International Journal of Information Systems for Crisis Response and Management
Volume 11 • Issue 1 • January-June 2019

Is Ubiquitous Technology for Needs Data Management a Game Changer in Humanitarian Arena?

Punya Prasad Sapkota, UNOCHA, Lalitpur, Nepal
Kashif Siddiqi, UNOCHA, Rawalpindi, Pakistan

ABSTRACT

One in every 70 people around the world is caught up in a crisis (natural disasters, conflict, climate change, etc.) and urgently needs humanitarian assistance and protection according to the OCHA. The humanitarian community assists millions of people every year based on emerging humanitarian needs. Most of the time, the conditions inside the countries, once the humanitarian needs data is collected, are not very conducive and required simple ways to collect data like paper-based data collection with simple questions. This data is later entered into a database or spreadsheet using rigorous and time-consuming data entry efforts. Dynamic changes in needs of people; numbers of partners involved; the complexity of evolving processes; and emerging technologies over time has led to a change in processes for data collection and management. This article is an attempt to capture humanitarian data collection best practices and the use of different technologies in managing data to facilitate humanitarian needs assessment processes for the Syria crisis.

KEYWORDS

Data Cleaning, Data Management, Data Weighting, Humanitarian Data, KoBo Toolbox, MIRA, MSNA, Multisector Data Collection, Needs Assessment, Quality Assurance, R-Script, Syria Crisis

1. INTRODUCTION

To provide life-saving humanitarian assistance to the people affected by the crisis, whether natural or man-made disasters, it is widely recognized that availability of timely and accurate information is a foundation for informed decision making. It is evident that the number of people affected by humanitarian crisis has almost doubled in the past decade and is expected to keep rising whereas the cost has more than trebled putting pressure on global humanitarian response system (OCHA and DARA, 2014). Moreover, the number of crises receiving international-led humanitarian assistance almost doubled between 2005 and 2017 whereas the average length of crises is also increasing (OCHA, 2018). On the other hand, the gap between needs and the resources available to provide humanitarian assistance has widened resulting overstretched global humanitarian system (World Humanitarian Summit, 2016). In order to prioritize the limited resources available, it is widely recognized that

DOI: 10.4018/IJISCRAM.2019010105

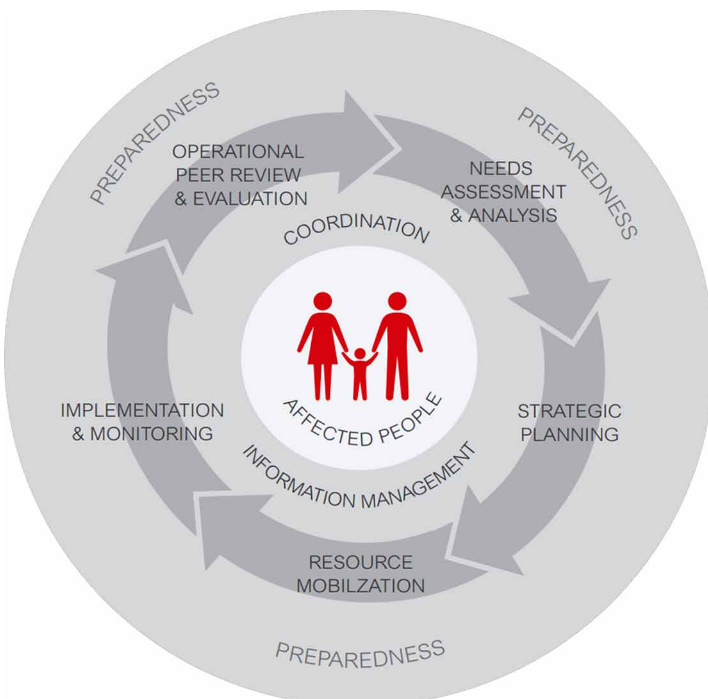
humanitarian agencies require timely and reliable information to understand, define, identify and measure the needs more accurately (World Humanitarian Summit, 2016). Need assessment is regarded as a vital tool to collect data for evidence-based planning and prioritization by aid organizations (Dysvik & Rohatgi, 2017; Inter-Agency Standing Committee (IASC), 2015; OCHA and DARA, 2014).

2. MULTI-SECTOR NEEDS ASSESSMENT (MSNA)

A multi-sector needs assessment (MSNA) is a collaborative effort of the humanitarian community to collect primary data to identify the number of People in Need (PiN), where they are located and quantum and severity of their needs. The MSNA in Syria is based on the Inter Agency Standing Committee (IASC) Multi-Sector Initial Rapid Assessment (MIRA) (Inter-Agency Standing Committee (IASC), 2015) guidelines which were customized and improved for the Syrian context over the years. The MSNA is a key element of the Humanitarian Program Cycle (HPC) that is used to develop the evidence-based Humanitarian Needs Overview (HNO) (Figure 1).

Syrian Arab republic, with nine years of continuous hostilities and conflict, constitutes one of the biggest displacements and protection crises in the world in recent time, with more than 5.6 million refugees and 6.2 million internally displaced people (IDPs). The Inter-Agency Standing Committee (IASC)[1] has classified Syria humanitarian crisis as Level 3 (L3) emergency which is activated in the most complex and challenging humanitarian emergencies when highest level of mobilization of resources is required across the humanitarian system. In such a large-scale humanitarian crisis, the humanitarian system requires notable efforts to gather timely and accurate information for the country-wide analysis of needs and informed prioritization of resources.

Figure 1. Humanitarian Programme Cycle



3. METHODOLOGY FOR MSNA

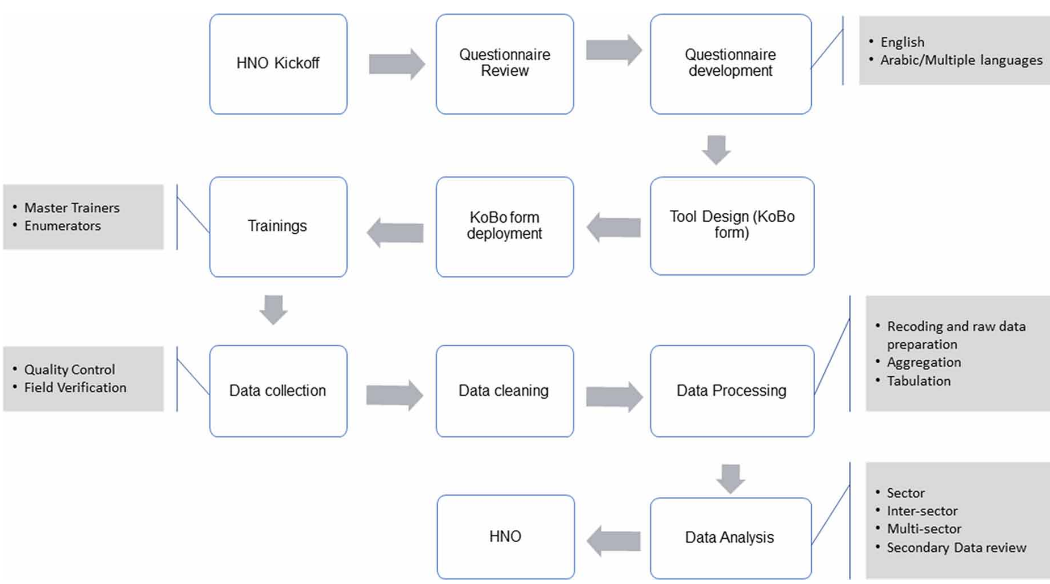
To support effective humanitarian response in different countries like Afghanistan, South Sudan, Yemen and Syrian Arab Republic, the humanitarian coordination system has successfully designed and implemented MSNA to collect primary needs data at the lowest geographic level (i.e. community level) that follows the IASC guidelines and methodology (Inter-Agency Standing Committee (IASC), 2012; Inter-Agency Standing Committee (IASC), 2015).

The (WoA) Afghanistan assessment (OCHA Afghanistan, 2019) is a multi-stakeholder assessment conducted in all 34 provinces of Afghanistan based on a mixed methodology, including 22,142 household interviews in accessible districts of Afghanistan, 1,395 key informant interviews (KIIs) in 70 hard-to-reach (HTR) districts, and 68 focus group discussions (FGDs) (two per province) to assess more sensitive topics (such as Gender Based Violence) and the impact of trauma on family/community dynamics), while simultaneously providing a better understanding of surprising quantitative findings. South Sudan (OCHA South Sudan, 2019) has an under developed communication technology infrastructure and most logistically challenging places in the world. The challenges in providing the concrete data in South Sudan is well known. The humanitarian community used projections from previous years data with wide margins of error; MIRA based data collection; cross-referencing; biometric registrations; village assessment surveys; SMART nutrition surveys and multitude of other methods at higher geographical levels. Whereas the Yemen (OCHA Yemen, 2019) Multicluster location assessment (MCLA) is a coordinated data collection exercise through Key Informants to fill information gaps at district level analysis of needs per population group.

Figure 2 shows the different stages of the MSNA process.

The importance of comprehensive, cross-sectorial, and impartial assessments to enable joint analysis of humanitarian needs has been widely recognized. At the same time, there is a widely recognized tendency that the individual agencies conduct assessments to implement their own programs (World Humanitarian Summit, 2016). The MSNA process in Syria is based on pillars of collaboration, ownership, rigorous planning, technical expertise, trainings, effective coordination, technology and quality information management. The process is aimed at collecting primary data from key informants or community focal points in face-to-face or remote interviews to support the

Figure 2. Assessment (MSNA) process



analysis and preparation of comprehensive overview of humanitarian needs. For the data collection, in-person/face-to-face interviews are prioritized when direct access to the community is possible. Remote interviews are only conducted where direct access is not possible due to security or other reasons. Data collected through MSNA process is expert judgment/perception based on experiences of desirable people or group in communities (i.e. identified as key informants or community focal points) for specific elements/parts of MSNA questionnaire.

4. USE OF TECHNOLOGY

The use of current data is paramount for evidence-based humanitarian programming. To collect timely and accurate primary data, the MSNA became an instrumental tool for the humanitarian community. Traditionally, the data from the field used to be collected on printed paper questionnaire. The data, in the subsequent process, was manually entered into a database or spreadsheet using a rigorous and time-consuming manual data entry efforts to provide data in a structured data format for further analysis. The advancement in technologies and needs in the field have led to the adoption of digital processes for data management. Due to rapidly changing field dynamics and involvement of many entities in Syria, it was crucial to identify and deploy the appropriate tool to successfully and efficiently implement MSNA. Moreover, to ensure the quality of data, it was necessary to consider the efficient implementation of complete quality assurance measures such as error prevention, data monitoring, data cleaning, and documentation (Broeck, Cunningham, Eeckels, & Herbst, 2005) during the entire process. The digital process of data management enabled imbedding such measures appropriately in the process thus reducing problems, improve the validity and reliability of the data (Maduka, Akpan, & Maleghemi, 2017).

A plethora of tools are available to implement a solution for the mobile data collection from the field (NOMAD - Online Selection Tool, n.d.). The Open Data Kit (ODK) based solutions have emerged as the prevailing options for many similar assessments. As an open source platform, ODK has inspired commercial entities and non-profit initiatives to provide ODK based digital data collection and data management solutions. The KoBo toolbox, based on ODK standards, provides a suite of tools that facilitates field data collection in challenging environments (KoBoToolbox, 2018). The KoBo toolbox has been successfully implemented in many crises environment to collect primary data from the field. The toolbox includes an integrated set of tools for building data collection form and collecting interview responses. The toolbox provides features and functionalities that enable both online/offline data collection using mobile and web apps; and submit data to the central online server. It also supports data collection form designing and deployment in multiple languages. On the other hand, the number of trainings were organized over the years to build capacity of the humanitarian community for the use of KoBo toolbox in their field data collection process. These were important consideration for successful data collection in Syria. The reduced training efforts required due to familiarity and experience of using KoBo tool by the partner organizations, free and open source tool built for data collection in the challenging humanitarian settings, the KoBo toolbox became a de-facto tool of choice for the MSNA. To implement the MSNA in Syria, KoBo toolbox and other related tools such as XLSForm, KoBoCollect app, R-script program were used. In the process, the limitations in the available features of KoBo toolbox were also identified and solutions were developed. These are elaborated in the subsequent sections.

5. RAW DATA PREPARATION

5.1. KoBo Form Design and Deployment

The KoBo form design is based on multi-sector needs assessment questionnaire that is specifically developed considering the information gaps to support the analysis of humanitarian needs and priorities. Each sector/cluster developed a set of sector-specific thematic questions to include in the multi-sector data collection tool. The questions were reviewed to ensure alignment with the information need and gap; and consistency across sectors using an iterative and consultative process. The questionnaire was then translated to Arabic language which was the primary language of interview with key informants in the field during data collection.

KoBo toolbox was used for mobile and web data collection from the field. The paper questionnaire was converted to XLSForm[2] in both English and Arabic languages for field data collection. The use of XLSForm to design the data collection form became very useful since the process of iterative revision of the questionnaire was used to finalize the tool. In addition, the use of XLSForm enabled to implement measures to prevent erroneous data entry by using conditional checks, implementing constraints and skip logic in the form design. Once the tool was finalized, it was uploaded to the KoBo toolbox and deployed for data collection.

For Syria MSNA, questionnaire design itself was challenging as it included some 200 questions from multiple sectors. In addition, the MSNA was planned to collect data at the community and neighborhood level which has over 7,655 units. Moreover, a big number of partners were involved for the data collection from different hubs[3]. During the design phase and deployment of the data collection form in the KoBo toolbox, following issues were considered:

- A large number of questions;
- Bilingual questionnaire (English and Arabic languages);
- Community/neighborhood level data collection with more than 7,655 units;
- Sharing of data collection forms with partners collecting data.

The KoBo toolbox features such as multiple languages support and sharing of the deployed projects to other users; were very useful to address some of the challenges. However, due to the size of questionnaire, number of administrative units for data collection, details in the questions and bilingual questionnaire, slow loading of the form and frequent request time-out in the web app were experienced. The user experience was not better in the KoBoCollect app[4] in the mobile which took several minutes to load the form and frequently crashed while navigating through the questions even in the recent high-end android phones which were used for testing. These inherent technical limitations in the KoBo toolbox and related tools were identified and solutions were developed to address these challenges in the process.

In addition, the multiple partners involved were managing their field visits and collecting data from the locations assigned to them[5] independently. The individual partners requested access to their data only to efficiently manage data collection, field planning and progress monitoring process. It was requested that their raw data was not to be shared with other partners before it was properly reviewed and cleaned by the respective partners. While KoBo toolbox has a feature that allows to share a project with another KoBo user account and grant different level of access permissions[6], it is not granular enough to address the data sharing provisions that were needed to manage coordinated MSNA of this scale. For example, the assigned access permission was for all the records in the project while it was not possible to limit the access of subset of raw data submitted by one user to another in the same project.

To address the limitations as mentioned above in the KoBo toolbox, R-script was developed for the following:

- Multiple forms in the Arabic language were prepared for deployment to collect data. It was developed from the main XLSform originally designed in English and Arabic;
- Separate forms were prepared and deployed for each partner for each governorate where they were collecting data.

As a result, it helped in reducing the size of the form ensuing better user experience while collecting and editing survey data. In the process, more than 100 forms were created that added other challenges such as efficiently uploading forms, deploying and sharing with respective partner's KoBo account, and downloading the data. R-Script program[7] was developed utilizing the New REST API feature[8] to automate the whole process of:

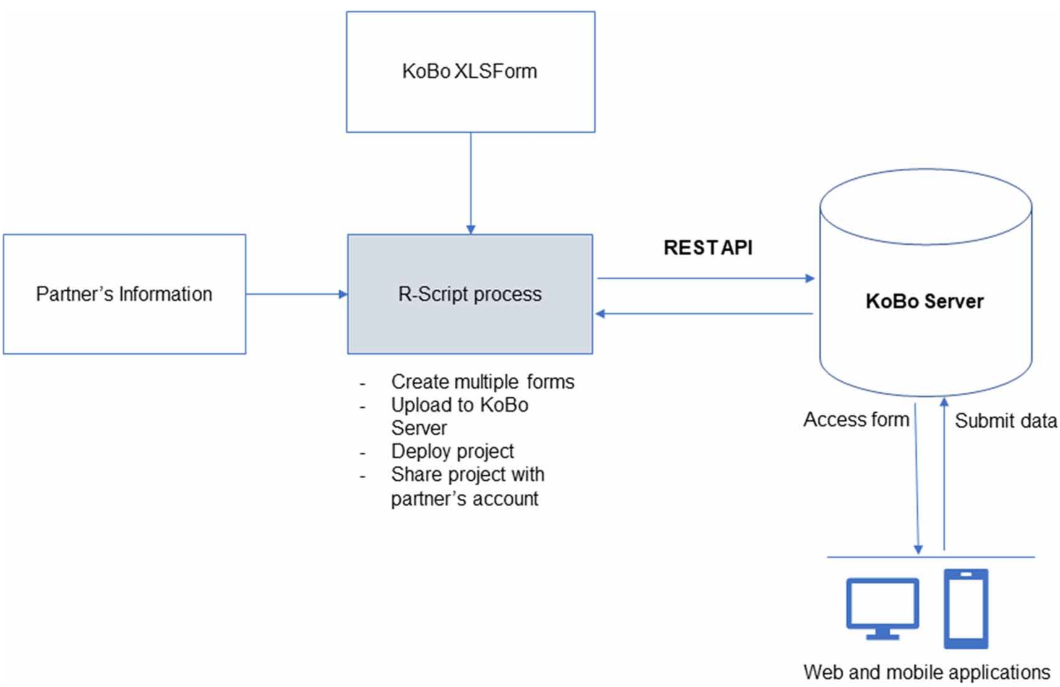
- Creating a form for each partner and governorate in Arabic language;
- Filtering out community/neighborhood names;
- Managing form name to easily identify in the long list of forms;
- Uploading form;
- Deploying project(s);
- Sharing with respective partners with appropriate access permission;
- Downloading data.

The API feature in the KoBo toolbox and use of R-script programming became very useful in automating the whole process. The diagram below shows the stages of R-Script and REST API feature used in the form management process (Figure 3).

5.2. Data Collection and Field Progress Monitoring

For the data collection, individual partner planned and managed field interview process independently following standard process and timeline. The profiles of the list of priority key informants for each sector were provided to enumerators to help them identify key informants and arrange interview. In addition, while designing the data collection XLSForm, some aspects of error prevention measures were implemented during the form design process by using conditional logic checks, constraints in responses and use of skip logic. In some cases, paper-based forms were used during the interview of

Figure 3. Use of REST API available in KoBo toolbox

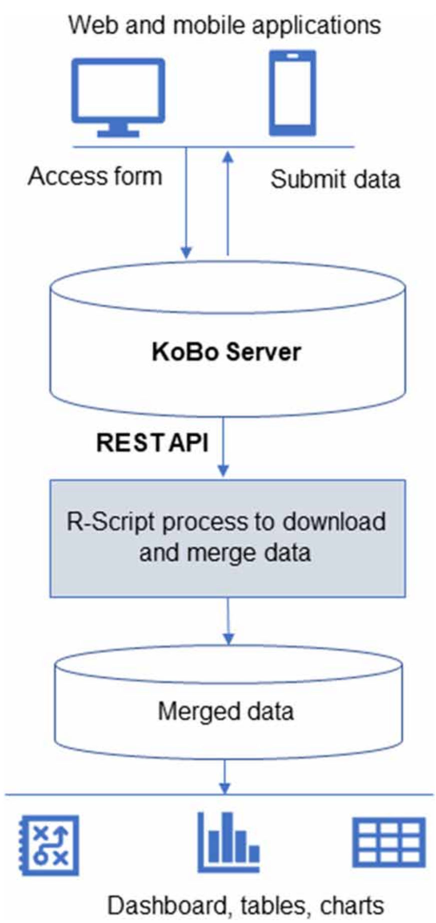


key informants when it was deemed insecure to use digital data collection. The data was later entered to the KoBo system using the mobile or web applications. Moreover, there was a need to monitor progress on few key indicators such as monitoring of gender of enumerator and respondents with the aim of increasing number of interviews of female respondents than rather low previous years records. In addition, monitoring of quality of submission including completeness of the questionnaire, identification of nature of data errors, was important to provide guidance to enumerators for necessary correction and improvements while data was still being collected. This was incorporated as a necessary step to avoid surprises at the end.

The enumerators from the partner used separate projects/forms shared with their respective accounts to collect and submit the survey data. In this way, data was submitted through over 100 forms by multiple partners. This led to a situation where downloading data using normal KoBo process became cumbersome and unmanageable for daily progress monitoring. For the assessment of this nature, it was necessary to routinely download the submitted data and analyse for field progress monitoring, provide early feedback to enumerators and re-allocation of coverage where necessary.

An efficient process was required to achieve the tasks mentioned above. KoBo API feature was used, and R-script program was developed to automatically download data from all the deployed projects (Figure 4).

Figure 4. KoBo data management process



The data from individual projects were merged into a single file. The information related to monitoring indicators was extracted using R-script program for further use. The near real-time dashboard using Google Spreadsheet and Google Data Studio to visualize and monitor the progress of key indicators was developed [9]. In the dashboard information like number of communities covered, the percentage of male and female respondents, the percentage of male and female respondents by sectors, number of communities covered by partners and coverage gaps etc., were included. The dashboard particularly became very useful to provide timely feedback and guidance to individual partners and enumerators.

5.3. Data Cleaning and Raw Data Preparation

Data cleaning is a process which involves screening, diagnosing and editing of the identified errors or abnormalities in the data. While data error prevention measures help to minimize the data errors, it cannot eliminate them. The data cleaning process deals with the data problems once they have been submitted (Broeck, Cunningham, Eeckels, & Herbst, 2005; Aldo Benini, 2013b). No matter how well data is collected, there is always a chance of some level of errors (ACAPS, 2016). Therefore, data cleaning is an important and crucial step in assessment data collection to ensure the quality of data before sharing raw dataset with relevant stakeholders for further analysis. Moreover, it is a challenging process when dealing with data of surveys/assessments involving large number of questions and records (Maduka, Akpan, & Maleghemi, 2017).

For Syria MSNA, the following process was employed:

- Cleaning of the data records by the organizations collecting and entering data for entry errors and logical inconsistencies;
- Reviewing of information recorded as ‘Others’[10];
- Additional cleaning for logical inconsistencies, outliers, completeness and anomalies in the data by a technical team managing the assessment process.

Since multiple partners were collecting data independently, it was important to provide guidance to ensure a consistent approach of error identification and correction used by all the partners. A detailed guideline was developed outlining the process to review data for data entry errors, completeness, logical inconsistencies checks and steps to clean data. Training was organized to partner’s focal persons to build a common understanding of the process and guide them through the process.

The first phase of the data cleaning process (i.e. by partners) was run parallel to data collection and data entry to minimize delays in sharing cleaned raw data. To help identify data entry errors and logical inconsistencies, an Excel-based Add-in was developed for partners to use. The editing in the data was directly done by accessing the data in KoBo web interface. In addition to the cleaning process adopted by partners, R-Script program was developed to download data from partner’s projects, merge all data to a single file and perform additional checks in the data. In case of any issues found in the data, respective partners were notified to check and rectify the issues. The editing of the data records was documented in the data cleaning log.

The second phase of data cleaning was performed once partner notified the completion of data collection, data entry and first round of data cleaning. In this phase, logical inconsistencies, outliers and abnormalities in the data records were checked and edited as necessary following a standard guideline. In addition, information entered as ‘open text’ under other category were reviewed against the existing list of categories. This process was done in the downloaded data file. After completion of data cleaning, some information such as the name of enumerators and other personal identification information was anonymized. The variable codes were recoded to English labels using the KoBo XLSForm as a reference. This process was automated by developing a R-script program. Anonymized and recoded raw data was shared with sectors. It became evident that the KoBo form (data collection form) designed by properly implementing error prevention measures helped to improve the quality of information submission thus less time required for data cleaning.

6. DATA AGGREGATION PROCESS

Multiple partners and initiatives collected data in the populated communities in Syria. The majority of the communities were covered by more than one partners. Each partner assigned to a location managed their data collection process independently with other partners. Therefore, many communities had multiple community-level records available for triangulation. This approach of collecting multiple data points, which represents perspective or understanding of different key informants or community focal persons, for the same community was crucial to maximize the credibility of information (Peersman, 2014). The responses from multiple key informants or community focal persons may or may not indicate the same results or situation of the same community. Therefore, these multiple records per community needed to be triangulated and aggregated to produce one record per community for further analysis. It was necessary to implement a systematic and efficient process of data aggregation to maintain transparency of the methods used, avoid subjective biases in the process and generate reproducible results. This was a vital step in the Syria MSNA to increase credibility and wider buy-in of the information.

For the aggregation of the multiple community-level records into one single record, a systematic approach was developed which mainly involved three steps:

1. XLSForm dictionary preparation;
2. Confidence level and weight calculation;
3. Weighted aggregation of data.

Figure 5 shows the overall workflow of the process. R-script program was developed to automate the data aggregation process.

6.1. XLSForm Dictionary Preparation

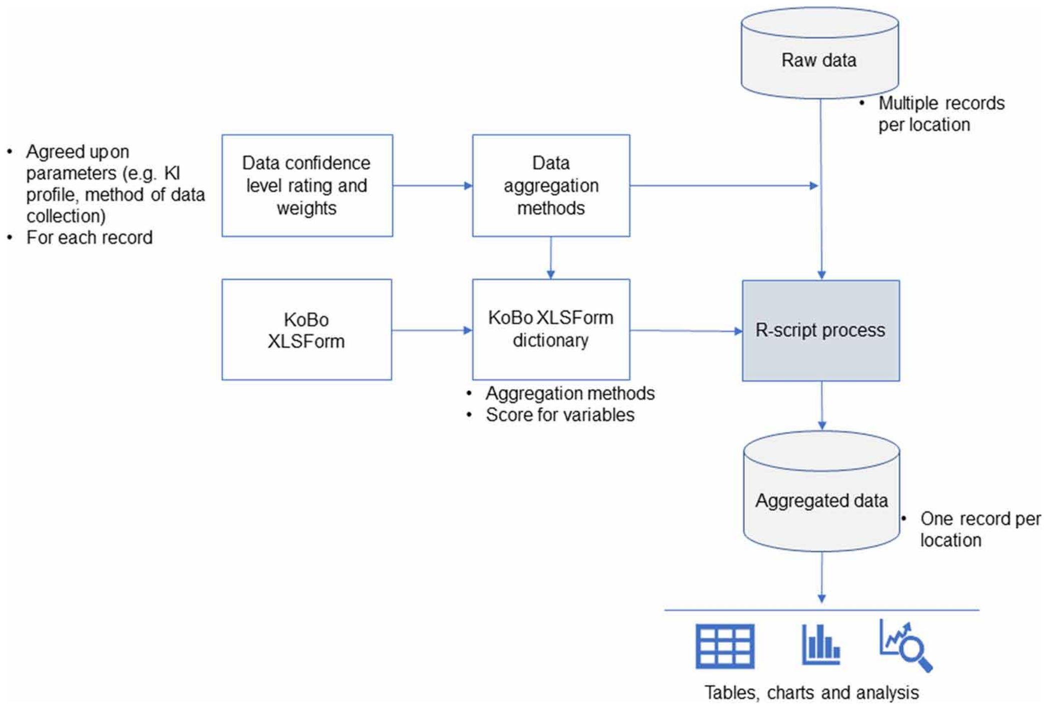
To make the aggregation process iterative and generate reproducible results, the aggregation plan was built within the XLSForm. The ‘survey’ and ‘choices’ sheets[11] of the form were configured with additional information required to feed into the aggregation methods. In the ‘survey’ sheet, information such as the aggregation method, rank score, the group identifier of the questions and sector names were added. Similarly, in the ‘choices’ sheet, additional information like variable type identifier, ordinal score, the weight of the variables was added. This additional information was used in the R-script program to decide the aggregation methods and use related parameters within each method. This approach helped to add additional methods or review existing methods and related parameters as required to produce the correct aggregation plan.

6.2. Confidence Level and Weights

When multiple questionnaires by different partners were submitted for the same community, it was important to measure confidence level of the information so that the responses from the most reliable sources and method is given more weight when triangulating and aggregating data. The confidence level was estimated using the parameters related to the methodology applied for collecting data and the source of information. The confidence level score was used as a weighting parameter for the aggregation.

For the questionnaire development, each sector suggested the list of priority profile of respondent to select for an interview. During the interview, the information related to the method of data collection, type, gender and age of the respondents were recorded for each sector in the metadata

Figure 5. Data aggregation workflow



section. To estimate confidence level, the score was assigned for each parameter as outlined in the Table 1 below.[12] This allowed estimation of separate confidence level score for each data record of the respective sectors.

The sum of the scores was used to evaluate the confidence level for each data record and the composite score was also recorded in the dataset for future reference.

6.3. Aggregation

An overwhelming majority of the questions developed for interview in MSNA were qualitative in nature with a predefined list of answers to select from. It was required to convert the categorical variables (nominal or ordinal in nature) to numerical scores for weighted aggregation which is a common practice in qualitative data analysis (Kitchenham & Pfleeger, 2003). The design of specific data aggregation method was dependent on the type of question and the nature of the responses recorded.

Table 1. Criteria and scores for data weighting

Parameters/Criteria	Options	Score
Key Informant or Community Focal Point (CFP) Role or knowledge in subject matter	Expert	3
	General	2
	Limited	1
Data Collection modality	Face to Face	3
	Remote	1

To develop aggregation methods, first, the type of questions was identified which mostly fall under continuous/numerical, categorical and ordinal categories depending upon the type of answers recorded. Thereafter, the aggregation method was developed considering the nature of each type of question. The logic used to systematically aggregate each type of question present in the Syria MSNA are outlined below. To efficiently run the aggregation process, R-script program was developed.

6.3.1. Numeric/Continuous Data

The questions which capture quantitative information, for example, number of internally displaced people living in the community, the average price of bread or travel time to the nearest health facility etc, falls under this group. To estimate number closer to the answer provided by the most reliable source and method, the weighted average of the answers from different questionnaires was calculated. To aggregate 'n' number of data records in a location to a single record, following expression was used:

$$\text{Weighted Average Calculation} = \frac{\text{Data1 x Score1} + \text{Data2 x Score2} + \ldots + \text{Data n x Score n}}{Score1 + Score2 + \ldots + Scoren}$$

where score 1, score 2 and score n refers to confidence level score for records holding data 1, data 2 and data n respectively. During the process, unknown values (Do not know or No answer), which were recorded as -1 or -5, were recoded to NA (NULL) and excluded from the calculation.

Results were rounded to a nearest integer value if integer result was required.

6.3.2. Categorical Data

In the needs assessment, categorical responses were collected to capture communities' perception of problems, situations, preferences or needs. For each question of this type, a predefined list of options was provided to record answers. The answers were recorded without the order of priority or importance. In the assessment design, the following types of questions were designed to collect categorical responses without any order of preferences or priorities:

- Select one answer;
- Select the top three (or top five) answers;
- Select all relevant answers.

Answers from different questionnaires for the community were aggregated by identifying the answer provided a) most often with b) the highest reliability score.

The following steps were followed to aggregate multiple records into one record per location:

- Create a binary variable for each of the categories of answers;
- Convert YES/NO or TRUE/FALSE to 1 or 0, respectively;
- Multiply 1 or 0 by confidence level score;
- Sum the score for each category and identify the category with highest score;
- Add the category with the highest score as a final result.

For 'select one' answer type question, single answer with highest score was returned as a result. If multiple answers were found to have the same top score, further verification and triangulation to was done to establish the final result for a location. If a decision could not be made between multiple answers, "no consensus could be found" was returned as the final aggregated result. Similarly, for 'select three' answer type question, answers with top three highest score were returned as a result.

On the other hand, for 'select all' relevant answers type questions, all answers were retained in the aggregated dataset.

6.3.3. Categorical Data Where Ranking Was Recorded

In MSNA, multiple questions were designed to capture the answers in the order of priority or importance (i.e. ranked order). During the interview, the enumerator asked the respondent to rank their answers. For the aggregation of ranked data records, Borda count method outlined in references (Aldo Benini, 2013a) and (Minu Limbu; Leticia Wanyagi; Berryl Ondiek; Benoit Munsch; Kioko Kiilue, 2015), which is derived from the method used in election systems, was used.

Additionally, during the KoBo form design, answer to each rank was captured in a separate 'select one' type question. For example, three "select one" questions were used to capture first, second and third-ranked responses. During the pre-processing stage, data entered in multiple single select responses were grouped together, transformed and the appropriate rank score was assigned as shown in the Table 2. The process is elaborated in the subsequent section.

Following steps were used to aggregate multiple records into one record per location:

- Scoring the ranked answers – assign score 3 to the first rank, 2 to the second rank and 1 to the third ranked options (in situation of three ranked options);
- Multiply the rank score by confidence level score;
- Sum the score and identify the categories with the top three highest total scores (in case of rank three question);
- Add these top three categories as the result together with associated rank;
- If multiple categories were found to have same highest score within the 3 highest ranked scores, all were retained. Further verification and triangulation were done to establish the final result for the location (Table 3).

6.3.4. Ordinal Data

In the assessment questionnaire, ordinal variables were used mainly to gather interval estimates of counts, percentage intervals and preference or priority measures for problems or needs. For example, 1-25% of total children not attending school in the community.

Table 2. Data transformed to store ranked response (1 = first ranked, 2 = second ranked, 3 = third ranked)

Data Record				Restructured Data With Rank				
Location	Rank 1	Rank 2	Rank 3	Location	Variable 1	Variable 2	Variable 3	Variable 4
A	Variable 1	Variable 4	Variable 3	A	1	NA	3	2
A	Variable 3	Variable 1	Variable 4	A	2	NA	1	3

Table 3. Aggregated result

Location	Confidence Level Score	Variable 1	Variable 2	Variable 3	Variable 4
A	6	1=>3	NA	3=>1	2=>2
A	5	2=>2	NA	1=>3	3=>1
Total score		3x6+2x5 =28	NA	1x6+3x5 =21	2x6+1x5 =17
Result rank		1	NA	2	3
		First rank	NA	Second rank	Third rank

Answers from the different questionnaires were aggregated by giving more weight to answers provided with a higher reliability. It is a common practice to convert the ordinal responses to numeric scores and analyze the data as numeric values (Kitchenham & Pfleeger, 2003).

The following steps were followed to aggregate multiple records into one record per location:

- Recode the ordinal categories into 1, 2, 3, 4 scores etc.;
- Multiply ordinal score (e.g. 1, 2, 3, 4) by the corresponding confidence level score and sum the scores;
- Divide the sum of scores by the total sum of confidence level score to obtain a weighted average score;
- Recode average score back into the corresponding ordinal category.

Following expression elaborates the calculation method for aggregation of 'n' number of records to one record:

$$\text{Weighted Category Score} = \frac{\text{OC1 x Score1} + \text{OC2 x Score2} + \ldots + \text{OC n x Score n}}{Score1 + Score2 + \ldots + Score n}$$

where 'OC' refers to ordinal category score and 'Score' refers to confidence level score defined in section 6.2. If the weighted score falls between 2 ordinal categories, worst-case scenario was adopted. The worst-case scenario for the relevant questions was defined in consultation with sectors.

7. CONCLUSION

Humanitarian crisis across the globe is affecting more people than before. It is becoming imperative to find timely and quality information to address the needs of people. The types of crisis and dynamics of situation is making data collection a challenging process. As a result, the humanitarian community is addressing the challenges by combining innovations in use of technology with technical capacity. The MSNA in Syria has been a testing ground of best practices in data management and use of technology to achieve better results by the humanitarian community. KoBo toolbox, XLSForm, KoBo Collect, R-script and API features etc are some of the examples of the use of technology that is referred to in the paper. The aggregation discussed in the paper provides an exhaustive detail about triangulation of data using mathematical models. This was needed to address aggregation of multiple data points collected at same location by different partners to ensure transparency and avoid subjective judgement. This also helped to implement different methods of aggregation depending on type of data collected for different groups of questions. The use of technology helped to manage the complex process efficiently and with good quality. The analysis of processes from use of technology, data collection, cleaning and data analysis reveals that there is room for customization of these process. The evolution of these processes for Syria took few years and they can be adapted for other assessments by humanitarian, social and development sectors. The MSNA process is reviewed with a lens of technical implementation. The reason behind the technical lens is to elaborate the process for readers to systematically replicate this complicated process. It helped, in increasing efficiency; doing business easily and systematically; brought in transparency to the processes; and generated reproducible results.

REFERENCES

ACAPS. (2016). Data Cleaning. Retrieved from https://www.acaps.org/sites/acaps/files/resources/files/acaps_technical_brief_data_cleaning_april_2016_0.pdf

OCHA Afghanistan. (2019). Reliefweb. Retrieved from https://reliefweb.int/sites/reliefweb.int/files/resources/afg_2019_humanitarian_needs_overview.pdf

Benini, A. (2013a). Severity and priority - Their measurement in rapid needs assessments. Geneva: ACAPS. Retrieved from http://aldo-benini.org/Level2/HumanitData/Benini_forACAPS_SeverityAndPriority_2013.pdf

Benini, A. (2013b). Technical Brief - How to approach a dataset. ACAPS. Retrieved from https://www.acaps.org/sites/acaps/files/resources/files/how_to_approach_a_dataset-part_3_analysis_march_2013.pdf

Broeck, J. V., Cunningham, S. A., Eeckels, R., & Herbst, K. (2005). Data Cleaning: Detecting, Diagnosing, and Editing Data Abnormalities. Retrieved from https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0020267

Dysvik, P. C., & Rohatgi, M. A. (2017). *Needs assessment; a duality in humanitarian practice? How does current needs assessment practices facilitate information-sharing among humanitarian actors?* Universitetet I Stavanger.

Inter-Agency Standing Committee (IASC). (2012). Operational Guidance for Coordinated Assessments in Humanitarian Crises. Inter-Agency Standing Committee (IASC). Retrieved from https://www.humanitarianresponse.info/sites/www.humanitarianresponse.info/files/documents/files/ops_guidance_finalversion2012_1.pdf

Inter-Agency Standing Committee (IASC). (2015). Multi-Sector Initial Rapid Assessment Guidance - Revision July 2015. Inter-Agency Standing Committee (IASC). Retrieved from https://www.humanitarianresponse.info/sites/www.humanitarianresponse.info/files/documents/files/mira_revised_2015_en_1.pdf

Kitchenham, B., & Pfleeger, S. L. (2003). Principles of Survey Research, Part 6: Data Analysis. Retrieved from http://www.inf.puc-rio.br/~inf2007/docs/artigos/Principles%20of%20Survey%20Research.pdf

KoBoToolbox. (2018). Retrieved from KoBoToolbox: https://www.kobotoolbox.org/

Limbu, M., Wanyagi, L., Ondiek, B., Munsch, B., & Kiilue, K. (2015). Kenya Inter-Agency Rapid Assessment Mechanism (KIRA): A Bottom-up Humanitarian Innovation from Africa. *Procedia Engineering*, *107*, 59-72. Retrieved from https://www.sciencedirect.com/science/article/pii/S1877705815010127

Maduka, O., Akpan, G., & Maleghemi, S. (2017). Using Android and Open Data Kit Technology in Data Management for Research in Resource-Limited Settings in the Niger Delta Region of Nigeria: Cross-Sectional Household Survey. *JMIR mHealth and uHealth*, *5*(11), e171. doi:10.2196/mhealth.7827 PMID:29191798

NOMAD - Online Selection Tool. (n.d.). Humanitarian Operations Mobile Acquisition of Data (NOMAD). Retrieved from https://humanitarian-nomad.org/online-selection-tool

OCHA. (2018). *World Humanitarian Data and Trends 2018*. OCHA.

OCHA. (2019). Global Humanitarian Overview 2019. Retrieved from https://www.humanitarianresponse.info/sites/www.humanitarianresponse.info/files/documents/files/gho-2019-web-en.pdf

OCHA and DARA. (2014). *Saving lives today and tomorrow - Managing the risk of humanitarian crisis*. OCHA Policy and Studies Series.

Peersman, G. (2014). Overview: Data Collection and Analysis Methods in Impact Evaluation. UNICEF Office of Research, Florence. UNICEF. Retrieved from https://www.unicef-irc.org/publications/pdf/brief_10_data_collection_analysis_eng.pdf

OCHA South Sudan. (2019). Reliefweb. Retrieved from https://reliefweb.int/report/south-sudan/2019-south-sudan-humanitarian-needs-overview

World Humanitarian Summit. (2016). Too important to fail - addressing the humanitarian financing gap. ReliefWeb. Retrieved from https://reliefweb.int/sites/reliefweb.int/files/resources/%5BHLP%20Report%5D%20Too%20important%20to%20fail%E2%80%94addressing%20the%20humanitarian%20financing%20gap.pdf

OCHA Yemen. (2019). Reliefweb. Retrieved from https://reliefweb.int/report/yemen/yemen-2019-humanitarian-needs-overview-enar

ENDNOTES

1 IASC - The Inter-Agency Standing Committee (IASC) is the primary mechanism for inter-agency coordination of humanitarian assistance. https://interagencystandingcommittee.org/

2 XLSForm is a standard created to simplify the development of data collection forms in Excel.

3 Syria Humanitarian operation hubs includes Turkey, Syria, Lebanon, Jordan and Iraq

4 KoBoCollect is a mobile data collection application which can access forms and submit responses to the KoBo toolbox.

5 Multiple partners were assigned to same location wherever available during the coverage assignment phase.

6 KoBo project sharing permissions are View form, Edit form, View submissions, Add submissions, Edit submissions and Validate submissions.

7 R is a programming language and environment for statistical computing and graphics https://www.r-project.org/about.html

8 KoBo API reference https://kobo.humanitarianresponse.info/assets/

9 Dashboard example: https://goo.gl/2kwtaH

10 In the questions where pre-populated list of answers was provided, enumerators were able to enter additional answer in the 'other' field.

11 The XLSForm usually includes three sheets named survey, choices and settings to build the data collection form.

12 A general framework of data weighting system for HNO 2017 was developed by "Weighting work stream" of HNO process.

Punya Prasad Sapkota, a national of Nepal, has extensive experience in information management, GIS and data analytics in humanitarian and development sectors. He has worked in Nepal, Kyrgyzstan, Afghanistan, and currently working with the United Nations OCHA Turkey. There he has been working on providing information management services for the humanitarian response to the Syria Crisis since 2013. He has an interest in providing innovative information management and GIS solutions to address complex issues.

Kashif Siddiqi, a national of Pakistan, has extensive experience in development and humanitarian sector and worked in Pakistan, Yemen, Philippines, Jordan, Central Asia, South Sudan and specifically on Syria crisis (with OCHA Turkey) since 2012. The author published research papers and technical guidelines on the use of geo-information technologies for disaster management and humanitarian response and publishes in peer-reviewed journals and international conferences. Institutionalizing change in organizations, transforming processes and managing programs are his key strengths for over two decades and has worked with academia, public, private and humanitarian sectors.