an "OK" in the column "field_officer_validation". Then, FO sends filled-out report to DO.
g. DO saves all filled-out verification reports in a dedicated folder on a drive shared with AOs (such as onedrive) and checks whether enumerator feedback is logical. If necessary, DO reverts to FOs, with AO in CC, and makes corrections to the feedback in the verification report.
h. DO specifies the *action* to be done for each issue (in the column "action") and the *new value* of each variable, if there needs to be a change (in the column "new.value"). The new.value needs to correspond exactly to the xml names found in the "choices" tab of the KoBo questionnaire. Below the different possible actions:

    1. nothing: No action will be performed
    2. remove: Survey will be removed
    3. change: old.value will be replaced by new.value for question.name
    4. recode: parent.other.answer will be replaced by new.value for parent.other.question
    5. recode_all: All answers of question.name that match old.value will be replaced by new.value

i. AOs double-check all filled-out verification reports found in the shared drive regarding whether the action is justified based on the enumerator feedback. If it is not, they revert back to the FOs for clarification and make necessary changes. If the report is ok, AOs write "OK" in the column "assessment_officer_validation" and saves it in a different folder of the same shared drive.
j. DO saves all filled-out verification report which were validated by AOs in the local folder ("input/verification_reports_filled_out".
k. DO manually goes through the daily raw data excel sheet in the folder "output/raw_data" to check if any types of error can be spotted that are not already accounted for in the logical checks excel sheet (so which are not checked for in the automatic checks in R). If a value in the data needs to be changed and issue was not included in logical checks excel sheet in the input folder, DO will add it to that sheet and will add the cleaning log entry to a separate excel sheet "additional_corrections" (having the same structure as the verification reports) in the "input/verification_reports_filled_out" folder

The verification reports contain the following columns:

Columns already filled out automatically in R – should NOT be changed by anyone:

a. uuid: Unique identifier
b. question.name: XML question name (the name contained in the KoBo questionnaire)
c. old.value: Current response value

d. parent.other.question: The cleaning log will list all "other" answers, this column will have the parent xml question. Example: question.name: food_source_other, parent.other.question: food_source
e. parent.other.answer: The cleaning log will list all "other" answers, this column will have the answers of the parent.other.question. This is helpful to decide if we should keep other as an option, or if a similar option was already selected or if we have a similar option in the list of choices and other should be replaced. Example: question.name: food_source_other, parent.other.question: food_source, parent.other.answer: own_production food_aid other
f. problem: A brief explanation of the logical inconsistency issue detected through the automatic data quality checks in R, as well as explanation in case an "other" response needs to be checked for recoding.

Column to be filled out by enumerator:

g. enumerator_feedback: Here, the enumerator explains why the issue, as mentioned in the "problem" column, occurred and what action he/she would recommend (hence, whether to leave the value, change the value, or recode other response to existing choice) etc.).

Column to be filled out by FO:

h. field_officer_validation: Here, the FO puts an "OK" for validation, if the enumerator feedback is logical.

Columns to be filled out by DO:

i. new.value: If applicable, the new correct XML response value (as found in the choice tab in the Kobo questionnaire).
j. action: we specify how to treat each row of the cleaning log.

Column to be filled out by AO:

k. assessment_officer_validation: Here, the AO puts an "OK" for validation, if the action recommended in light of the enumerator feedback is logical.

2. Weekly and final data cleaning
   a. DO makes sure that for every report sent to enumerators ("output/verification_reports_empty"), there is a filled out report ("input/ verification_reports_filled_out") – meaning that enumerators have provided feedback on all reports they received
   b. DO cleans the data, using R. R automatically compiles all filled out verification reports to one cleaning log, which is used for the cleaning. Clean data is automatically exported to folder "output/clean_data" and cleaning log is exported to folder "output/cleaning_log"
   c. DO runs the post-cleaning data quality checks verifying:
      i. Whether all issues previously detected are now solved
      ii. Whether all select multiple answers have corresponding values in the binary variables
      iii. Whether all "other" responses were translated into English
      iv. Whether any "other" responses selected in the select multiple question are missing
   d. DO manually goes through the clean data excel sheet to check if any obvious error can be spotted.
   e. For each issue detected in the manual and automatic post-cleaning checks, DO adds an entry to the excel sheet "additional_corrections", which serves as cleaning log for issues not included in verification reports.
   f. Once no issues are detected in the post-cleaning checks, DO sends clean data to AO and AM for in-country validation.
   g. Once data is validated in country, DO sends it to HQ data unit for validation.

ANNEX 3: STEPS TO DATA ANALYSIS

1. Daily data cleaning
   - Using the data cleaning log
   - Keeping track of suspicious results per Enumerator
   - Check for minimum duration of surveys
   - Daily (data cleaning/check) and daily email to the Field Manager and AO reporting
     - ✓ The # of KIs per settlement (2 times per week)
     - ✓ the LGA coverage per settlement (1 time per week, unless requested for 2 time per week towards the end of the month)
     - ✓ The enumerator patterns (1 time per week)

   - Responsible persons:
     - **Database Officer** to do the cleaning and sending the email to report issues that need to be double check by the FM and AO.
     - **Field Manager and AO** to keep an overview on recurring issues.
     - **Senior Assessment Officer** to crosscheck weekly.

2. Download data from reach kobo server
   - Responsible persons:

   Database Officer

3. Data cleaning of the entire period in question
   - Horizontal and vertical cleaning
   - General data cleaning process
   - Responsible persons:
     - **Database Officer** to do the cleaning
     - **Database Officer** to crosscheck by looking at the cleaning log and the raw data if the changes made on the cleaning log reflect on the raw data.
     - **Senior Assessment Officer** 2nd crosscheck and validation

4. Run data cleaning script
   - The script ensure cleaning log names are the same as the original (folder", "dataset","sheet","enumerator_id","question", "old", "new", "reason", "_uuid", "modified", "cleaner_name")
   - The correction is based on the new value by uuid and question from the cleaning log, the script looks for what to clean in the raw data based on the new value, question and UUID from the cleaning log.
   - For questions, it replaces "/" with "_"
   - Remove multiple uuid columns (duplicates)
   - Remove if not a header in the data
   - Delete records when the cleaning log header new value=="NULL"
   - Find outliers
   - Remove unneeded indicators like b__geopoint_precision, deviceID, b_geopoint_altitude, index
   - Responsible persons:
     - **Database Officer** to run the script using the cleaning log and raw data in excel and check results by looking at the cleaning log and the cleaned data and replace all missing values with NAs.
     - **Senior Assessment Officer** 2nd crosscheck and in-country validation
     - **Database Officer** to send email with clean dataset, cleaning log and deleted records to relevant people in HQ, with country coordinator, the Assessment Officer and the GIS Manager in cc for final validation

5. Run settlement aggregation script
    - This script aggregates the settlement from the clean dataset to consensus aggregation based on 5% threshold of reporting LGA using the function file.
    - Aggregating function to pick "yes" over "no" responses “AOK_YES”. This checks to see if we have any response other than yes or no, return NC, this is for a response for one settlement, but for a settlement that has more than one response AOK yes then it will pick yes over the other response. Eg yes, I don’t know, no, no response, then it will pick yes. This is for the variables which priorities a YES response. Aggregating function to pick "no" over "yes" responses “AOK_NO”. This checks to see if we have any response other than yes or no, return NC, this is for a response for one settlement, but for a settlement that has more than one response AOK no then it will pick no over other response. Eg yes, I don’t know, no, no response, then it will pick no. This is for the variables which priority is NO.
    - Aggregating function to calculate mode, while outputting NC (no consensus) if we don't have a clear winner. This checks to see if we have more than one mode (a tie), return blank if so. “AOK”
    - Aggregating function to pick most recent responses over others. For frequency questions. This checks to see if we have more than one mode (a tie), return blank if so “AOK_RECENT”
    - Aggregating function to pick most recent responses over others. For duration questions. This checks to see if we have more than one mode (a tie), return blank if so “AOK_RECENT”
    - The script removes columns we don't need. Notes first, no data in those!!
    - Remove "category_ok_" from column headers
    - Creating the settlements dataset
    - Define and It represents all the skips logic with SL
    - Filling in all blank values settlement [settlement == ""] <- "NC"
    - Counting KI coverage per village
    - Remove "other" wards
    - Count settlements per-ward
    - Counts unique settlements
    - Join h2r+ward data and calculate the proportion above threshold
    - Percent of settlements in wards
    - Subset gis+settlment merge and join to main dataset
    - Responsible persons:
        - **Database Officer** to create variables list based on the tool used for data collection
        - **Database Officer** to update, run the script and check results by doing a PIVOT table of the CONSENSUS data
        - **Senior Assessment Officer** to crosscheck results and send to HQ

6. Manual deletion of any settlements that do not meet the 5% LGA threshold for the data collection period in question. Note, for LGA with 5% threshold of less than 5 settlements, a minimum of 5 settlements must be assessed. This is done after running the settlement aggregation
    - Responsible persons:
        - **Database Officer** to do the deletion

7. Run the Ward & LGA aggregation script for analysis on higher admin levels
    - This script analyzes the data from the settlement aggregation call CONSENSUS DATA by LGA or WARD.
    - Pivot/aggregate settlement-level data (settlement proportions)
    - Remove "SL" from aggregation
    - Remove special characters and "consent_yes" from headers
    - Create combined lga and group variable to aggregate
    - Responsible persons:
        - **Database Officer** to run the script and check results

- **Senior Assessment Officer** 2nd crosscheck and validation to crosscheck and validatecheck results by doing a PIVOT table of the CONSENSUS data
- **Database Officer** to send email with clean dataset, settlement aggregation and LGA aggregation to the GIS Manager, with country coordinator, Assessment Officer in cc to HQ

Summary

- Load raw data and cleaning log to run the cleaning script
- Use the cleaned data to run the settlement script for generating the consensus data
- Use the consensus data to run the LGA _ward script to generate the GLOBAL RESULT
- The Global result for reporting.

1. Any other issue flag by the AO.
2. Common Sources.
   - If the settlement coverage has reached the need percentage the settlement should be flagged to the FO.
3. Data records.
   - All issues should be added to the cleaning log, and a backup should be kept in a folder. Eg (All changes done on other option, all logical issues, all deleted survey)

Blanks: if there is a skip logic between two or more questions, the questions following the skip logic trigger are likely to contain multiple blanks given that it doesn't apply to respondents who didn't select yes to the first question