*Mátyás Bódi – SurveyCloud – Hungary*

Data collection on the Middle East

This paper is dedicated to provide insight for the technical background of a grandiose a firm-level survey, conducted in 10 countries with the research platform in question. Among my goals the introduction of research findings or any theoretical issue about the survey don’t appear, partially because World Bank’s Enterprise Survey is still an ongoing project, and on the other hand, here the innovative data collection methodology is concerned, and intended to be presented. The paper has a special focus on the challenges and methodological requirements, set by the client, as well as on the software solutions offered. In the second part, some of the advanced monitoring tools will be explained from technical point of view, while its broader adaptability is going to be examined as well. Finally I will give a full picture about how data quality was ensured through the questionnaire with more than 400 questions and data entry post.

**CAPI** (*Computer Assisted Personal Interviewing*) **data collection**.

Deployment of laptops, smartphones and tablets is not a new phenomenon in market/marketing research industry, nor in the world of traditional and academic opinion research. Nowadays such methodologies are in the spotlight of the respective professional discussion and joint events as mobile solutions, online research with its ‘*gamification’* approach, but mention must be made of bid data and other hardware supported means as eye-tracking, etc. These methodologies are considered to be sexy by market researchers, globally, even though some of the important circumstances are not taken into their account. These factors are those can enable the usage of the latest technologies in data collection, but their presence is not self-evident.

To sum up, a remarkably big gap must be acknowledged between different parts of our world; even country by country inequalities can be recognized in term of data collection opportunities, and the feasibility of certain tools.

Nevertheless, face-to-face interviews with pan and paper data entry still have a significant role to play, especially on the Middle-East. ESOMAR’s last MENA Research Forum, held in Dubai last March, highlighted some major obstacles, researchers need to face with. Organisations, dealing with data collection, need to divide the majority of their research budget for enhancing data consistency and quality checks, although introducing CAPI not just dramatically reduces the time lag between data collection and data analysis, but data validation can be done in the time of data collection, so the information is ready for statistical analysis as soon as surveying is completed.

A long list of other benefits can be yielded by the adaptation of CAPI; number of coding errors as well as incidence of missing data can be eliminated by the automated routing programming.

Flexibility is another provided benefit among many others. Changes in the structure of the questionnaire or the contact sheet can be instantly reflected on the interviewers' devices. This allows for last-minute updates or error corrections. CAPI technology simplifies conducting surveys with dynamic structure, where the questions to be asked will vary depending on the answers given by the respondent.

**Challenges**

In the region of our operation (Middle-East & North-Africa – MENA), local makings with the client’s requirements set a wide range of challenges that the platform had to handle;

* 3 kind of firm level questionnaire by business type with significant overlapping and advanced routing logic (hard/soft checks, skips, preconditions, etc.),
* Establishing a bilingual (Arabic/French) working environment for Android devices.
* Preliminary screening of companies over the phone in order to establish eligibility. The joint WebCATI system of Gallup Hungary provided eligibility information for the CAPI platform that sent back feedback about the field results.
* Collection of accurate GPS coordinates of each interview (data entry point), including surveyors’ route information.
* Time-stamp recording and interview length measurement.

Among the biggest obstacle, the absence of phone numbers must be mentioned, even though similar firm-based data collection projects on local businesses have been conducted in several countries (i.e.: Egypt, Palestine). The unblocking of the phone-based screening in the respective countries was managed by the tablet adaption of the screener questionnaire in order to establish company eligibility for further investigation (interviews). In these cases the screener interview was completed face-to-face instead of phone calls, while the two questionnaire (screener and main) was linked to each other via embedded scripts.

**Project implementation, software employment**

The issue of online and offline data collection also raises several questions. Online work could be carried out with data SIM cards, but due to the poor 3G coverage of the majority of the countries (or even the delay of technology introduction and high-speed mobile broadband) made it unfeasible. As a consequence, supervisors (local vendors – research companies) and the client were informed on a daily bases by the field-work team, while sample updates, amendments were received by interviewers. When the device synchronizes with the server it sends/receives the surveys, their results and application updates if any exist. The communication had two directions; the headquarters were informed about the data collection process; exhausted or fulfilled quota divisions, but on the other hand they can released and assign new addresses instead, in order to make the local vendors go closer to the target numbers of the research.

The above mentioned tablet based screener was applied on the same way, as it was described by the device synchronization. The defined eligibility codes of respective firms were sent by the interviewers on daily bases. In accordance with the adjusted sample, screened companies were assigned to interviewers in order to re-visit the firms and go on with the main questionnaire.

As it was mentioned, in term of its complexity, length and nature, the questionnaire itself was a challenge both for vendors and their field-work team. Constant check on midterm survey data was inevitable for the success. Not just the functioning of the built-in routing logic had to be checked, but refusal rate of certain sensitive questions (addressing business-government relations and bribery-related topics) monitored as well. The observations have generated several changes in the questionnaire structurally. Surveys, quotas and survey assigning to field surveyors are centrally managed as well as those alterations in the questionnaire, justified by the survey results, constantly coming back from the field.

Another significant group of information that could be monitored throughout the data collection stage enables supervisors to make background checks of each interview. The application saved the exact time and length of the interviews as well as the timestamps (question by question) was recorded. Interviews whose length value didn’t reach a certain threshold got forwarded to further monitoring, and the time stamps got double checked as well.

GPS coordinates, including route information of interviewers are saved. Watching each twist and turns of the surveyors on the field can bring some useful information for management. Most importantly, false interviews (conducted not at the place of the businesses) can be detected and excluded. The possible cheat cases can be sorted not just by random selection, but the built-in distance enumerator tool that provided information about the spatial gap (if existed) between the address, assigned to interview prior the actual visit and the exact location of the interview.

In several cases, local vendors had no information about the GPS coordinates of the companies in sample, thus geographical interpolation was employed by address (street level information) attributes.

What all the above mentioned background data have in common is the nature as they are being recorded. Timestamps, length values, even GPS coordinates are recorded during offline work; however devices are needed to be synchronized with the server to upload all those information by the end of each day. As a consequence, no matter where the field surveyors were, the solution allows the vendors’ staff to conduct b2b surveys, while being fully offline o the field.