Research Terms of Reference

Afghanistan Prolonged IDPs assessment

REACH Support to Assessment and Information Management in Afghanistan

September 2016
Final

REACH Informing more effective humanitarian action

1. Summary

Country of intervention	*Afghanistan*										
Type of Emergency		Natural disaster		Conflict	X	Complex Emergency					
Type of Crisis		Sudden onset		Slow onset	X	Protracted					
Mandating Body/ Agency	*OCHA*										
Project Code	*02 iABU 0U4*										
REACH Pillar		Planning in Emergencies	X	Displacement		Building Community Resilience					
Research Timeframe	15 July 2016 – 14 January 2017										
General Objective	The overall project objective is to promote evidence-based planning by decision-makers in Afghanistan to support the allocation of limited humanitarian financial resources to priority areas, sectors, and target groups.										
Specific Objective(s)	Specifically, the project will support humanitarian actors to enhance their understanding of the vulnerabilities and needs of prolonged IDPs.										
Research Questions	- What are the numbers and main locations of prolonged IDPs? - What are the key demographic and socio-economic characteristics of prolonged IDPs ? - What are the future prospects of prolonged IDPs ? - What are the humanitarian needs of prolonged IDPs ? - How do humanitarian needs differ between prolonged IDPs and population groups living in the same area?										
Research Type		Quantitative		Qualitative	X	Mixed methods					
Geographic Coverage	Nationwide										
Target Population(s)	Prolonged IDPs (displaced for more than 6 months)										
Data Sources	- Secondary data collection: 1) ERM; 2) IOM RAF; 3) UNHCR IDP tracking; 4) other relevant datasets; - Secondary data verification: 1) Government authorities; 2) NSP/CDC; 3) ERM partners - Primary data collection: household level sample representative of IDPs at identified IDP locations across Afghanistan										
Expected Outputs	- 1 assessment ToRs developed - 1 dataset of locations and estimated numbers of prolonged IDPs - 1 initial map set identifying the location of prolonged IDPs. - 1 clean dataset on prolonged IDP needs and vulnerabilities - 1 updated set of IDP maps										

	- 1 report on the needs, vulnerabilities, and locations of prolonged IDPs.		
Key Resources	REACH Country Focal Point REACH GIS/DB Officer		
Humanitarian milestones			
	Milestone		Timeframe
		Cluster plan/strategy	
	X	Inter-cluster plan/strategy	HNO 2016 (in November 2016)
		Donor plan/strategy	
		NGO plan/strategy	
		Other	
Audience			
	Audience type		Specific actors
	X	Operational	92 UN agencies and NGOs
	X	Programmatic	92 UN agencies and NGOs
	X	Strategic	92 UN agencies and NGOs
		Other	
Access	*X*	Public (available on REACH research center and other humanitarian platforms)	
		Restricted (bilateral dissemination only upon agreed dissemination list, no publication on REACH or other platforms)	
		Other	
Visibility	*CHF, OCHA, REACH*		
Dissemination	*Country level joint analysis workshop, presentation of findings during inter-cluster and/or HCT meetings, country level mailing list, upload on web portals (REACH, Humanitarian Response Info, ReliefWeb, and other relevant platforms)*		

2. Background & Rationale

In addition to being prone to natural disasters—including earthquakes, floods, and avalanches—Afghanistan has suffered from on-going conflict for over three decades. These natural and man-made disasters have engendered tremendous population movements, both out of the country and internally, which in turn have led to increased poverty and the heightened vulnerability of Afghans.

As of 31 March 2016, more than 335,400 Afghans were displaced (UN figures), of whom 90,100 were newly displaced due to conflict in 2016. Of Afghanistan's 34 provinces, 23 provinces recorded some level of forced displacement. Major triggers for recent displacement include the October 2015 earthquake in Badakhshan, flash floods, the Taliban's renewed spring offensive, and continued insecurity. Most humanitarian actors focus their efforts on the immediate needs of those recently displaced as a result of emergencies, leaving a significant gap in both information and support/funding for IDPs in prolonged displacement.[1]

As of May 2012, UNHCR reported that there were still over 74,000 IDPs in Afghanistan who had been displaced before 2003 (UNHCR, Conflict-Induced Internally Displaced Persons in Afghanistan, July 2012). A 2013 IOM survey in several provinces found that in Herat, prolonged IDPs are displaced for seven years, on average (IOM, Displacement Dynamics,

[1] While there is no established distinction between "newly-displaced" and "prolonged" IDPs, when considering this population, REACH will consult with OCHA and other key stakeholders regarding policies on providing assistance to displaced persons, and when IDPs become ineligible for emergency assistance.

2014), and an NRC-commissioned 2012 survey of IDPs in five provinces found that nearly 30% of respondents had been displaced for three to seven years, 10% had been displaced for 8 – 11 years, and 11% had been displaced for over 11 years (NRC, Challenges of IDP Protection, 2012). While it is commonly assumed that people who are displaced for long periods of time eventually find ways to adapt, field data point to the opposite conclusion. Prolonged IDPs often do not have access to livelihoods, land tenure, or regularly provided services, and "require significant assistance to achieve durable solutions" (NRC, Afghanistan – Fact Sheet, July 2014); the aforementioned IOM survey showed that nearly two out of three prolonged IDPs still needed shelter kits and permanent housing (IOM, Displacement Dynamics, 2014). Furthermore, the final NRC report noted that the conditions of IDPs displaced for over ten years were the same as those displaced more recently. More specifically, IDPs in prolonged displacement were equally or more likely than recent IDPs to not have access to land, and both populations had the same food needs. Overall, the data suggests that "IDPs in prolonged displacement…lacked the adaptive advantages they are often believed to possess" (NRC, Challenges of IDP Protection, 2012).

Information gaps and subsequent assessment needs have been initially established through discussions in country with relevant aid stakeholders, including OCHA, sectors coordination teams (ES/NFI and WASH in particular), donors (ECHO, OFDA) and implementing actors (both ERM and non-ERM partners). The 2016 HNO acknowledges the lack of data, sectoral assessments, coordination between assessments, and common data collection standards and methodologies. There are clear gaps and imbalances in the current geographic distribution of assessments, with only one cluster, FSAC—of which ACTED is a member—covering all provinces; several clusters did not conduct assessments in dozens of provinces. Four clusters conducted no assessments in the Central Highland region, three clusters conducted no assessments in the North Eastern and Northern regions, and four clusters conducted fewer than ten assessments each. As of November 2015, only seven needs assessments were planned for the next 12 months, including just three inter-cluster assessments.

There is also a lack of information about prolonged IDPs in particular--including their geographic distribution, the length of their displacement, their needs, and their vulnerabilities. The 2016 HNO does not include any data about prolonged IDPs, and monthly updates from OCHA include province-specific information about newly-conflict-displaced persons, but not prolonged IDPs. This assessment aims to clarify and address these gaps specifically, adding to the knowledge base available to humanitarian actors and improving their capacity to serve this vulnerable and marginalized population.

The humanitarian response to the needs of displaced and crisis-affected communities—and in particular prolonged IDPs—has been limited by numerous factors, including insecurity and access, which hamper the capacity of stakeholders to gather reliable and regular information, monitor needs, and address gaps. Most humanitarian actors focus their information gathering and response on the immediate needs of those recently displaced as a result of emergencies. This leaves a significant gap in both information and support/funding for IDPs who remain in situations of protracted displacement, although the humanitarian and protection needs of this population group can drastically increase over time. Furthermore, numerous challenges related to assessments and information management in Afghanistan hinder the capacity of humanitarian actors and donors to ensure an evidence-based allocation and prioritisation of resources. The lack of comprehensive information on the humanitarian context across affected areas, population groups and sectors poses a challenge towards effective resource allocation, potentially precluding the most vulnerable populations from being assisted. Main challenges are as follows:

• Lack of relevant, reliable, and comparable data and analysis across sectors, populations, and geographical areas: To date, beyond the Emergency Response Mechanism (ERM), the aid community in Afghanistan lacks a systematic and comprehensive approach to the collection and analysis of primary and secondary information as a basis for aid planning and delivery. Even the information available through ERM is not currently used to understand the evolution of IDPs' needs and vulnerabilities over time as displacement becomes prolonged, or to inform cluster planning and key humanitarian

milestones.

• Limited capacity to adopt and roll-out standardised tools and methodologies for data gathering and analysis: Humanitarian actors in-country are in the process of finalising standard data collection tools and approaches for household level multi-sector assessments to be administered on a census basis in the event of new displacement (based on the ERM approach). However, the tool has not yet been rolled-out and systematised across relevant humanitarian partners, and an approach for joint assessments for larger emergencies has not been adopted. Moreover, there is limited in-country analysis capacity to use and process such information as a basis for decision making related to coordination, aid planning and delivery, hindering the efficiency of the information system and its linkages to decision making processes.

In light of the abovementioned challenges, establishing information mechanisms to provide reliable and comprehensive evidence to decision makers on a continuous basis is crucial to enable an effective allocation of resources across sectors, population groups, and areas in Afghanistan. While addressing the assessment and information challenges mentioned above, REACH will specifically focus on gathering reliable information and data on the needs, vulnerabilities, and gaps to support prolonged IDPs.

3. Research Objectives

- To understand vulnerabilities and needs of prolonged IDPs in Afghanistan.

4. Research Questions

- What are the numbers and main locations of prolonged IDPs?
- What are the key demographic and socio-economic characteristics of prolonged IDPs ?
- What are the future prospects of prolonged IDPs ?
- What are the humanitarian needs of prolonged IDPs ?
- How do humanitarian needs differ between prolonged IDPs and population groups living in the same area?

5. Methodology

5.1. Methodology overview

REACH will coordinate with facilitating partners - I/NGOs, UN, and national/regional clusters and will promote inter-agency cooperation through coordination platforms, including the ICCT and the Assessment Working Group (AWG) to ensure the buy-in of clusters. ACTED is active in clusters at regional (North-East/West) and national levels. To ensure coordination with the ICCT, REACH will attend relevant coordination and cluster meetings. Where methodologies/tools are already available through clusters or other partners, REACH will not duplicate efforts. When it is necessary to create new tools/methodologies, these will be developed with the cooperation of clusters and the AWG. When finalized, these products will be handed over to IMs to promote consistency across sectors. REACH will use secondary data available through existing reports, databases, maps, partners, and the cluster system, particularly during phase I of the project to identify location and population estimates of prolonged IDPs.

Primary data will be collected directly from the field by REACH, security and access permitting. Probability sampling methods will be used to enable generalisations to the population of interest (prolonged IDPs) with a specified level of precision. When primary data collection is infeasible due to security, REACH will work with partners already in the field to collect data, either by embedding staff in partners' teams, or providing support from the capital or provincial level. The primary data collection will be conducted at household level, where possible using a mobile data collection platform, which enables geo-localisation of data points, allowing REACH to monitor proper implementation of the random sampling strategy and conduct geo-spatial analysis. When mobile data collection is impossible, data will be collected using paper forms and action plans revised to reflect time/resources needed for data cleaning/processing.

5.2. Population of interest

The population of interest to this assessment include all prolonged IDPs across Afghanistan. The exact definition of 'prolonged' will be established during the secondary data review, in consultation with OCHA and the clusters, to ensure findings can be generalised to a clearly defined population. This definition will likely include the duration of displacement.

Please find definition in Annex 2

5.3. Secondary data review and verification

In coordination with OCHA, UNHCR, ERM partners, clusters, the ICCT, and the Assessment Working Group, REACH will collect and gather all available secondary data (quantitative and qualitative) on prolonged IDPs. The data sources will be triangulated with the aim to:

1) Estimate prolonged IDP numbers
2) Identify & map prolonged IDP locations across Afghanistan
3) Inform the primary data collection methodology and tool design
4) Provide a basis for triangulation with primary data

Key data sources that will be requested from partners for inclusion in the triangulation include:

- ✓ ERM
- ✓ IOM RAF
- ✓ UNHCR IDP tracking
- ✓ Any other IDP tracking data sources

It is expected that the triangulated data will need verification to ensure best possible estimates and may in addition reveal geographical coverage gaps where no data currently exist. A short primary data collection exercise may therefore be undertaken both to verify and to fill gaps in the secondary data. This data collection would focus on sources that are believed to hold the most up-to-date information about displacement, including:

- ✓ National/provincial level government authorities
- ✓ NSP/CDC
- ✓ ERM partners

Please find Aggregation methodology in Annex 3

When security and access permits, this data will be collected through key informant interviews and direct observation at the field level. When primary data collection is not possible due to access and security, remote data collection will be used. If relevant, UNOSAT remote sensing techniques will confirm the presence of prolonged IDPs in hard-to-reach areas.

Please find Verification and Identification methodology in Annex 4

5.4. Primary Data Collection

The primary data collection methodology will be finalised based on findings from the secondary data review. In particular, the secondary data review will identify necessary parameters for the sampling framework (e.g. estimated number of locations / number of prolonged IDPs / strata of interest); inform development of indicators (e.g. depending on type of geographical setting where prolonged IDPs are living); and help establish the definition of 'prolonged IDP' (e.g. duration of displacement).

REACH will liaise with OCHA, the clusters, ICCT and Assessment Working Group on the data collection tools and analysis framework to strengthen assessment ownership by key stakeholders, and consider access constraints to develop methodologies appropriate for each target area.

The aim is to use a probability sampling strategy to randomly select households at identified IDP locations, thereby enabling generalisation with a specified level of precision of findings to the population of interest. The target precision of the sampling strategy is generalisation to the prolonged IDP population at each strata level, with a minimum of 90% level of confidence and a maximum of 10% margin of error. Relevant strata will be identified through the secondary data review (e.g. rural/urban IDPs; natural disaster/conflict IDPs). In order to address the final research question on difference in needs compared to neighbouring populations, a sample will be drawn of households directly neighbouring sampled prolonged IDP households.

Depending on the number of locations identified to be hosting prolonged IDPs, a cluster sampling approach may be used to minimise the number of locations that need to be accessed by the data collection team. A larger sample would in this case need to be interviewed to counter the impact of the resulting design effect on level of precision. As mentioned above, this decision rests on secondary data review findings, where IDP locations are identified.

Please find the primary data colletction sample summary table in Annex 5

Once the sample strategy has been identified, implementation of sampling can be conducted, here again the method of implementation will depend on secondary data findings. It is essential to ensure as equal probability as possible for households in the population of interest to be selected for inclusion in the sample. Possible methods include:

- In the unlikely event of IDP household lists existing in locations, these can be used to randomly select households up to the target sample at each location
- IDP household lists could be constructed at location, should the IDP population represent a relatively small number and proportion of the community population
- Randomly generated GPS points could be used should the IDP population represent a relatively large proportion of the community population (or be located within a defined geographical area within the community)

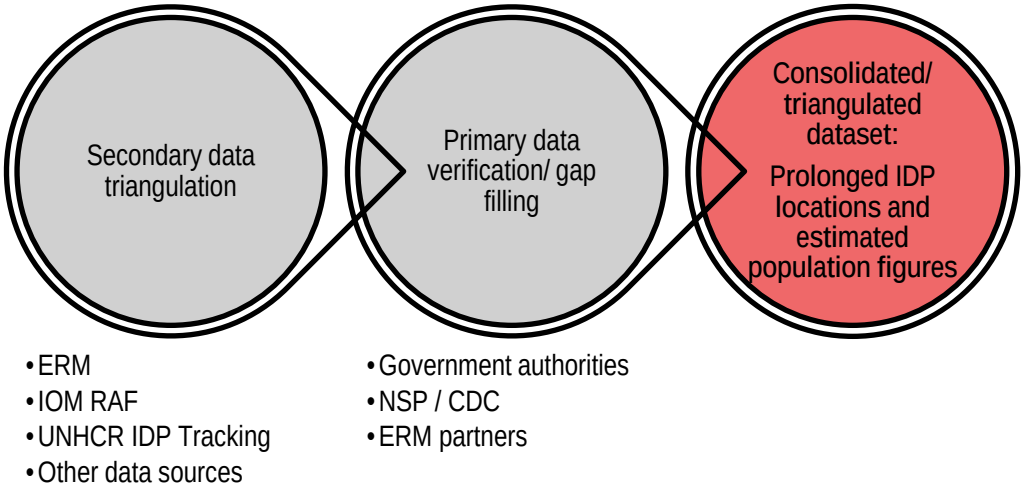
A data collection tool will be developed based on indicators proposed to and reveiwed with, OCHA and the clusters. Indicators should be based on existing indicators as far as possible, to maximise comparability with existing datasets. A questionnaire will be built on a mobile data collection platform (e.g. Kobo) and downloaded on tablets/smartphones to be used to record interviews in the field. Completed forms are uploaded to a dedicated server and monitored on a daily basis for any errors or misunderstandings that can be corrected and fed back during daily debriefings with the data collection team. GPS coordinates recorded for each form at the location of the interview can be plotted to monitor compliance with the sampling strategy. Once data collection is completed the data collection team conducts a thorough end of assessment debriefing with the Country Focal Point & GIS/DB Officer to review and aid interpretation and analysis of collected data. In total, a team of 30 enumerators will be recruited for 2 months to conduct the secondary data review verification and primary data collection exercises. In addition to data collection ,this period must include training, travel between locations and debriefing sessions.

5.5. Data Analysis Plan

Once data is collected and any cleaning resulting from the debrefing sessions is completed, the final dataset will be analysed and triangulated with secondary data sources with a view to address the research questions (see above). Upon completion of the analysis, REACH will present preliminary assessment findings in a dedicated workshop in Kabul with all aid actors that have been involved in the assessment process, including clusters and members of the ICCT, to enable contribution to the analysis and discussion around preliminary findings. REACH will consolidate workshop outputs and

recommendations into a final assessment report, including a set of prolonged IDP location maps. Finally, upon validation by OCHA Afghanistan, REACH will upload a final clean assessment dataset and other products on OCHA's HDX web-portal or on other appropriate platforms identified in consultation with OCHA, the ICCT, clusters, and the Assessment Working Group. The data analysis plan will be finalised based on the indicators agreed with humanitarian actors during the initial finalisation of the assessment terms of reference.

Figure 1: Data analysis methodology



6. Product Typology

Table 1: Type and number of products required

Type of Product	*Number of Product(s)*	*Additional information*
Assessment TOR	1	
Dataset	2	1 x Location and estimated numbers of prolonged IDPs 1 x Prolonged IDP vulnerabilities and needs
Report	1	
Presentation	1	
Map	2 (sets)	1 x Estimated locations of prolonged IDPs 1 x Updated locations of prolonged IDPs

7. Risks & Assumptions

Table 3: List of risks and mitigating action

Risk	*Mitigation Measure*
Secondary data to enable estimation of IDP location and numbers is not available	Collection of primary data from sources believed to hold most up-to-date information about prolonged IDPs
Partners and/or ministries are unwilling to cooperate	Clear communication and outreach to key stakeholders conducted during initial stages of the

	project
IDPs are unwilling to participate in assessment	Clear communication about the objectives of the assessment to help relay the value of the data collected; replacement sample strategy in case of refusal to particiate
Security concerns prevents field data verification and collection	Daily monitoring of security situation to identify safe periods of data collection; outsourcing of data collection to local partners

8. Documentation Plan

The following key documents and outputs should be filed for future reference:

- Assessment Terms of Reference
- Consolidated secondary data review matrix
- Primary data collection tracking matrix
- Clean primary dataset
- Primary data cleaning log

9. Annexes

1. Data Management Plan
2. Definition of Prolonged IDPs
3. Aggregation Methodology
4. Verification and Identification Methodology
5. Primary data collection sample summary

Annex 1 : Data Management Plan

V3

Checklist Section	*Questions to consider and guidance*
Administrative Data	
Project Name	Afghanistan Prolonged IDP Assessment
Project Code	02CPN
Donor	CHF
Project partners	ACTED, IMPACT initiatives and UNITAR//UNOSAT
Project Description	The overall project objective is to promote evidence-based planning by decision-makers in Afghanistan to support the allocation of limited humanitarian financial resources to priority areas, sectors, and target groups. Specifically, the project will support humanitarian actors to enhance their understanding of the vulnerabilities and needs of prolonged IDPs. In coordination with OCHA, UNHCR, ERM partners, clusters, the ICCT, and the Assessment Working Group, REACH will collect and gather all available data (quantitative and qualitative) on prolonged IDPs, who are reported at risk of prolonged displacement. REACH will then analyze this secondary data to: 1) Identify and assimilate available information on prolonged IDPs' numbers, needs and vulnerabilities; 2) Identify prolonged IDPs' locations across Afghanistan and, in coordination with OCHA and the clusters, classify these locations according to levels of access. As a result, a first mapping of the location of prolonged IDPs will be available; 3) Inform the assessment analysis framework and data collection tools design; 4) Provide the REACH team with data enabling a future triangulation of the assessment findings.
Project Data Contacts	vincent.annoni@impact-initiatives.org; elisabeth.vikman@reach-initiative.org thomas.stork@reach-initiative.org
DMP Version	27 August 2016 V3
Related Policies	
Data Collection	
What data will you collect or create?	This research contains two stages of data collection 1. Secondary data review 2. Primary data collection. 1. Secondary data review: REACH will request to use all relevant currently available data sources (ERM, UNHCR, OCHA and other project partners) to identify prolong IDPs locations. 2. Primary data collection: Primary data will be collected by REACH data collectors through household interviews using probability sampling, with parameters identified through the secondary data review (e.g. number and types of strata; probability sampling method). The primary data will be collected using ODK forms (where security permits) and the processed data will be shared with all stakeholders.
How will the data be collected or created?	REACH will deploy data collection team to identified prolonged IDP locations according to the specified sampling strategy. Data will be sent to ODK-based server by data collection teams. The REACH database officer will review incoming data for potential errors; check and verify any possible corrections with the data collection teams. Once a data collection is completed, the final

	dataset is exported and checked for errors, with any verifications and corrections made, recorded in a data cleaning log. Raw and master databases are saved on the REACH server using REACH file name & document title standards. Electronic File Name: REACH_Countrycode_typeofdocument_Crisisname_mandatingbody_MonthYear Document Title: Countrycode_producttype_crisisname_monthYear
Documentation and Metadata	
What documentation and metadata will accompany the data?	For better understanding and reuse of this assessment result as secondary data by stakeholders, REACH will produce a package of data, which contains cleaned database, factsheet, analysis, and maps. REACH will also add meta-data in the data-set of this assessment which contain: 1. Methodology of the assessment 2. Limitations of the methodology 3. Year of the survey 4. Geographical coverage of the survey 5. Tag of sectors/thematic covered by the assessment 6. Description of any composite variables created 7. Data cleaning log
Ethics and Legal Compliance	
How will you manage any ethical issues?	In accordance with the Code of Ethics and Conduct, REACH will ensure that every person from whom data is gathered for the purposes of research consents freely to the process on the basis of adequate information. They will also be able, during the data gathering phase, to withdraw freely or modify their consent and to ask for the destruction of all or part of the data that they have contributed. Throughout training of assessment teams, it will be emphasized that participants are not obliged to provide information they feel poses a risk to their well-being or if they feel this may cause a threat to their personal safety. Through constant feedback, such instances are reported to inform continuous improvements to training. Personal identifiable information will not be publicly disseminated to minimize/eliminate protection concerns for the assessed population. All data will be aggregated to a location and no household identifiers will be publicly visible from the reports and maps. If agencies request the raw data containing household identifiers, then the sensitive name/contact details will be removed and replaced with a unique key so that the identifier information can be re-connected at a later date, based on protection standards. If there is a further request for the names, REACH will contact OCHA (or in the case of secondary data, the discloser of this data) to examine whether the requesting agency can receive the identifier data or not, and if so, under which conditions. REACH will not collect personally identifiable data. REACH will include a request for consent within assessments to share household level information collected, which will only occur under the following circumstances: - For REACH internal use for the purposes of data management (i.e.

	avoiding duplicate data collection) - With donor / partner organization where data-sharing agreement exists, and only for humanitarian purposes - With partner organization and/ or agency where specific and urgent need occurs, for example urgent medical needs or for other referral mechanism. - With the owner (i.e. the assessed party) of the information if requested. REACH will work closely with relevant clusters to ensure assessment methodologies, indicators and analysis is sensitive to gender concerns and wider protection issues. Data collected will be disaggregated by age and gender, with the ability to identify vulnerable households. REACH will also work with relevant agencies in protection and health to ensure the referral of urgent cases. Specifically, all questionnaires and assessments will be designed in coordination and collaboration with relevant cluster leads, including Protection, Food security, WASH, Education, Health and displacement, with close coordination with OCHA. The sensitivity of questions is reviewed in coordination with protection colleagues. Specific protection assessments will be defined in close liaison with the protection cluster at a later stage.
How will you manage copyright and Intellectual Property Rights (IPR) issues?	The anonymized dataset will be uploaded on the OCHA HDX web portal under open data license. Any anonymized secondary data that is incorporated in datasets uploaded on the OCHA HDX web portal will be fully referenced acknowledging the original data source.
Storage and Backup	
How will the data be stored and backed up during the research?	REACH will be responsible for data storage, back up, and data recovery. Multiple data storages will be used to maximize data security, as outlined below: 1. ODK-based server: The ODK server will be administrated by Impact HQ GIS team in Geneva, to which a limited number of REACH staff will have access (the device setting will only contain the URL of the forms and no password) and whenever any data is requested as per guidelines, it will be extracted from ODK-based server. 2. REACH country server: a. Pre Assessment: Before starting any assessment, specific separate folders will be made for each assessment (considering REACH documentation system) and will be protected by passwords b. During Assessment: A daily backup will be extracted from ODK server into and saved as an xls file in the specific assessment folder. c. Post Assessment: After completion of data collection REACH database officer will clean the data according to data cleaning guidelines and stop accepting submissions into ODK server for the specific assessment. Raw and cleaned data sets will be stored on the REACH country server xls format. 3. REACH global cloud: The final cleaned database of the assessment will stored by REACH HQ Geneva on the REACH Global Server in the

	CERN.
How will you manage access and security?	The access policy to the data will differ by the time of assessment (during assessment and after assessment). 1. During assessment: only designated technical staff (REACH GIS officer, Assessment officer, and database officer) will have access to ODK-based server. 2. After assessment: when data collection process is completed the cleaned anonymised data will be uploaded on the OCHA HDX web portal under open data license.
Selection and Preservation	
Which data should be retained, shared, and/or preserved?	REACH will not destroy any of the data set included in this research but will apply information anonymization policy (replacing sensitive fields in the data into codes) to ensure the sensitive information of households will not be shared with irrelevant parties.
What is the long-term preservation plan for the dataset?	Due to data security REACH will not keep any paper form (hard filling) from this assessment's data-set, The data set of this assessment will be archived virtually on the REACH country server, and global cloud as REACH primary data. REACH or other stakeholders can benefit from this information in future assessments, reports, and proposals.
Data Sharing	
How will you share the data?	The processed data (completed, cleaned, analyzed, and validated data) will be shared with all stakeholders through HDX web portal, humanitarianreponse.info website and REACH resource center. A copy of the dataset will be provided to cluster Information Managers upon request. Any anonymized secondary data incorporated in these datasets will be fully referenced, acknowledging the original data source.
Are any restrictions on data sharing required?	REACH will apply restrictions only on those data-sets which contain sensitive information such as beneficiary contact details, personal information and complainant identity. REACH will apply an anonymization policy, unlinking all sensitive information from the dataset while ensuring a unique record identifier is in place that enables reconnection of the information. If there is a further request for the sensitive information, REACH will contact OCHA (or in the case of secondary data, the Discloser of the data) to examine whether the requesting agency can receive the identifier data or not, and if so, under which conditions. No data will be disseminated before completing the data process (data cleaning and data validation).
Responsibilities	
Who will be responsible for data management?	REACH will be responsible for the assessment's data-flow and implementation of DMP and every single step of data collection and data process will be managed by REACH database officer. REACH will take lead in collaborative researches across partners and every server of partners will share their forms to REACH server.

Annex 2 : Definition of Prolonged IDPs

Methodological note – 22 August 2016 (updated 30 September)

Afghanistan Prolonged IDP assessment

1. Detailed definition of Prolonged IDPs for the purpose of this assessment

1.1. Internally Displaced Persons

Internally Displaced Persons will here be defined in line with the UN Guiding Principles on Internal Displacement[2] and the National Policy on Internally Displaced Persons:[3]

"…. 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 generalized violence, violations of human rights or natural or human-made disasters, and who have not crossed an internationally recognized State border."

"Homes or places of habitual residence" will here be specified as village (rural) or neighborhood (urban) of origin, hence persons that have been forcibly displaced from their village or neighbourhood to another village or neighbourhood within Afghanistan will be considered as internally displaced persons for the purpose of this assessment.

"Forced or obliged to flee or to leave" will here be specified as the status at the time of the original displacement, regardless of factors triggering subsequent movements, following the original displacement.

1.2. Prolonged Internally Displaced Persons

Further refining the definition outlined above for the purpose of this assessment, Prolonged IDPs will here indicate IDPs that have been displaced:

- After 1 January 2014
- Before 1 March 2016

These parameters are set to meet the overall objectives of the assessment:

- ✓ To address information gap in consolidated information on vulnerability and needs of longer-term IDPs, as identified by OCHA and humanitarian actors. Several ongoing data collection initiatives cover IDPs within 6 months of displacement, hence the key information gap starts once IDPs have been displaced for more than 6 months, which the primary data collection component of this assessment will focus on.
- ✓ To identify estimated numbers and locations of prolonged IDPs based on available secondary data. After reviewing core secondary data sources (UNHCR PMT; IOM RAF; ERM), the secondary data available prior to 2014 was concluded to be too incomprehensive to serve as a basis for number and location estimates.
- To triangulate secondary and primary data. With 6 months and longer identified as the start of the overall information gap and 1 January 2014 identified as the start of reliable secondary data, this period was adopted for both secondary and primary data to enable triangulation of findings. Furthermore, to ensure cleaned and aggregated secondary data covering IDPs displaced for at least 6 months as of 1 October 2016 could be directly

[2] UN Guiding Principles on Internal Displacement (1998); Natural disasters as a cause of internal displacement were added to the Guiding Principles through the Inter-Agency Standing Committee's (2006) "Operational Guidelines on Human Rights and Natural Disasters"

[3] The Government of the Islamic Republic of Afghanistan, Ministry of Refugees and Repatriation, 25 November 2013

matched with the full primary data sample, primary data collection will also focus on households displaced before 1 March 2016.

1.3. Returnees

Returnees (persons returned to Afghanistan from other countries) will be considered prolonged IDPs for the purpose of this assessment when:

- ✓ They have since returning been displaced from a location in Afghanistan to a location in Afghanistan between 1 January 2014 and 1 March 2016.

Annex 3 : Aggregation methodology

V2

Methodological note – 21 September 2016

Afghanistan Prolonged IDP assessment: Aggregation Methodology

A. Objectives

1) Generate one dataset aggregating UNHCR PMT; IOM RAF; and ERM data on persons displaced to locations between 1 January 2014 and 1 March 2016
2) Aggregate the data sets to obtain a lower and upper estimate for 2014, 2015 and January/February 2016 respectively, for each village where persons were reported displaced to.

B. Overall methodology

The aggregation methodology was first conducted within each data set for each year individually.

- Step one: Aggregate recorded locations within each dataset to village level.
- Step two: Join all datasets to create one aggregated village level dataset

The result was a dataset with a unique record for each village (within each unique province and district), with values (where existing) for each dataset, across three years. This was complemented by basic calculations to present an estimated number of persons recorded as displaced to each village/site. The following sources were included in the aggregation:

Table 1: Data sources included in the aggregation: variables included[4]

Data Source	Lowest Geographical Level	Province name variable	District name variable	Village name variable	Lat Column	Lon Column	Population individuals variable
ERM	Village	Province	District	Village	XDEST	YDEST	Total members of HH
PMT	Village	Province of Displacement	District of Displacement	Village of Displacement	-	-	Individuals
OCHA	District	#admin1+name+x_destination	#admin2+name+x_destination	-	-	-	#affected+displaced

[4] Some entries in the aggregated datasets seemed to refer to a more granular level than village (e.g. camps). These entries have been included retaining the name entered under Village with a view to identify the Village nearest to the site during the Verification and Identification exercise that followed the Aggregation.

RAF	Village	Province	District	Village or Nahya	-	-	IndividualsDisplaced
PMT – returnees	Village	Province of Displacement	District of Displacement	Village of Displacement	-	-	Individuals

C. Preparing the datasets for aggregation: data cleaning

Provinces

To complete step two in the aggregation process (joining all datasets) a unique and consistent identifier was needed for provinces and districts. Ideally, the province and district names in each record would be matched with the standardized list of provinces and districts obtained from OCHA (the Common Operational Dataset). However, the spelling of province, district and village names was found to be inconsistent, both within and between datasets. To enable step two in the aggregation process, province and district names were cleaned. However at village level, given that many villages have very similar names, which could easily be mistaken for different spellings of the same village, names were retained as per the original datasets and included as individual entries in the aggregated dataset, to enable field teams to verify which records were in fact differently spelt names of the same village, during the verification and identification exercise. For each dataset, a csv file cleaning log was created to record all corrections carried out on the original datasets, as seen below in table 2 for the ERM dataset:

Table 2: Aggregation cleaning log

old_province	new_province	province_code
Baghlan	Baghlan	9
Balkh	Balkh	18
BADAKHSHAN	Badakhshan	15
BADGHIS	Badghis	29
Badkhshan	Badakhshan	15
Badakhshan	Badakhshan	15
BAghlan	Baghlan	9
Baghaln	Baghlan	9
Baghla	Baghlan	9
BAGHLAN	Baghlan	9

Each occurrence in the old_province column was given a correct value, either through an excel VLOOKUP or through manual cross checking of the names. If a correct value could not be found, “n/a” was added to the “new_province” and “province_code” columns. This cleaning process was followed for all datasets, for both province and district. A short section of code was then created in Python to create a new variable for province and district names which was populated with the correct spelling of all provinces and districts for each entry in each dataset.

1. New column is created – either “Province_Match” or “District_Match”.
2. New column selects the values from the original, unclean province or district column.
3. New column matches the unclean value with the column “new_province”.
4. The process is then repeated for the match between “new_province” and “province_code”.

This resulted in four columns being added to all datasets: Province_Match, Province_Code, District_Match and District_Code. The process was then repeated, matching the newly cleaned Province with the Region they belong to.

D. Preparing the datasets for aggregation: filtering by Year

The datasets that were aggregated covered a number of years, as outlined in the table below:

Table 3: Years with recorded displacement – by dataset included in aggregation

Data Source	2014	2015	2016 (until 1 March 2016)
ERM		X	
PMT	X	X	X
OCHA			X
RAF	X	X	X
PMT - returnees		X	

A column called "Year" was manually added to the datasets, either sourced from a date column in the dataset, or from information gathered on the year the dataset came from.

Each dataset was then divided by year, resulting in one distinct dataset per source and year, as seen for PMT below:

1. "PMT_2014"
2. "PMT_2015"
3. "PMT_2016"

E. Aggregation within datasets

The aggregation was first conducted *within* each individual dataset, to obtain one record for each village.

For each aggregation, several indexes were used. For example, several village records were only aggregated into one where a perfect match was found for Region, Province, District and Village across records. Where a match was missing, the records remained separated. This process was followed to ensure that where the same village name, e.g. Village_1 was found in Districts 1 and 2; these were not aggregated into one single record.

F. Joining datasets

The datasets were then joined (merged) into one aggregated dataset. As each dataset may or may not contain villages (or even districts or provinces) found within other datasets, we used an "outer" join. This means, for example, if we were to join two datasets such as:

Dataset 1

Village	Value
a	12
b	54

Dataset 2

Village	Value
b	31
d	45

The resulting joined dataset would be:

Village	Value
a	12
b	67
d	12

When joining the datasets, the join was based on the sum of the values (as opposed to the maximum value or the average value). The joined datasets resulted in a final aggregated dataset, with a column for each dataset and year.

Table 4: Aggregated dataset structure – aggregated variables

ID	Type	Region	Province	District	Village	PMT 2014	RAF 2014	ERM 2015	PMT 2015	RAF 2015	Returns 2015	ERM 2016	PMT 2016
1													

G. Calculations

For each year across all datasets, the following calculations were conducted for each village entry:

- ✓ The lowest recorded value from all sources that cover each year
- ✓ The highest recorded value from all sources that cover each year
- ✓ The source of the lowest and highest value for each year
- ✓ The sum of the lowest and highest values across the three years
- ✓ The sum of the lowest and highest (minus the number of persons recorded as returned) across the three years

Table 5: Aggregated dataset structure – calculated variables

2014 Lower Est.	2014 Upper Est.	2015 Lower Est.	2015 Upper Est.	2016 Lower Est.	2016 Upper Est.	TOTAL Individuals reported displaced TO location Lower estimate	TOTAL Individuals reported displaced TO location Upper estimate	TOTAL Individuals reported displaced TO location MINUS RETURNS Lower estimate	TOTAL Individuals reported displaced TO location MINUS RETURNS Upper estimate

H. Repeating the Process

The following steps should be followed to repeat this process:

- ✓ For existing datasets, new data can be appended to the bottom of the datasets aggregated here. As long as the required columns are populated with the appropriate values, the existing script could be applied to regenerate the aggregated dataset including new data.
- ✓ The values province and district name variables need to be compared to the names in the existing cleaning log. If new names appear in the new data, they can be added. The cleaning sheets mean that the more data added the less likely new names will appear. For datasets not already included in the aggregation, the same steps are required, with the datasets added to the "Joining datasets" and a new cleaning sheet created.

Annex 4 : Verification and Identification Process

V6

Methodological note – 21 September 2016

Afghanistan Prolonged IDP assessment: Verification and Identification Process

I. Objectives

1. Verify if people that were recorded displaced during the period of interest (1 January 2014 to 1 March 2016) are still displaced at the respective location (Prolonged IDPs); and
2. Identify any additional locations to where people were displaced during the period of interest, where they are still displaced

J. Tools

Please print and save the following tools for each data collection team:

- ✓ Verification list – one and the same copy used for all rounds: listing provinces/districts/villages where IDPs have been recorded displaced to, between 1 January 2014 and 1 March 2016, with spaces for recording verification information
  - o One paper copy; A3 landscape; fit all columns to one page.
  - o One soft copy (Excel)
- ✓ Identification list – one and the same copy used for all rounds: template to record locations that do not exist on the verification list
  - o One paper copy; A3 landscape; fit all columns to one page.
  - o One soft copy (Excel)
- ✓ District maps – one and the same copy used for all rounds: showing the locations of all displacements listed on the verification list and all villages in the district, to enable identification of any locations of previously unidentified displacements
  - o One paper copy per district, A3.

K. Overall Verification and Identification methodology

Round 1 (Regional/Province level): The teams first visit Regional and Province level offices of UN, NGO, MoRR and other government bodies to conduct the first round of Verification and Identification, following these steps:

- The Verification list paper copy; Identification list paper copy; and District maps are used to record information gathered from each office.
- All answers recorded on the Verification list paper copy are entered into the soft copy version of the tool on a daily basis.
- The Verification soft copy is emailed to REACH coordination at the end of every day.
- The paper copy of the Verification list; Identification list; and District maps are scanned and emailed to REACH coordination at the end of every week.

Round 2 (District level): For each district where entries have not been possible to verify, the teams visit UN/NGO and Government district offices to conduct a second round of Verification and Identification, recording answers on the same tool paper copy used in Round 1 and following the same procedure outlined above.

Round 3 (Village level): Where entries remain unverified after the Regional/Province/District level verification, village level key informants will be contacted to obtain clarification on whether prolonged IDPs remain. If verification remains unclear after consulting key informants, visits to specific sites may be undertaken, recording answers on the same tool paper copy used for Round 1 and 2 and following the same procedure outlined above.

L. Verification and Identification process to follow in each Round

The same process is followed and the same copy of the tools are used during each Verification and Identification Round, until all entries on the Verification list are verified

Verification: for all entries where Village names are listed on the Verification list

1. Verification list: Find out if people listed as displaced to each Village during the period of interest (1 January 2014 to 1 March 2016), remain in each Village.
    a. For each Village (each row in the Verification list), record the answer to each of the questions listed in the red columns:
        i. Out of the total persons displaced to each Village during the period of interest (1 January 2014 to 1 March 2016):
            1. How many remain in the Village at this time? (Q1)
                a. Probe about Lower and Upper Est. (estimates) – and any reported returns (D1; D2; D3)
            2. Which source is verifying this? (Q2)
            3. What evidence does the source have for these figures? (Q3)
                a. Documented records by the office
                b. Documented records by a secondary source (to be specified)
                c. General knowledge only
            4. Comments and explanations (Q4)
        ii. To avoid accidentally merging two different villages with similar spelling, all village names with recorded displacements are listed on different lines in the Verification list. This is because the:
            1. Same village name is sometimes spelt in different ways
            OR
            2. Different villages may have very similar names
                a. When you review the list with your Sources, please confirm where the Same village has been spelt in two or more ways, and enter the correct spelling under the 'Village name corrected' column (G1) where appropriate

i. If the same village has been spelt in two or more ways, identify the number of individuals remaining (Q1) out of the total listed on each of the lines for the same village.

iii. Some district names are unidentified although the Village name has been recorded. Please review these Villages names and enter the District names that your Sources can confirm under the 'District name corrected' column (G2).

****If an office report that a recorded displacement is incorrect – i.e. they do not believe it exists – Enter this in the Comment section (Column XX) and confirm with other offices****

Figure 2: Verification list – example

			VERIFICATION		VERIFICATION	2014		2015		2016					VERIFICATION			
ID	Province	District	Q1: District Name Corrected *Enter when applicable*	Village	Q2: Village Name Corrected *Enter when applicable*	Lower Est.	Upper Est.	Lower Est.	Upper Est.	Lower Est.	Upper Est.	D1: TOTAL Individuals reported displaced TO location Lower estimate	D2: TOTAL Individuals reported displaced TO location Upper estimate	D3: TOTAL individuals reported to return FROM location out of those displaced (2015)	Q3: Individuals remaining at location, out of those displaced to location between 1 January 2014 and 1 March 2016 only (D1-3) *Enter Number*	Q4: Source providing information about Q1 *Enter Source Name*	Q5: Type of Source knowledge information about Q1 is based on *Enter type of Source knowledge (e.g. own documented records; other organisation documented record; general knowledge)*	Q6: Comments *Explain higher figures or previously unidentified locationsor other details*
854	Badakhshan	Argo		AB BAREK		4900	4900					4900	4900					
855	Badakhshan	Argo		ALI MANKO								0	0					
856	Badakhshan	Argo		*Unidentified*								0	0					
857	Badakhshan	Baharak		Baharak		927	927	3850	3850			4777	4777	3294				
858	Badakhshan	Baharak		Dasht Faragh		215	215	143	143			358	358	1055				

2. District map: Check that all Villages where prolonged IDPs are confirmed to still remain in the Verification list, are marked on the District map.
    a. If a Village is not marked on the map identify the approximate location of the Village on the District map and mark the following on the map:
        i. Village location – mark with "X"
        ii. Village name – write name next to the X on the map using same spelling as Verification list

Identification: for all entries where Village names are recorded as *"Unidentified"* on the Verification list

3. Identification list: Find out if there are any additional villages in the district not listed on the Verification list (or in Districts not listed on the Verification list, if District name is also 'Unidentified), where people were displaced to during the period of interest (1 January 2014 to 1 March 2016), who are still displaced;
    a. Record any additional Villages where sources say Prolonged IDPs are currently living, on the Identification list:
        i. Village name/Province/District – use the same spelling as what you see on the District Map for each Village
            1. If Village name does not exist on District map:
                a. Mark location with "X" on District map and write Village name next to the X
                b. Record the Village on the Identification list using exactly the same spelling as you used on the map
            2. Village longitude and latitude – if already known by your source
            3. Number of IDPs displaced to the Village during the period of interest (1 January 2014 to March 2016)
            4. Number of IDPs displaced to the Village during the period of interest, that still remain at the location
            5. Source and type of Source knowledge

Figure 3: Identification list - example

Q1a: Village Name (English)	Q1b: Village Name (Dari /Pashto)	Q2a: Village Latitude (if known)	Q2b: Village Longitude (if known)	Q3: Marked on the map *Enter: 'Exists' if village exists on map OR 'Marked' if village does not exist and was marked with 'X' on map*	Q4: Province	Q5: District	Q6: TOTAL Individuals displaced to location between 1 January 2014 and 1 March 2016	Q7: Individuals remaining displaced to location out of TOTAL (Q6)	Q8: Individuals remaining displaced from location out of TOTAL *Enter Source*	Q9: Individuals remaining displaced from location out of TOTAL *Enter type of Source knowledge (documented records, general knowledge)*	Q10: Comments
Qadzyan		34.8746	71.1468	Exists	Kunar	Asadabad	38	38	Asadabad District Office	General knowledge	

Annex 5: Primary data collection target sample

ESTIMATED RATE OF INTERVIEWS	Quantity
Weeks	5
Travel / introduction to local authorities days	1
Interview days	4
Interviews / day	4.5
Data collectors	48
Total interviews	4320

	Sample
PROLONGED IDPS - Regional level	
Geographical strata (Regions)	8
Level of confidence per strata	90%
Margin of error per strata	5.8%
Actual simple random sample required per strata	200
Adjusted sample for estimated design effect per strata	400
Total sample to be collected on Prolonged IDPs	3200
PROLONGED IDPS - National level	
Actual simple random sample accounting for design effect	1600
Level of confidence at national level	95%
Margin of error at national level	2.5%
NEIGHBOURING POPULATION GROUPS - National level	
Level of confidence at national level	95%
Margin of error at national level	5%
Actual simple random sample required at national level	385
Adjusted sample for estimated design effect at national level	770
TOTAL REQUIRED SAMPLE	3970
TOTAL REQUIRED SAMPLE with 10% buffer (non-responses / errors)	4367