Data Handling Manual
Guidance to Survey Firms

Data Entry Program, Data Management & Integration, & Field Based Data Entry Training and Support

Kimetrica International Limited
2020 Pennsylvania Ave, NW #715
Washington, DC 20006
www.kimetrica.org

Tel: +1 202 470 6974
Fax: +1 202 280 1272

Authorized Negotiator:
William (Rob) Rose, Director
Email: rob.rose@kimetrica.org

Date: 27 September 2010

Prepared for:
The World Bank
Attention: Bertha Briceno & Claire Chase
1818 H Street, NW
Washington, DC 20433

Tel: 202-473-4111
Tel: 202-473-2928

Kimetrica

Table of Contents

List of Abbreviations and Acronyms

DHS	Demographic and Health Survey (a MEASURE project)
IRB	Internal Review Board
LSF	Local Survey Firm; i.e., the in-country firm who conducts the fieldwork and data processing tasks for the baseline and/or endline surveys
MICS	Multiple Indicator Cluster Survey (a UNICEF-funded survey)
PI	Principal Investigator
PDA	Personal Digital Assistant
PUF	Public Use File
ToR	Terms of Reference
TSSM	Total Sanitation and Sanitation Marketing
WSP	Water and Sanitation Program

1 Introduction

On 29 April 2010, Kimetrica International Limited was awarded a contract with the World Bank to support the Data Entry Program, Data Management & Integration, and Field-Based Data Entry Training and Support for the Bank's impact evaluation surveys related to the Water and Sanitation Program (WSP)'s programming and the scale-up of the Total Sanitation and Sanitation Marketing (TSSM) project. This document, the Data Handling Manual, is the second in the series of seven deliverables for the project. It is based on the Terms of Reference (ToR), and will present such topics as:

- Ensuring data, document, and specimen integrity
- Steps to guard against data, document, and specimen loss
- Procedures for secure batch (cluster) handling and transport
- Field enumeration and data processing procedures to protect respondent privacy in accordance with local and global IRB regulations

2 Data Handling Basics

Sound data handling procedures will address data captured and/or stored both electronically (such as on laptop computers, PDAs, flash drives, CDs, and cell phones) and non-electronically (for example, biomarker samples, paper questionnaires, and field listing sheets), and should be drafted and in place by the start-up of any data collection activity. This document will elaborate on clear guidelines to ensure that the data are collected and processed in a safe, secure, and professional manner.

Below are a list of data handling basics that apply to all survey staff, though those engaged in direct fieldwork or data processing will find them most relevant.

2.1 Make It Identifiable

The following issues ensure that survey materials, in particular survey questionnaires, are clearly identified with proper administrative codes, and that the author of all markings on a questionnaire is known:

1) It is of paramount importance that Field Supervisors receive all necessary information to find and identify each cluster *before* entering the cluster. They must be given household listing sheets with all cluster identifiers designated, maps indicating cluster boundaries and households selected therein, along with instructions on how to find and identify each cluster and each selected household therein. Supervisors need to review these materials before entering the cluster to ensure the materials are adequate for the task.

2) It is equally import that all questionnaires have correct and complete administrative and personal identifiers written on all questionnaires where indicated. Field Supervisors must ensure that (a) Enumerators fully complete all identifiers where needed; that (b) when Enumerators move to the next cluster they "reset" themselves and begin using the correct identifiers for that cluster; and that (c) there are no duplicate household numbers within a cluster (i.e., that two households share the same household number).

3) It is of further importance that anyone, and in particular the data processing staff (since they are the last to work with the questionnaires), be able to identify who has written on the questionnaire during its movement through the survey operation. When questions arise during office editing or entry, it is critical for the person to know who has written what text on the questionnaires; i.e., whether it was the Enumerator, Field Supervisor or Editor, or office staff member. Therefore, it is strongly urged that different colored pens be used depending on the person's position. Suggested colors are black or blue pens (Enumerators), red pens (Field Supervisors and Editors), and green pens (office/data processing staff).

2.2 Keep It Confidential

The following issues ensure the confidentiality of survey materials both in the field and in the survey office:

1) All persons engaged for the survey must take oaths of confidentiality and agree to follow general guidelines set forth therein. An example of an oath is as follows:

> *"I will not disclose any information contained in the schedules, lists, or statements obtained for or prepared by the [name of survey firm] to any person or persons either during or after employment."*

2) Incorporate confidentiality concepts, and the need for secure data handling, in all staff training.

3) Enact a system for reporting concerns on data handling procedures. Provide cell phone numbers or other contact information of survey managers or their designee, so that anyone within the survey structure can report suspected violations.

In addition, the following guidelines will ensure the confidentiality of the survey data if administered to all survey personnel:

1) Survey staff (both in the field and office) should only discuss the details of individual household interviews with other team members to resolve problems or concerns—they should not be the topics of idle conversation.

2) Field staff should never discuss specific questionnaire responses where they could be overhead by others unconnected with the survey. It is recognized that field staff, being away from typical office environments, will have a more difficult time to find appropriate venues to discuss problems. Nevertheless, they must take efforts to avoid discussing questionnaire specifics in public venues.

3) Likewise, office staff should never discuss the specific questionnaires responses outside of the office/work environment. For example, office personnel going to lunch together should not discuss the content of the questionnaires that could be overhead by others unconnected with the survey.

4) Under no circumstances should office staff take questionnaires, listing forms, or other confidential materials to their home or other off-site location.

2.3 Keep It Secure

1) On a day-to-day basis, field staff must be attentive to the physical security of the questionnaires. Questionnaires should never be left unattended in unlocked vehicles; nor should they be left carelessly lying around the field staff's lodging, where hotel workers could easily peruse the materials.

2) Physical office spaces (such as the data entry room, questionnaire storage room, and any other office location containing confidential materials) must have the ability to be secured during non-use, and monitored during use.

3) Access to the data processing center should be restricted to those involved in the survey—colleagues not working on the survey, or friends, relatives, etc., of the data

processing or survey management staff should not be given access to the processing center (this includes the storage room).

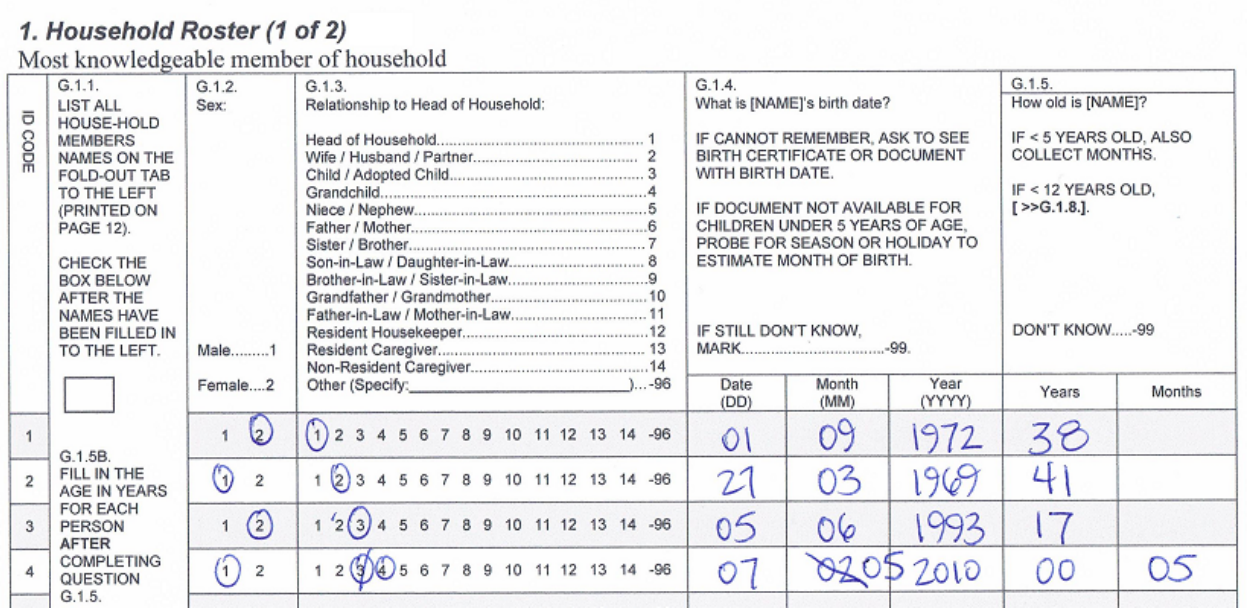
4) Data processing staff, Survey Management, and other survey staff coming into contact with electronic data files must not share these files with anyone other than the LSF's internal staff affiliated with the survey, World Bank WSP officials, and Kimetrica, unless explicit permission is given to do so by either World Bank officials or Kimetrica.

2.4 Keep It Legible

There is nothing so frustrating when collecting data as to find one's best efforts thwarted by sloppily recorded responses. The following example illustrates two common problems in the recordation of data and how they should be properly resolved:

- Circled responses: Should the Enumerator find they circled the incorrect response, a single diagonal line should be neatly drawn through the incorrect response, and the correct response should be circled **once** (not multiple times for emphasis). See the relationship field for the fourth person below in Screenshot 2-1 for an example of how this type of correction should be made. In addition, Enumerators must make their best effort to neatly circle **one** response, rather than quickly circling a response only to find they have partially circled two responses, such that it is unclear which response is the desired response.

- Written responses: Should the Enumerator find they entered the wrong data, a single diagonal line should be neatly drawn through the incorrect numeric (or text) response, and the correct response should be written adjacent to the original entry (either above or below the crossed-out response, or to the left or right if space does not permit entries above or below). At **no** time should the Enumerator scratch out or otherwise obliterate the original response, nor should they overwrite the correct response overtop the incorrect response (for example, changing a 5 to an 8 by drawing the 8 forcefully overtop the 5 numerous times). See the month of birth field for the fourth person below in Screenshot 2-1 for an example of how this type of correction should be made.



1. Household Roster (1 of 2)

Most knowledgeable member of household

ID CODE	G.1.1. LIST ALL HOUSE-HOLD MEMBERS NAMES ON THE FOLD-OUT TAB TO THE LEFT (PRINTED ON PAGE 12). CHECK THE BOX BELOW AFTER THE NAMES HAVE BEEN FILLED IN TO THE LEFT. ☐ G.1.5B. FILL IN THE AGE IN YEARS FOR EACH PERSON AFTER COMPLETING QUESTION G.1.5.	G.1.2. Sex: Male.....1 Female....2	G.1.3. Relationship to Head of Household: Head of Household....1, Wife / Husband / Partner....2, Child / Adopted Child....3, Grandchild....4, Niece / Nephew....5, Father / Mother....6, Sister / Brother....7, Son-in-Law / Daughter-in-Law....8, Brother-in-Law / Sister-in-Law....9, Grandfather / Grandmother....10, Father-in-Law / Mother-in-Law....11, Resident Housekeeper....12, Resident Caregiver....13, Non-Resident Caregiver....14, Other (Specify:_____)...-96	G.1.4. What is [NAME]'s birth date? IF CANNOT REMEMBER, ASK TO SEE BIRTH CERTIFICATE OR DOCUMENT WITH BIRTH DATE. IF DOCUMENT NOT AVAILABLE FOR CHILDREN UNDER 5 YEARS OF AGE, PROBE FOR SEASON OR HOLIDAY TO ESTIMATE MONTH OF BIRTH. IF STILL DON'T KNOW, MARK....-99.			G.1.5. How old is [NAME]? IF < 5 YEARS OLD, ALSO COLLECT MONTHS. IF < 12 YEARS OLD, [>>G.1.8.]. DON'T KNOW.....-99	
				Date (DD)	Month (MM)	Year (YYYY)	Years	Months
1		1 (2)	(1) 2 3 4 5 6 7 8 9 10 11 12 13 14 -96	01	09	1972	38	
2		(1) 2	1 (2) 3 4 5 6 7 8 9 10 11 12 13 14 -96	27	03	1969	41	
3		1 (2)	1 2 (3) 4 5 6 7 8 9 10 11 12 13 14 -96	05	06	1993	17	
4		(1) 2	1 2 3 (4) 5 6 7 8 9 10 11 12 13 14 -96	07	~~02~~05	2010	00	05

Screenshot 2-1: Correcting Responses

2.5 Keep It Clean

In an effort to keep questionnaires clean from excessive wear and tear that could lead to difficulty in their processing, the following guidelines should be followed:

- Survey staff should not eat or drink (except water) when handling the questionnaires.
- Survey staff should not smoke cigarettes, cigars, pipes, or other tobacco products, nor chew tobacco when handling questionnaires.
- In the field, questionnaires should be kept in bags or satchels whenever possible, or placed atop tables, chairs, or other furniture, rather than placed on the floor or ground.
- In the office, questionnaires should be kept on desks, chairs, or shelves whenever possible, rather than placed on the floor.
- In the office, the storage room shelving should be cleaned of all excessive dirt and debris before being placed into use.

2.6 Keep It Organized

In order to more easily find individual household questionnaires within a cluster, whenever the cluster is not actively in use (i.e., the cluster is in transit from the field to the main office, is in

the storage room, or is awaiting keying/editing on a person's desk), the cluster should be kept tied up or otherwise securely bundled together, with the questionnaires in ascending order according to household number (i.e., lowest household number on top).

Additionally, all modules belonging to a specific household should be kept in ascending order according to module number, such that Module 1 (the Household Roster) is first/on top, followed by Module 2 (Education Roster) underneath, Module 3 (Labor) below Module 2, and continuing in this fashion for all modules collected for the household. This facilitates data entry, as well as consistency editing.

3 Confidentiality Agreements

One key element in preserving confidentiality is to have all persons involved in the survey sign a non-disclosure statement before entering into employment. In addition to providing example terms of reference for the various survey team positions, we have included in Screenshot 3-1 below a draft non-disclosure statement for local survey firms (LSF) to use for field staff. The non-disclosure agreement should:

- Restrict the Survey Staff (field and data entry) from discussing responses with colleagues unless expressly authorized by the Supervisor.
- Restrict the Survey Staff from discussing the responses with any intentions of malice.
- Restrict the Survey Staff from discussing the responses with non-survey team members.
- Restrict the Survey Staff to discussing responses only with the Respondent or the Supervisors only when seeking clarification.
- Clearly outline the consequences of non-adherence to the rules of confidentiality.

SUPERVISOR/ENUMERATOR CONFIDENTIALITY AGREEMENT

In consideration of my employment by *[Survey Firm]* for the duration of *[the Survey],* and the compensation now and hereafter paid to me, I hereby agree as follows:

1. RESPONDENTS INFORMATION

I understand that the *[Survey Firm]* has received, and in the future will receive, from Respondents confidential or proprietary information ("Third Party Information") subject to a duty on the *[Survey Firm]* part to maintain the confidentiality of such information and to use it only for certain limited purposes.

During the term of my employment and thereafter, I will hold Third Party Information (e.g., from Respondents) in the strictest confidence and will not disclose their information to anyone (except in connection with my work for the *Survey Firm*), unless expressly authorized by an executive office [*Survey Manager/Supervisor]* of the Company in writing.

2. RECOGNITION OF *SURVEY FIRM'S* RIGHTS; NONDISCLOSURE

At all times during the term of my employment and thereafter, I will hold in strictest confidence and will not disclose, use, lecture upon, or publish in any way any of the *[Survey Firms]* proprietary Information except as such disclosure, use, or publication may be required in connection with my work for the *[Survey Firm]*, or unless an executive officer(Supervisor/Survey Manager) of the *[Survey Firm]* expressly authorizes such in writing.

The term "Proprietary Information" shall mean trade secrets, confidential knowledge, software code, data or any other information of the *[Survey Firm]* . By way of illustration but not limitation, "Proprietary Information" includes

(a) inventions, mask works, trade secrets, ideas, processes, formulas, source and object codes, data, programs, other works of authorship, know-how, improvements, discoveries, developments, designs and techniques (hereinafter collectively referred to as "Inventions");and

(b) information regarding plans for research, development, new products, regulatory matters, marketing and selling, business plans, budgets and unpublished financial statements, licenses, prices and costs, suppliers and customers; and information regarding the skills and compensation of other employees of the *[Survey Firm]*.

3. CONSEQUENCES OF NON-ADHERENCE

I understand that violation of respondent confidentiality will result in immediate termination of my employment as a Survey Supervisor/Enumerator. If the survey has been completed, violation of respondent confidentiality can result in legal proceedings against me.

Screenshot 3-1: Sample Non-Disclosure Agreement

4 Survey Management Guidelines

This section discusses the overall practices and procedures that survey management should follow for the entire survey operation.

4.1 Who's In Charge?

Within each LSF there needs to be one person who assumes overall authority for the project. This "point" person will be the one to whom the World Bank, the Principal Investigator, and Kimetrica confer with on survey issues. This person may or may not be the same person with whom the field and data processing staff, and local laboratory staff, work with. This person (let's call them the Survey Manager), needs to be engaged on a day-to-day basis with the entire survey operation, in particular the field operation.

This person needs to be the central "go to" person for all problems encountered in the field, whether it be with logistics, staffing, or questionnaire errors. This person likewise needs to be the one to get the solutions out to the field staff, either personally or through their designee.

4.2 Reporting Process

There needs to be a clear reporting hierarchy to transmit field concerns, such as supplies running low, transportation issues, questionnaires lost, errors uncovered in the questionnaire, etc. Field staff need to know who to call, and likewise the Survey Manager will need to know who to contact in the field. A complete list of all persons involved in the fieldwork, with their job titles/roles and cell phone numbers given, needs to be distributed to the Survey Manager, all Field Supervisors, and drivers.

Another important topic is the method of transmitting questionnaire revisions to the field staff. If the revisions are minor, the Survey Manager (and/or their designee) can contact the Field Enumerators directly to discuss the problems and their corrections. However, should the list of revisions prove long, the Survey Manager should follow up phone conversations with a printed list of the corrections, to be delivered via team driver or questionnaire delivery person.

4.3 Field Site Visits

Survey Managers should made unannounced site visits to survey teams to ensure they are following best practices for the physical and verbal safety of the data, i.e., to ensure proper methods of questionnaire storage are being followed, and that field staff are not discussing questionnaire details in unsecured public areas. The Survey Manager should also be alert to any

other transgressions or problems occurring, as well as allow field staff to voice any concerns they have with the survey.

Principal Investigators should also make site visits, as a method of verifying the local survey firms' overall work performance, but also to serve as another support person to the field staff.

4.4 Computer File Backups

Although the Data Processing Supervisor is responsible for conducting weekly backups of all computer files, the Survey Manager is responsible for ensuring this critical activity takes place. To facilitate this activity and serve as a redundancy to the operation, the Survey Manager should request one of the weekly backup copies be sent to their offices. Ideally, a reusable medium such as a pen drive should be used, preferably two. In this fashion, pen drive #1 can be backed up one week, pen drive #2 can be used for the second week, and from that point forward the pen drives can be alternatively sent to the DP Center, such that if a pen drive were lost or be rendered inoperable, the previous backup exists. This pen drive should be stored in a secure place (such as lockable desk drawer), and should not be used for other file work.

5 Field Guidelines

This section presents the guidelines to follow for each stage of field collection. This encompasses Field Supervisors, Field Editors (if a special position is created for this task), Enumerators, Biodata Collectors (if a special position is created for the collection of anthropometric data, anemia testing, and stool collection), team drivers, and anyone else directly coming into contact with field-collected data.

5.1 Training Components of Data Handling

-

5.2 Enumerator Observation

While Field Supervisors should always be attentive to their Enumerators' performance, they should be especially so during the Enumerators first week of actual fieldwork. Unless Enumerators are full-time interviewers, they will not have had enough practice interview sessions during the training to be confident interviewers, nor know how to resolve problems they might encounter in the field. Field Supervisors should be alert during this initial period to determine who might need additional assistance, or who might even need to be dismissed, should their conduct or rate of errors prove unacceptable.

During the first two days of fieldwork, each Enumerator should be observed for at least one complete interview. If the house is unoccupied or the respondent refuses to be interviewed, then the Enumerator must be observed in another household. During this observation period, the Field Supervisor must be monitoring to ensure the Enumerator is adhering to the survey methodology taught during training (in particular, completing the household roster correctly, making the correct child selection, and completing the correct modules), is maintaining respondent confidentiality during the interview (i.e., that the respondent's family members or other individuals are not within audible range during the interview), and is conducting the interview in a professional manner (speaking with deference to the respondent, not using profanity, not chewing gum, etc).

During the observed interview, the Field Supervisor should keep notes about the Enumerator's performance, both good and bad. Field Supervisors should refrain from correcting the Enumerator during interview, unless the Enumerator is having problems.

After leaving the household, the Field Supervisor and Enumerator should find a place where they can talk freely and comfortably discuss the interview. The Field Supervisor should first make a quick perusal of the questionnaires to ensure completeness, accuracy, and neatness of the information obtained. For example, if there had been three children under the age of 5 years in the household, the appropriate modules should be completed. The Field Supervisor should also check that skips are being properly followed. At this time a full review of the questionnaire is not possible and should be saved for the evening; rather, the Field Supervisor should take 15-20 minutes to look for the most egregious errors.

Field Supervisors should then discuss with the Enumerator the problems they found in the questionnaires and how to resolve them. They should then give feedback on any problems they noted during the interview, or commend them on things they did well. No matter how poorly the performance of the Enumerator, the Field Supervisor should always find something to compliment the Enumerator on—receiving only negative feedback will be self-fulfilling.

In the latter portion of the first week, the Field Supervisor should again accompany the Enumerator to at least one complete interview to note their performance, which will hopefully be improving. The Field Supervisor should repeat the same steps used for the first observation.

5.3 Field Editing

Questionnaires should be reviewed the same evening of the interview, or the morning subsequent to it, so that if there are questions or problems with the information collected that

cannot be resolved by speaking with the Enumerator, a return visit to the household can be made while the team is in the same cluster.

During this activity, Field Supervisors (and/or Field Editors) should always "edit lightly", that is, they should not edit indiscriminately, and they should **never** guess or otherwise make up an answer. Even seemingly simple to correct errors such as missing identification codes should always be reconfirmed with the Enumerator first; it could be that questionnaires from different households were intermixed, or it could be a more serious problem such as the Enumerator went to the wrong household/place.

5.4 The Interview

As mentioned elsewhere in this document, and in particular in Section 2, the Enumerator needs to adhere to the following guidelines at all times:

1) Keep the confidentiality of the respondent in mind

2) Keep the questionnaires organized by household number, and by module number within each household

3) Not smoke, eat, or drink (except water) when handling the questionnaires

5.5 Re-Interviews

Within each cluster, the Field Supervisor should re-interview at least one household, using the household questionnaire only (Module 1). This should occur 1-2 days after the initial Enumerator's visit. The Field Supervisor should then compare their household questionnaire to the Enumerator's household questionnaire, reviewing to ensure such items as:

(1) The Enumerator is finding the correct household

(2) The number of household members on the household cover sheet matches the total number of persons listed in the household schedule

(3) There is no age misreporting (i.e., dropping children under the age of 5 years, or changing the ages of children age 0-4 years to 5 or older in order to reduce the Enumerator's workload)

Any differences between the two versions should be brought to the attention of the Enumerator. Depending on the severity of error found, the Field Supervisor should pay closer attention to the Enumerator's work for the succeeding days.

5.6 Monitoring Performance

One way to keep Enumerators engaged throughout the fieldwork, to make them feel part of the process, is to hold daily team meetings. They'll most likely be longer at the beginning of the survey (30-60 minutes), as Enumerators discuss problems they're having and ask more questions, but within fairly short order they should only last 15-30 minutes a day. These team meetings should be a place for the team to discuss the problems they're facing (which other team members are most likely facing as well), for the Field Supervisor to make announcements (such as passing on corrections to the questionnaires), to adjust schedules as needed, etc.

After several weeks in the field and interest levels wane, this is also a good venue for the Field Supervisor to give pep talks to keep the team on focused and on track.

5.7 Preparing Questionnaires for Shipment

While the fieldwork is underway, questionnaires will need to be returned to the office on a regular basis so that the data entry operation stays in sync with the fieldwork and does not fall behind. Questionnaires should always be submitted to the home office on a cluster basis, and should be bundled together as such.

Utmost care should be devoted towards maintaining the condition of the questionnaires. Clusters should be kept bundled together, either in bags or satchels, or tied together with twine. If the clusters are to be transported in open vehicles such as trucks or jeeps, then the clusters must be completely covered by tarps or other waterproof material to protect them from dirt or rain.

If drivers are forced to spend the night en route to the office and an open vehicle is being used, then either the vehicle must be placed in an area that can be secured or guarded, or else the contents of the vehicle must be off-loaded to a secure location for the night. Likewise, if drivers are using closed vehicles they must ensure that the vehicle is locked, windows are up, and if possible that the vehicle is guarded or otherwise monitored during the night. If there is any question about the safety of the contents, then they must be off-loaded to a secure location for the night.

5.8 Preparing Biodata for Shipment

6 Data Processing Guidelines

This section discusses items of concern for data handling within the data entry operation. This includes questionnaire receipt from the field, questionnaire storage while in process, data entry, editing, and the overall management of the data processing operation.

Throughout the survey period, questionnaires should be kept bundled according to cluster, either in a satchel, bag, or tied with string, so that the cluster designation is clear. While we refer to questionnaires and clusters interchangeably in this section, in general we are referring to the movement or processing of a cluster.

6.1.1 Control Systems

Control systems assure sample integrity, reconcile the sample selection with the returned clusters, track, and control the data entry operators' work. The DHS and MICS have developed control systems that allow data entry supervisors to easily check in clusters, verify that there are no duplicates or omissions in the sample selection, ensure operators enter all anticipated questionnaire modules, and through the use of double-entry, ensure no keying errors occur. The control system should provide the following functionality:

- Tracking: allows the Data Entry Supervisor to chart data entry progress
- Reporting: allows the Data Entry Supervisor to run status reports that show number of questionnaires entered, speed, etc. for each data entry clerk
- Quality control: allows the Data Entry Supervisor to run field check tables and other consistency checks
- Backup: allows backup of the data files and programs to a remote server (if available) or flash drive/other external media on a daily basis.

6.1.2 Questionnaire Receipt

Immediately upon receipt of the questionnaires from the field (which must be submitted on a cluster basis), a person must be assigned to verify the contents of the shipment. Depending on the volume of questionnaires received during the survey operation, one to two full-time persons will be needed for this task (henceforth referred to as the Questionnaire Administrator).

Normally a driver will return several clusters to the office at a time from one or more teams. Upon arrival, there should be a designated area within the storage room, or a small room adjacent to the storage room, that can accommodate these questionnaires. Preferably, shelving should be established clearly marked "Newly Arrived Clusters" (or words to that effect) which the Questionnaire Administrator can direct them to for unloading.

For each new cluster received, the Questionnaire Administrator should conduct the following checks, in the order indicated:

1) Ensure all household questionnaires within the cluster are sorted in ascending order by household number (i.e., lowest household number on top), and that all modules belonging to that household are in ascending order by module number immediately following the household questionnaire.

2) Ensure all geographic and administrative identification codes are correctly completed on the cover sheet of the household questionnaire, and that no household shares the exact combination of identifying codes with another household.

3) Ensure the proper number of household questionnaires exist for the cluster; i.e., how many households should have been interviewed within the cluster? All household questionnaires need to be returned, even if the interview was partially completed, refused, or otherwise unfinished.

4) Ensure that the total number of household members stated on the household cover sheet agrees with the number of members recorded in the household roster.

5) Ensure that modules have been completed (or entries within a module have been completed) for each eligible child listed in the household roster.

Once these checks have been made and there are no discrepancies, the cluster can be placed on shelves in the storage room, in numerical order according to cluster number. The shelves should be clearly marked to facilitate cluster storage and retrieval.

If computer access is possible, it is suggested that the Questionnaire Administrator register the cluster within the CSPro tracking system at this time. This helps in several ways, as it allows the cluster to be trackable. The Data Processing Supervisor can then run reports to:

- Show the number of clusters returned to the office, so the Data Processing Supervisor can determine whether or not they are running out of unprocessed questionnaires

- Show the number of clusters returned by team, so that Survey Management can determine how teams are performing in relation to one another, whether certain teams are lagging behind (indicating the team might be facing problems in the field), and whether certain teams are finishing ahead of time (which isn't always good, as it could indicate they are rushing through the interview cycle).

- Show the overall progress of data processing. If clusters are entered as they come into the office, progress can be shown by (a) clusters received (b) clusters finished or undergoing main entry (c) clusters finished or undergoing verification entry and (d) clusters saved after main and verification entry show no discrepancies between the two files. Should editing be made on the clusters, then an additional two steps can be shown, namely (e) clusters being edited and (f) clusters finished the editing phase.

Occasionally questionnaires will be returned from the field that will be deemed incomplete or otherwise unkeyable. These questionnaires should never be discarded; rather, they must be retained with the cluster until the project's end when all questionnaires are ready for disposal.

Note: If for a given country there will not be a fixed number of households interviewed within each cluster, i.e., from cluster to cluster the number of households selected for interview will vary, then it is suggested that (1) a master listing of the anticipated number of households within each cluster be supplied to the Questionnaire Administrator, so they may know if the household questionnaire count is short or over; and (2) that a control sheet be affixed to each cluster indicating the actual number of household questionnaires for that cluster.

6.1.3 Questionnaire Security

Section 2.3 above discusses the following points that should be adhered to within the data processing center:

1) Physical office spaces (such as the data entry room, questionnaire storage room, and any other office location containing confidential materials) must have the ability to be secured during non-use, and monitored during use.

2) Access to the data processing areas should be restricted to individuals involved with the survey—professional or personal visits to data processing team members should occur outside the office area.

3) At no time should office staff remove questionnaires or other confidential material from the data processing center, unless instructed by Survey Management.

In addition, the storage area used to house the questionnaires must have shelving so that questionnaires can be kept off the floor—this is important should flooding be a concern, either due to rainy seasons within a country, or more mundane problems such as flooding caused by plumbing problems within the building.

On the other hand, in case of potential fires from cigarette smoking (which should of course be banned), electrical box or wiring malfunctions, or other sources, one or more fire extinguishers should be placed within the data processing center. In particular they should be placed within the storage room (since at any given time the bulk of the questionnaires will therein reside and they would pose the largest potential source of "fuel" to the fire), as well as in the data processing room to protect the computer equipment (another reason to regularly back up files to minimize the loss of data). If fire extinguishers are not available in-country, then pails of sand with shovels should be placed strategically throughout the data processing center.

In the unlikely chance of a fire or accident, staff must be able to quickly evacuate the premises. To facilitate this, questionnaires should never be stacked in hallways or corridors where they can block egress or pose tripping hazards, nor should they be piled up near doorways. They should also never be used to prop open doors, for in the event of a fire, doors must be able to be closed to help block the flow of smoke and/or flames.

6.1.4 Questionnaire Handling

Sections 2.3, 2.5, and 2.6 above discuss several items of note, and should be read in its entirety. The key points are summarized below:

1) In an effort to keep questionnaires clean from excessive wear and tear, data entry staff should not drink (except water), eat, smoke, or chew tobacco when in contact with the questionnaires.

2) Questionnaires should be kept on desks, chairs, or shelves whenever possible, rather than placed on the floor. Shelves should be cleaned of all excessive dirt and debris before being placed into use.

3) Whenever the cluster is not actively in use (i.e., it is not being keyed by a data entry operator), the cluster should be kept tied up or otherwise securely bundled together, with the questionnaires in ascending order according to household number. For each household, modules should be kept in ascending order according to module number.

6.1.5 Data File Backup

Backups are an integral part of any data processing activity, and needs to be done to ensure against loss of data from fire, flood, computer failure, and even theft, should the computers being used find they suddenly have legs. Correspondingly, once these copies are made, every effort needs to be made to ensure their safety from both loss and theft.

Backups should occur on a regular basis. Timing for this activity should be influenced by one primary concern: how much data can the agency afford to lose? If losing a week's worth of data would not pose a hardship, either financially (for the extra effort involved to rekey up to a week's worth of work) or in scheduling (can the agency afford to now set back data entry operations by a week), then backups can occur on a weekly basis. Ideally, they should occur on a bi-weekly or daily basis. However, under no circumstances should backups not commence at least weekly.

Backups can be made to either removable devices (such as flash drives or R/W CDs) or stationary devices (servers). Servers usually allow for a more secure backup, as unlike removable devices, the device cannot be lost. It also allows for better security, as generally servers have passwords and cannot be accessed by the general public. Contrast this with removable devices, which can be easily lost, or read by unauthorized persons should they come into contact with the removable device. Therefore, advice will not be made as to which device is the better; rather, this should be left to the local agency to decide.

However, note that if the server is located on the same premises as the data entry operation, then an effort should be made to periodically make and keep copies of the data in a secure off-site location as protection against fire, flood, or other physical disturbances. In addition, at least two devices should be maintained—the first device to be used for the initial backup, the second device for the second week's backup. From that point forward, the devices should be rotated into use for the backup. In this manner, if a device was lost or rendered inoperable, the previous backup exists. The devices should not be used for any other file work.

6.1.6 Data File Security

Many factors need to be taken into consideration to ensure the safe storage and retrieval of data files and the software programs used to create them (which, for the WSP survey, will

primarily be CSPro). The following list[1] presents issues of primary concern that should be addressed by the local survey firms:

1) Provide unique login accounts and passwords for the Data Entry Supervisor, and each Data Entry Operator.

2) Use and regularly update virus protection software on all computers involved in the data processing activities, whether or not the computer it connected to the Internet.

3) Ensure data and software recoverability in the event of emergencies.

4) Backup multiple copies in multiple secured locations on a regular basis.

5) Physical access to the data processing room should be limited to those involved with the survey.

6) Electronic access to the survey computers should ideally be kept isolated from other computers connected to the LSF's network. In addition, only the Data Entry Supervisor's computer should have access to the Internet.

7) File sharing should only occur among the LSF's Survey Management and Data Processing Staff, World Bank WSP officials, and Kimetrica, unless explicit permission is given to do so by either World Bank officials or Kimetrica.

8) During the data processing operation, should any computer need to be removed prematurely from the premises due to obsolescence, equipment failure, or seconding to another project, data must be properly removed from the computer. See Section 7.3 below for details.

6.1.7 Lab Procedures

Protocols will be presented on how the data processing staff should handle the questionnaires and any biological specimens collected.

[1] Modified from the Responsible Conduct of Research's website, located at Northern Illinois University (http://www.niu.edu/rcrportal/datamanagement/dhtopic.html).

7 Post-Processing Guidelines

This section discusses data handling procedures for the LSF, following the completion of their data entry activities. This section covers questionnaire storage, questionnaire disposal, and computer file removal.

7.1 Short-Term Questionnaire Storage

Once the LSF's data processing activities are complete, the LSF should retain the questionnaires for a fixed period of time (six to 12 months being the norm, but to be determined by the World Bank), in the event the World Bank, Kimetrica, or other agencies affiliated with the survey need to revisit them.

During this short-term storage period, questionnaires can either remain in their current location (provided the room can be secured and is not used by any person unaffiliated with the survey), or can be boxed up and stored elsewhere. Should this latter course of action be taken, then this alternate storage location must meet the following standards:

- Allow for secured lockup
- Keep the clusters in sequential order—should more than one cluster be placed within a box, then the clusters must be kept in sequential order, but also the boxes as a group must be clearly marked on the box with the cluster contents, and the boxes must be arranged within the room in sequential order.
- Must be visually accessible—a person entering the storage area must be able to visually see all boxes, so that they can quickly pinpoint the location of a given cluster.
- Must be safely stored within the room—preferably no boxes would sit on the floor (in the event of flooding), and likewise fire extinguishing materials would be readily available (i.e., fire extinguishers and/or sand with shovels).

7.2 Questionnaire Disposal

Once the World Bank has given approval for the questionnaires to be disposed of, they must be destroyed in such a way so as to maintain the confidentiality of the respondents; that is, the entire questionnaire does not have to be physically destroyed; rather, only those pages with identifying information such as person or place names/codes need be destroyed. For the WSP, this would entail removing the following pages from each household set:

- the cover sheet of the household roster
- the household roster pages
- all rostered pages within the subsequent modules that list persons by name

These pages should be shredded, burned, or otherwise rendered unreadable. The disposal method must be approved by World Bank or Kimetrica representatives. This should be done in a timely manner, i.e., within a month or two of the request. This is an critical step in the survey cycle, for retaining these paper copies when not needed after a project is over can lead to unauthorized access of confidential data.

7.3 Computer File Removal

During the data processing operation, should any computer need to be removed prematurely from the premises (for example, due to obsolescence, equipment failure, or seconding to another task), the entire folder structure that was used for the survey must be deleted, preferably using a zeroing out software program (such as Eraser) that physically removes the data from its location on the hard disk. Of course, before proceeding with this activity, it must be ensured that all data files have been properly saved to the Data Processing Supervisor's computer. If there is any doubt on the part of the LSF whether or not this has been done, they should contact the Kimetrica offices.

After the data processing operation is finished by the LSF, and at a specific time as chosen by the World Bank, and after confirmation by Kimetrica that all necessary files have been copied onto the Kimetrica FTP site and are in proper order, all computers utilized by the data entry operation should have files purged. This means the entire folder structure that was used for the survey must be deleted, preferably using a zeroing out software program (such as Eraser) that physically removes the data from its location on the hard disk.

8 Ensuring Respondent Confidentiality in the Final Datasets

Due to the nature of the WSP survey (a longitudinal survey where data was gathered at several points in time from the same household, rather than being taken at one point in time), it was necessary to collect and key in individuals' names to facilitate the matching up of households. Similarly, to facilitate field operations, other direct identifiers such as telephone numbers and street addresses were collected for the survey that may be entered into the data files. In all these instances, this information must be suppressed for the final dataset.

Should GPS data be collected for the survey, and should the final data files be made publicly available either in whole or in parts (as PUFs), then this GPS data must be either perturbed, or saved into a separate file before being suppressed from the primary survey data file.

In addition to global IRB standards, most countries have statistical laws that specify the need to preserve respondent confidentiality. In developing final guidance materials, local laws and regulations must be reviewed and incorporated into country-specific policies to ensure both respondent confidentiality, and to ensure the LSF adhere to national survey procedures.