The World Food Programme's real-time monitoring systems:

Approaches and Methodologies

16 December 2021

The World Food Programme's real-time monitoring systems: Approaches and Methodologies

I. Context

In 2013, leveraging the global growth of cellular and internet technologies, the World Food Programme (WFP) established the mobile Vulnerability Analysis and Mapping project, or mVAM: an umbrella programme which explored the use of mobile phone technologies in the context of monitoring food security in areas with restricted humanitarian access.

With the success of the pilot surveys in refugee camps in the Democratic Republic of Congo, the mVAM programme grew exponentially and organically as demand for these technologies increased throughout WFP and the humanitarian community. The emergence of new global health threats such as Ebola Virus Disease (EVD) and prolonged conflicts and instability in the Middle East and parts of the Sahel further incentivized and provided new opportunities to test these data collection modalities. In fact, demand was so high that more than 100,000 surveys were administered by 2015.

As the popularity of mVAM increased, WFP invested a significant amount of time and resources validating remote data collection methodologies for household food security surveys. While there are multiple studies demonstrating the feasibility and validity of computer-assisted telephone interviewing (CATI) surveys for poverty and other socioeconomic indicators[1] the collection of food security information via phone surveys had not been thoroughly explored. As such, it was necessary to conduct robust mode experiments to understand the feasibility and validity of using CATI modalities to collect internationally recognized food security indicators, including most notably the Food Consumption Score (FCS), the Reduced Coping Strategies Index (rCSI), the Minimum Dietary Diversity for Women (WDD-W) and the Household Dietary Diversity Score (HDDS). Mode experiments were conducted in Kenya, Democratic Republic of Congo, Zimbabwe, Mozambique, South Sudan and in Mali. Each of these studies demonstrated the feasibility of collecting these estimates via CATI surveys. Results are shown in *Table 1*.

In the validation study conducted in Kenya, the magnitude of measured differences in women's dietary diversity scores between the face-to-face survey and CATI was small enough to make CATI a valid mode of data collection. The same study also showed that while women with mobile phones were generally more affluent than those without (in terms of assets, education levels and household head's employment status), no significant differences were observed in the dietary diversity scores of women in either category.[2]

[1] Dabalen A, Etang A, Hoogeveen J, Mushi E, Schipper Y, von Engelhardt J (2016) International Bank for Reconstruction and Development/ The World Bank Group. Mobile Phone Panel Surveys in Developing Countries.
https://openknowledge.worldbank.org/bitstream/handle/10986/24595/9781464809040.pdf?sequence=2&isAllowed=y

[2] Lamanna C, Hachhethu K, Chesterman S, Singhal G, Mwongela B, Ng'endo M, et al. (2019) Strengths and limitations of computer assisted telephone interviews (CATI) for nutrition data collection in rural Kenya. PLoS ONE 14(1): e0210050.

Table 1

Country	Year	Indicator	Modality	Estimates Modality 1	Estimates Modality 2	Findings	Links
DRC	2014	FCS	SMS vs F2F	SMS (February 2014) Poor (%): 57.5 Borderline (%): 37.7 Acceptable (%): 4.7% SMS (March 2014) Poor (%): 73.5 Borderline (%): 23.5 Acceptable (%): 2.9%SMS (March 2014) SMS (April 2014) Poor (%): 70.2 Borderline (%): 26.1 Acceptable (%): 3.7 SMS (May 2014) Poor (%): 61.6 Borderline (%): 32.9 Acceptable (%): 5.5	F2F (Nov 2013) Poor (%): 33 Borderline (%): 50 Acceptable (%): 17	The results of the monthly survey rounds carried out show an immediate deterioration of food security indicators when food assistance was halted, and also captured typical seasonal variations well in DRC.	http://www.cdacnetwork.org/i/20140722100657/jr29x/Click
DRC	2014	FCS	IVR vs SMS	IVR - Median rCSI: 21 Median rCSI (HHs led by men): 23 Median rCSI (HHs led by women): 19 Median rCSI (Flush toilet): 19 Median rCSI (Pit Latrine): 22 Median rCSI (Blair Pit Latrine): 18 Median rCSI (No Toilet): 24	SMS - Median rCSI: 15 Median rCSI (HHs led by men):14 Median rCSI (HHs led by women): 18 Median rCSI (Flush toilet): 11 Median rCSI (Pit Latrine): 17 Median rCSI (Blair Pit Latrine): 19 Median rCSI (No Toilet): 28	The SMS data was weighted by province populations, while the IVR data is unweighted. Thus, the absolute values of indicators cannot be compared. However, we look at the trends observed in both data sets to see whether using different modes for data collection give vastly different results. We find that responses from both IVR and SMS surveys show similar trends in rCSI where households headed by women were worse off (higher rCSI) compared to those headed by men. For rCSI by toilet type, both SMS and IVR follow a similar, expected, pattern where the economically worst-off households (without toilets) have the highest rCSI and the better-off households (with flush toilets) have much lower rCSI.	https://mvam.org/2016/12/15/mind-the-mode-ivr-vs-sms-in-zimbabwe/
Zimbabwe	2016	rCSI	CATI vs F2F	CATI R1 - Mean FCS: 45 P+B: 20% R2 - Mean FCS: 34 P+B: 51%	F2F R1 - Mean FCS: 34 P+B: 52% R2 - Mean FCS: 32 P+B: 56%	Observed substantial differences in terms of FCS between CATI and F2F due to the mode effect in round 1 but vanished in round 2. In round 2, the difference was not statistically significant. In fact, the slight remaining difference between the two groups (voice and F2F) was due to respondent households' socio-economic profile, not to the mode used to collect data.	https://mvam.org/2017/01/26/mind-the-mode-2-setting-the-food-consumption-score-in-south-sudan/
South Sudan	2016	FCS	CATI vs F2F (list and open-based method)	CATI: HDDS: Mean 6.39	F2F: HDDS: Mean 5.02	Open vs List method: There is not a significant difference between HDDS estimated using the "open" approach and the "list" approach. We would have expected a higher HDDS with the open approach as it really pushes respondents to recall their food consumption more accurately but according to the tests we run the two approaches are giving us the same results. CATI vs F2F: When comparing the results between the CATI and F2F surveys, we observed sample populations with significant differences, from the gender of the respondents to their primary income sources and shocks. The HDDS scores measured in the two samples also differ: HDDS is higher by 1.3 food groups for CATI compared to F2F, showing better access to diverse foods among CATI households. We can deduce that the higher HDDS score observed for the CATI survey reflects the generally better-off sample reached compared to those reached via the F2F population sample.	https://wfp.sharepoint.com/ww/fr/sites/mVAM571/Layouts/15/Doc.aspx?sourcedoc=%7B108B3E97-859A-4184-8CE8-DD43BD569211%7D&file=Mozambique-HDDS-test_03122017%20(3).docx&action=default&mobileredirect=true
Mozambique	2017	HDDS	CATI vs F2F	CATI: Mean FCS: 56.21 Poor (%):11 Borderline (%):17 Acceptable (%):71	F2F (resp who took part to the CATI SURVEY): Mean FCS:55.65 Poor (%):6 Borderline (%):17 Acceptable (%):77	Food Consumption Scores (FCS) collected via CATI tended to be slightly higher than those collected via F2F. However, when comparing the same pool of respondents who participated in both the F2F and telephone survey rounds, their food security indicators are more or less the same (Mean CATI FCS was 56.21 while the mean F2F FCS was 55.65, with no statistically significant difference between the two).	https://mvam.org/2017/06/08/mind-the-mode-and-the-non-response/
Mali	2017	FCS	CATI vs F2F	CATI: MDD-W(%): 26.4 MAD (%): 29.6	F2F: MDD-W(%): 24.9 MAD (%): 12.3	The differences in women's dietary diversity scores between face-to-face survey and CATI was small enough to make CATI a valid mode of data collection. The same study also showed no significant differences in the dietary diversity scores were observed among women with and without phones. However, for MAD, the magnitude of differences in score with mode was large, meaning further work is required to test the validity of using CATI for collecting and reporting point estimates on MAD.	https://doi.org/10.1371/journal.pone.0210050
Kenya	2018	MDDW-MAD		SMS (February 2014) Poor (%): 57.5 Borderline (%): 37.7 Acceptable (%): 4.7% SMS (March 2014) Poor (%): 73.5 Borderline (%): 23.5 Acceptable (%): 2.9%SMS (March 2014) SMS (April 2014) Poor (%): 70.2 Borderline (%): 26.1 Acceptable (%): 3.7 SMS (May 2014) Poor (%): 61.6 Borderline (%): 32.9 Acceptable (%): 5.5	F2F (Nov 2013) Poor (%): 33 Borderline (%): 50 Acceptable (%): 17	The results of the monthly survey rounds carried out show an immediate deterioration of food security indicators when food assistance was halted, and also captured typical seasonal variations well in DRC.	http://www.cdacnetwork.org/i/20140722100657/jr29x/

With CATI approaches validated for the collection of food security indicators, WFP expanded remote monitoring capacities significantly, creating robust high frequency surveys in many of WFP's largest operations, including in countries such as Yemen, Nigeria and Syria. These high frequency surveys proved to be invaluable in that they provided data in places that due to security reasons would otherwise not be easily accessible. Ad hoc surveys also continued in other countries as well, with mVAM approaches and methods adopted in more than 40 countries by 2018.

In 2018, however, WFP began shifting from the approach of high frequency surveys towards a model of continuous, real-time monitoring or tracking systems. The initial impetus for this transition was the realization that the dynamism of the food security situation in certain countries could not be fully captured even by monthly high frequency surveys. WFP recognized the need for a greater understanding of the drivers of food insecurity and the crucial ability to provide a more effective early warning platform that could trigger early action while mitigating the risks and impacts on communities and effectively responding to future similar shocks.

Furthermore, the transition from high frequency surveys to continuous, real-time monitoring systems also came at no additional costs. In fact, by transitioning to continuous, more systematized data collection, WFP was able to reduce the costs associated with high frequency surveys. This was possible by introducing more automation into systems, reducing the need for manual interventions, and thus reducing the human resources needed to manage, analyze and report on the findings from the surveys. The transition to real-time monitoring systems was a win-win- allowing WFP to gain a greater and more nuanced understanding of food insecurity, its drivers and the impacts of specific interventions while reducing financial strain on country operations.

The establishment of real-time monitoring systems also ensured that WFP had a call centre or a platform for data collection activities that was continuously in operation. This provided immense flexibility to country operations to immediately diversify or scale up in the event of significant shocks, natural disasters, or other important shifts within a country- meaning the type or amount of data collected and analysed could change in days rather than weeks or months. This was most evident at the onset of the COVID-19 pandemic, in which WFP country operations were able to include certain modules in their data collection in the days after the pandemic was announced. This enabled WFP, within days, to understand in real-time the food security effects of COVID-19 and associated government mitigation efforts. Likewise, as face to face data collection activities became more difficult, the platforms (i.e call centres) established for real-time monitoring systems were leveraged by WFP and multiple partners, to conduct other data collection activities seamlessly, supporting activities which spanned from operational monitoring to surveys supporting the Integrated Food Security Phase Classification (IPC) system.

The following document provides descriptions as to how WFP has established these systems, the methodologies and technologies employed and the processes by which WFP ensures the quality of the information collected, whether it be during the initial stages of validation at country level or the continuous quality assurance monitoring that extends the entirety of the life cycle for real-time monitoring systems.

II. How are real-time monitoring systems designed and implemented?

WFP's real-time monitoring systems are designed in the same manner as traditional surveys, following the key steps required in any data collection exercise, including:

- Survey design and planning
- Training and testing of the enumerators/ operators
- Survey management
    - Data cleaning and quality assurance
    - Analyzing and reporting results
    - Data validation at country level

The continuous nature of these systems alongside the CATI modality utilized requires certain design nuances and intricacies, as well as additional quality assurance processes to be taken into account in the design and survey management phases. Certain processes that typically occur one time in a stand-alone survey are repeated multiple times for real-time monitoring systems. These additional requirements will be highlighted in each of the sections below.

Survey design and planning

In the design of real-time monitoring systems, the first considerations are quite basic. What is the purpose of establishing this system? Is such a system needed in a particular country? Is it feasible? Once the answers to these questions are clear, then it's necessary to define the questionnaire and ultimately the sample size required given the strata needed and the analysis plan envisaged.

Purpose and feasibility: Real-time monitoring systems are established ultimately to "keep the finger on the pulse" of certain key indicators which are early indications of a potential deterioration in a country's food security situation. The aim is typically to monitor the situation at sub national level and often at the first administrative area, meaning in most cases the states, regions, provinces, or governorates. [3] Thus, in essence, these systems track the following household level indicators of consumption and coping[4];

- Food Consumption Score / the prevalence of insufficient food consumption
- Reduced Coping Strategies Index / the prevalence of crisis or above crisis food-based coping
- Household Hunger Scale / the prevalence with the moderate to severe hunger
- Livelihood Coping Strategies / the prevalence with crisis or emergency livelihood coping

These indicators represent many of the outcome level 1 indicators in the IPC analytical framework. The IPC analytical framework takes into account two levels of outcome indicators, with outcome level 1 indicators (as shown above) dealing with changes at the household level that would be indicative of food insecurity. IPC level 2 indicators, on the other hand, measure the extent to which the health and nutrition of communities is deteriorating due to food stresses, which often are outcomes that only manifest in the weeks and months after changes in household consumption and coping are observed. Therefore, real-time monitoring systems aim to warn as quickly as possible on deteriorations in household consumption and coping behaviors that may be significant enough to trigger widespread health and nutrition impacts.

As the purpose of these systems is to serve as early warning platforms and triggers for early action, countries that need real-time monitoring systems should be countries prone to food crises due to natural or man-made emergencies. This is often defined by international consensus on countries prone to food crisis which are

[3] There are certain countries where monitoring is conducted at the second administrative area, these include: Malawi and Sierra Leone.
[4] WFP VAM Resource Centre. Food Consumption Score (2019) https://resources.vam.wfp.org/data-analysis/quantitative/food-security/food-consumption-score. Reduced Coping Strategies Index (2019) https://resources.vam.wfp.org/data-analysis/quantitative/food-security/reduced-coping-strategies-index. Household Hunger Scale (2019) https://resources.vam.wfp.org/data-analysis/quantitative/food-security/household-hunger-scale-hhs-indicator-definition-and-measurement-guide. Livelihood Coping Strategies-Food Security 2021 https://resources.vam.wfp.org/data-analysis/quantitative/food-security/livelihood-coping-strategies-food-security

included in the Global Report on Food Crisis. It is, however, also defined by internal WFP data in certain instances.

Feasibility is defined according to several factors; 1) whether WFP has a presence in a country 2) whether there are partners in place to support CATI services; and 3) the level of mobile phone ownership and network connectivity across the country. WFP presence is very important in order to ensure that country level validation is possible. The importance of robust country context when validating data cannot be overstated. The lack of a CATI service provider, on the other hand, can be overcome but it would be more difficult to establish systems in these cases. Mobile phone ownership levels as well as network connectivity remain the biggest obstacles. Although WFP does not have a national phone ownership threshold in place to determine eligibility for real-time monitoring systems, it is necessary to rely heavily on in-country expertise to determine the feasibility in certain low coverage countries. In practice this means that WFP has not established systems in South Sudan, for instance, because it's not believed to be possible or reliable anywhere in the country. In other cases, in the Central African Republic for instance, real-time monitoring systems are only implemented and deemed feasible in certain parts of the country.

Sampling: After defining the purpose of the survey, the next step is to determine the sample sizes both nationally and per strata that would be required to report estimates for the key indicators. Taking into account the modality, real-time monitoring systems adhere to the guidance in the IPC technical manual that state the following:

- At least 150 households per strata is required for a good level of reliability (in countries with more than 75% of households owning at least one operating phone)
- At least 90 observations per strata is required for a good but limited level of data reliability (in countries with more than 60% of households owning at least one operating phone)

In most cases, WFP aims to adhere to the highest levels of reliability, requiring as close to 150 households per strata as possible. There are however cases when it's necessary to fall back on the good but limited level of reliability due to financial or practical constraints. This is, however, the exception rather than the rule. In practice, sample sizes tend to be as close as possible to 150 households per 1st administrative level, with sample sizes ranging from 1,200 households per country per analysis window to over 4,000 households per country per analysis window. Within strata, sample sizes are determined via Probability Proportion to Size (PPS) sampling, with quotas generated for 2nd administrative levels as well to ensure that the results are representative at the strata level.

As real-time monitoring systems utilize CATI modalities, there are some additional complexities that must be addressed. First, as CATI surveys are via phone, it's important to ensure that real-time monitoring systems have as representative a sample as possible, which means it's important to know the geographic location of the households surveyed to the extent possible. With shifting analysis windows (more on analysis windows below), this is even more important as every day represents a new survey that requires a specific geographic dispersal of the sample.

How do we do this? Let's start with how real-time monitoring systems reach households in the first place. Typically, there are several ways to reach households via CATI systems, these include:

- Random-Digit Dialing (RDD)[5]
- MNO, Government, WFP database of geo referenced phone numbers

In almost all cases, real-time monitoring systems initially utilize RDD as the preferred method of reaching households. This is done to ensure that the most random selection of respondents as possible are reached in the first stages of the real-time monitoring systems. Importantly, in this initial process, certain filters are used in the RDD process itself to ensure the geographic and socio demographic distributions needed are respected, meaning that not all households reached are actually interviewed, rather interviews are only

[5] Wolter K, Chowdhury S, Kelly J. (2009). Handbook of Statistics. Vol 29, Part A, pgs 125-154. Chapter 7- Design, conduct and analysis of random-digit dialing surveys. https://www.sciencedirect.com/science/article/abs/pii/S0169716108000072

conducted amongst those that are located in the areas and match the specific characteristics needed. This enables WFP, over time, to build a sample of households that match to the extent possible the specific characteristics of the population in the country as a whole.

With a representative sample constructed via RDD, WFP can transition to panel surveys, the preferred methodology for real-time monitoring systems after the initial months of implementation. A panel survey considering the CATI modality means that households meet certain inclusion criteria based on a pre-targeted set of variables, including where they live and certain socio-economic classifications. The paneled households comprise 80 per cent of the sample and are followed for 8-10 months. The additional 20 per cent of the sample is contacted via RDD and then included in the panel thereafter if the respondents agree.

The analysis window: Real-time monitoring systems usually operate on analysis windows of 30 or 60 days. What this means is that every indicator value reported is an estimate based on either a 30 day or 60 day sliding period of data collection. To illustrate this, when implementing new real-time monitoring systems that operate on a 30-day analysis period, the first estimates will be reported after the 30th day of data collection based on data from the previous 30 days. New estimates are then reportable daily by adding a day of data collection for consideration and dropping a previous day. It works the same way for countries that operate on a 60-day analysis window.

Questionnaire design: In designing the questionnaire for the real-time monitoring systems, WFP follows the optimal survey design guidelines, as well as the best practices for phone-based survey. First, the questions are designed in a very simple and clear manner so respondents can easily understand the questions. Whenever necessary, notes are inserted in the questionnaire script to guide operators with probing and to provide further explanations as needed. Second, the entire questionnaire is translated to local languages to ensure all operators follow a consistent way of asking the questions in the respective languages used. Third, following the guideline for phone surveys, the questionnaires are relatively short (at most 30 minutes) to avoid respondent fatigue and the risk of respondents dropping out of the survey.

The content of the questionnaire includes the same core modules as the traditional face-to-face food security surveys and are designed together with country experts, meaning that the standard modules are adapted for country-context. The standard questionnaire includes modules on:

1. Introduction and consent process
2. Socio-demographic questions such as education levels, age, location
3. Key IPC outcome level 1 indicators: Food Consumption Score (FCS), reduced Coping Strategy Index (rCSI) and/or Livelihood Coping Strategy Index (LCSI)
4. Other context-specific questions such as remittances, income source, access to basic services

Once the questionnaire is properly programmed in the data collection software, it is tested multiple times to ensure that the questions and the skip patterns are programmed accurately. This process repeats each time the questionnaire is revised.

Training of operators: Training of call centre operators for real-time monitoring systems also follows the same structure as traditional enumerator trainings, with extra focus on the use of the data management software. The training session is divided into two phases and conducted by content experts:

1. Training on the software and data collection tool: Typically conducted by the owners of the software which are usually the service providers themselves
2. Training on survey methods and the questionnaire administration: Conducted by WFP food security experts at the country level

The two phases of training are intensive and interactive, usually lasting 3-4 days. The training on the software and data collection tools are shorter than the training on the content of the survey itself. WFP experts thoroughly outline WFP's mission and mandate, the purpose of the survey itself, the survey process and the concepts, questions and definitions needed to fully understand the questionnaire. The content of the questionnaire is reviewed using a question-by-question approach to ensure that operators fully understand

what information each question is trying to capture and can deal with any misunderstandings that may arise among the respondents. Apart from this, WFP also spends time reviewing the ethical considerations, stressing the need for the respondents to provide consent prior to conducting the interview.

After the conclusion of the training, mock/practice interviews are conducted to allow the operators to practice administering the questionnaire and to familiarize themselves with a variety of situations likely to be encountered.

Pre-testing of operators: Once the initial training is completed, WFP conducts pre-test surveys before starting actual data collection. Pre-testing is essential to test if the respondents can understand and answer the questions and if the operators are able to properly administer the questions.

As part of the pre-test for real-time monitoring systems, each operator is required to complete 25 live interviews, which are then reviewed thoroughly by experts at country, regional and global level to assess performance and determine whether there are concepts within the questionnaire that need to be further clarified prior to approving the operator to proceed with data collection. To aid in this process, some of these calls are also recorded for quality control purposes and reviewed by experts in the country offices. Operators are also carefully supervised in-person by the call centre manager and usually by a WFP representative at the country level to ensure adherence to the protocol and to assist with unexpected situations and pending questions.

WFP follows a rigorous operator clearance process to avoid operator-related bias and ensure reliability of the data. After the first pre-test, only the operators who show full comprehension of the questionnaire and data collection methodology are cleared for data collection. A second round of training and pre-testing is done for operators who are not cleared in the first stage, and the same process is repeated. If operators are not cleared after the second pre-test, then they are removed from the project.

WFP also gathers feedback from the operators during the pre-test process, all of which is reviewed carefully and used to make necessary changes to the questionnaire, if needed.

Survey management

The management of real-time monitoring systems falls into roughly two stages. The first stage is referred to as the implementation phase. This covers the first 6-8 weeks of any new real-time monitoring system and covers the period from the beginning of the data collection to the country level validation and release of the first results to the HungerMapLIVE.

The second stage, maintenance, begins after the first results are released and continues for the duration of the real-time monitoring system. Maintenance basically means systematized oversight of the performance of the call centre, the quality of the data collected, as well as remaining vigilant to issues that may arise that would threaten the viability of the system. These issues could range from technical issues within the analytical pipelines to political issues in a country that could endanger the continuance of the real-time monitoring systems. Maintenance would also cover the need to scale up or modify data collection to address new challenges emerging in countries.

To effectively manage real-time monitoring systems, active support and participation is required at all levels in WFP, including WFP country offices and Regional Bureaux as well as the involvement and oversight of the WFP's Hunger Monitoring Unit (HMU) based in Rome. Each level has distinct responsibilities within the implementation and maintenance stages of the real-time monitoring systems.

Hunger Monitoring Unit: As real-time monitoring systems begin, the HMU is responsible for ensuring that the processing from collection to visualization is fully automated. In practice, this means that that HMU experts must set up the analytical pipelines to clean and process the data as well as make the data available to the teams in country. Importantly, as a part of the analytical pipelines, HMU experts ensure that automated information on call centre performance and overall data quality is available for country teams to view and,

when necessary, act on. Likewise, HMU experts must also make the necessary changes in the HungerMapLIVE and the Insight and Key Trends daily products to ensure that upon validation, the information is easily releasable. These responsibilities persist throughout the life cycle of the real-time monitoring systems.

WFP Regional Bureaux: Experts from WFP regional offices serve as an important pillar of support for country office staff. In practice, WFP regional experts backstop Country Offices in the day-to-day management of these systems, which could range from supporting in the regular data quality checks to smoothing over issues with service providers. They are also critical in appropriately positioning the information that emerges from these systems to ensure it fits seamlessly into the larger network of food security monitoring systems.

WFP Country Offices: Country office experts are critical to the management of real-time monitoring systems, as consistent, active involvement is needed to have effective systems in place. In the initial implementation phases, country offices continuously interact with the operators, call centre managers and service providers to ensure protocols are maintained, operators are performing optimally and the data that is emerging is of the highest quality possible. These experts also make important analytical decisions (i.e. determination of most appropriate weighting schemes) that require in-depth contextual knowledge of the situation in country. Most importantly, country office experts, after consultation with partners, make the final decisions on when data is officially released for the first time and they continually monitor the data after the release to ensure that it remains in line with expectations and that it is of high quality.

Implementation of a real-time monitoring system

To establish and effectively manage a real-time monitoring system in a country, multiple stakeholders within WFP country and regional offices as well as the global HMU are actively involved. The management of real-time monitoring systems falls into two main stages: **implementation**, which lasts roughly 6-8 weeks, and **maintenance**, which begins thereafter and continues for the lifecycle of the real-time monitoring system.

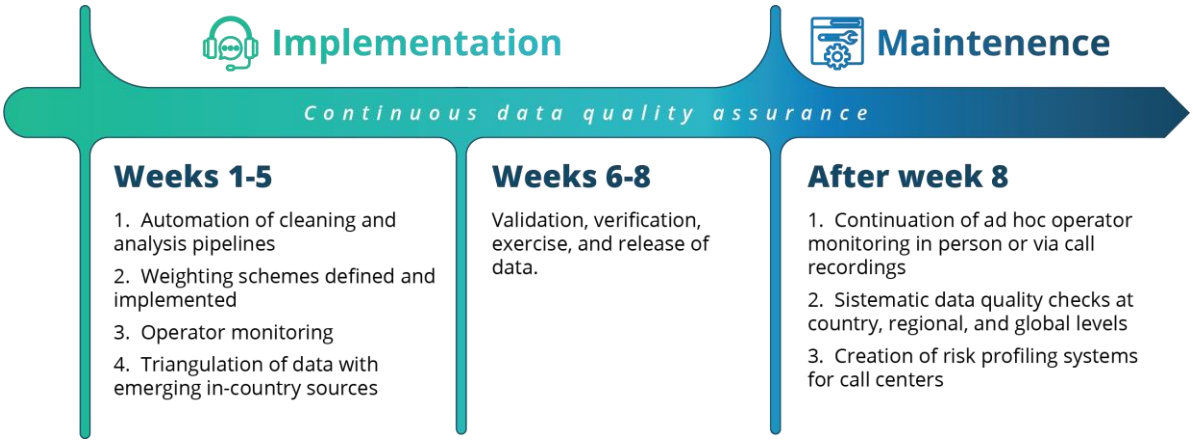

Figure 1

As *Figure 1* illustrates, there are several key processes in the implementation and maintenance stages that are critical and require more explanation. These include;

- Automation of the analytical and cleaning pipelines
- Weighting of real-time monitoring systems
- Quality assurance process

Automation of the analytical and cleaning pipelines: Real-time monitoring systems are possible given the automated systems that have been developed to pick up the data from the call centre, process and analyze it and then visualize it. These systems improve the quality and efficiency of the information while minimizing manual processes and thus reducing overall costs. *Figure 2* provides a schematic of how the systems operate.

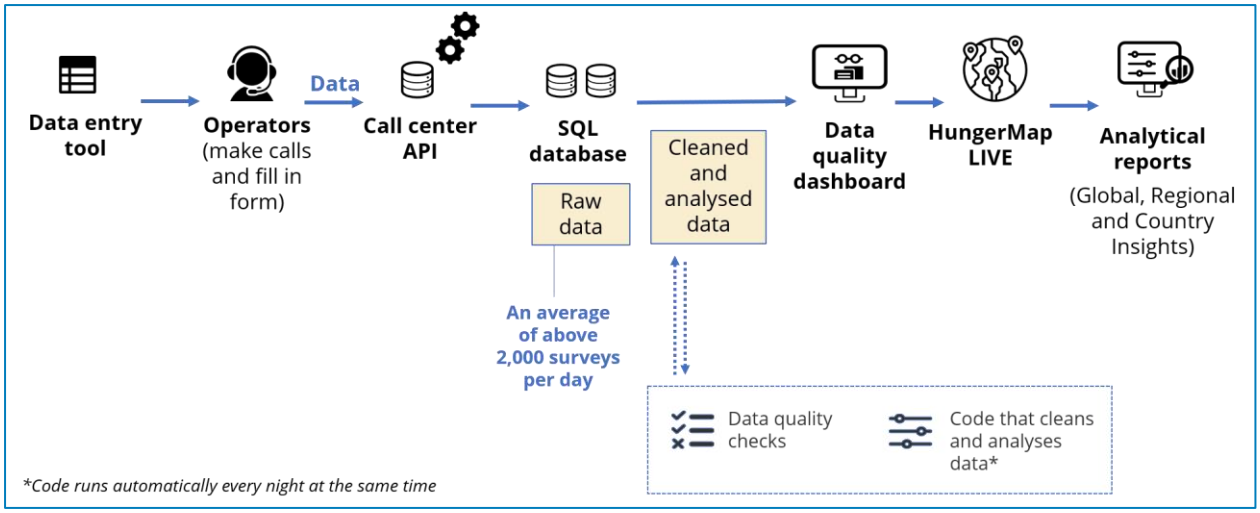

Figure 2

The data collected by the service provider is stored in the service provider database and shared with WFP through an Application Programming Interface (API). The data is received daily and is stored in a Microsoft Sequel Server database (MSSQL).

With slight variation in the questionnaires across monitoring systems, there are customized cleaning codes created per country. The cleaning process includes: re-coding of the response options when needed, generation of weights and the computation of the key food security indicators. At the same time, the key data quality indicators are parsed and visualized.

With cleaned data stored in the MSSQL database, the analysis code is run to process the results at strata and national level. The final, weighted results are computed using data collected in the last 30 or 60 days from any given day, and are stored in a table on MSSQL which contains results from all countries covered by the real-time monitoring and all days of data collection and analysis.

Weighting of real-time monitoring systems: Given the fact that real-time monitoring systems are mobile phone-based, there are inherent biases that must be mitigated to the extent possible[6]. The largest source of bias in CATI surveys is sampling bias.

Sampling bias, in CATI surveys, is introduced as not all households in any given country have access to phones, mobile network coverage, or, in certain cases, access to charging devices or power and are thus excluded or underrepresented in phone surveys. In short, the population eligible for selection in a mobile phone survey is slightly different from the actual population that resides in a particular country, often resulting in a sample that is artificially wealthier and more urban that the actual population in the country.

There are well tested and known measures to mitigate sampling bias in CATI surveys. This requires a robust weighting of the data within the sample to account for underrepresentation of certain demographics. As identified by WFP, the World Bank and others[7] the ideal method or "gold standard" for weighting CATI data is the Propensity Score Weighting (PSW) approach. The method uses multiple socioeconomic or demographic variables from both recent face-to-face surveys and real-time monitoring systems, to weight household responses in such a way to make the sample (and thus the findings) more reflective of the actual population in the country. PSW weighting has been utilized by WFP for ad hoc surveys in the past and has been validated

[6] Bowling A. Mode of questionnaire administration can have serious effects on data quality, Journal of Public Health, Volume 27, Issue 3, September 2005, Pages 281–291,

[7] Amankwah A, Ilukor J, Kanyanda S, Radyakin S, Sajaia Z, Shaw A, Wild M, Yoshimura K. (2020) High frequency mobile phone surveys of households to assess the impacts of COVID-19. Guidelines on CATI implementation.
https://documents1.worldbank.org/curated/ar/189691588696451053/pdf/Guidelines-on-CATI-Implementation.pdf

in countries such as Mali, Iraq, and Guinea. Implementing the PSW method, within the real-time monitoring systems, remains a work in progress as it introduces some complexities in the analytical pipeline that are currently being addressed by experts in the HMU.

As WFP works to implement PSW, current real-time monitoring systems utilize pre- and post-stratification weighting schemes. Population weights are used to compensate for over- or under-sampling of specific administrative areas, while demographic weights are used to mitigate selection bias and compensate for the under-represented households (i.e. households with low-income or a less-educated household head). Decisions on the application of weights are driven by certain parameters as outlined below:

- Population weights are generally not used if:
  - the number of mobile phone subscriptions per 100 people are above 100;
  - Sample size quotas at all levels are largely met with minimal over or under-sampling
- Demographic weights are generally not used if:
  - Various demographic patterns (education, water source) are in line with national averages from face-to-face surveys

If both weights need to be applied, they are multiplied into a final combined weight.

To note, due to the continuous nature of the data collection and the shifting analysis windows, when weights are required, they must be dynamic in nature. This means that weights are generated daily, as each daily estimate is based on sample characteristics that may differ slightly from one day to the next depending on the distribution of the selected households within the new analysis window. Put more simply, for real-time monitoring systems, each day is considered to be a separate survey and thus requires weights specific to the sampled households for that survey period.

Quality assurance process: The quality assurance process begins as data collection initiates and is prioritized throughout the life cycle of the real-time monitoring systems. Quality assurance and data quality checks happen at multiple levels, including at Country Office, Regional Bureaux and HMU. The distinct and parallel quality assurance processes are designed to look at different aspects of the monitoring systems while also building in purposeful duplication in certain processes to maximize the chance of identifying problems when they emerge.

As data collection begins, for instance, numerous processes are triggered at both HMU and country office level. At HMU, the first step is to pipeline the data collection exercise, so the data cleaning and analysis is automated. As a part of this analytical pipeline, certain data quality checks are conducted automatically on a daily basis and are displayed in data quality dashboards which are available for review at country, regional or global levels. While continuously iterating and improving on these automated capacities, the basic metrics calculated daily include:

- Adherence to the quotas at various administrative level
- Adherence to daily quotas (when applicable)
- Operator performance and potential bias
- Time taken to conduct the survey
- Number of calls per operator

Manual checks of each of the approved operators are also immediately triggered after the onset of data collection both at HMU and country office level. HMU statisticians review the data from the approved operators, looking for outliers or other evidence that there are systematic errors or misinterpretations. At the same time, country experts monitor the operators in person or monitor the call recordings of the operators to assess their professionalism and ensure they are asking the questions and probing appropriately. Country office experts are also assessing the headline findings emerging to assess whether findings are aligned with other recent data collection efforts in country. Informal exchanges are happening

frequently between HMU and country office experts to share information on their findings. Formal gatherings to assess the progress of the monitoring systems and quality of the data collected are usually arranged either weekly or biweekly for the first 6 to 8 weeks of the data collection.

Simultaneous to this, HMU and regional bureau experts are engaging weekly with the service provider and call centre to address any emerging technical or political issues that have the capacity to derail the monitoring systems. Common concerns discussed in this forum could range from difficulty achieving quotas in certain parts of the country due to network outages to discussions on the safety of the operators themselves due to insecurity or disease worries. While very important in the initial stages of data collection, the need for these routine check ins with service providers remains for the duration of the real-time monitoring systems. Thus, these weekly check ins are a persistent feature of WFP's monitoring systems.

At the 6-8 week mark since the initiation of data collection, a final meeting between country office, regional bureau and HMU experts is held to validate and release the findings. To validate and release findings, country office experts must be confident, given all of the above routine checks, that operators are asking the questions appropriately, have a consistent interpretation of the questions and are providing data in line with both the protocols required but also consistent with recent surveys in country. If so, the data is validated, the analytical pipelines are put into "production" and the results begin to flow to the HungerMapLIVE and its corresponding products.

From this point on, though the data is live on a day-to-day basis, quality assurance processes remain in place. Automated daily checks remain and periodic data quality checks continue at both country office and HMU level. Notably, after findings are initially released, HMU experts establish a risk profiling system that periodically assess the performance of the call centre based on a series of metrics including achievement of quotas, adherence to the real-time monitoring protocols and proper use of the operators. This allows WFP to understand which call centres are the highest performing and are extremely low risk or, by contrast, which call centres are struggling to perform adequately and are thus higher risk. To note, risk profiling systems have been only partially automated to this point, however by the end of January 2022, this risk profiling system will be fully automated and running continuously.

III. In Summary

To date, real-time monitoring systems have been established and are currently operating in 36 countries worldwide. HMU plans to expand the scope of these systems in 2022, subject to the availability of funds.