

*Federal Republic of Nigeria*



*National Bureau of Statistics*



***Basic Information Document***

***Nigeria  
General Household Survey–Panel  
2010 - 2011***

**February 2015**

## ACRONYMS

BMGF	Bill and Melinda Gates Foundation
CAP	Computer Assisted Personal Interviewing
EA	Enumeration Area
FCT, Abuja	Federal Capital Territory, Abuja
FMA&RD	Federal Ministry of Agriculture and Rural Development
FMWR	Federal Ministry of Water Resources
GHS	General Household Survey
GHS-Panel	General Household Survey-Panel (panel subcomponent of GHS)
GHS-Cross	General Household Survey-Cross (annual cross section of GHS)
LGA	Local Government Area
LSMS-ISA	Living Standards Measurement Study – Integrated Surveys on Agriculture
MDA	Ministries, Departments, Agencies
MSF	Master Sample Frame
NASS	National Agricultural Sample Survey
NBS	National Bureau of Statistics
NFRA	National Food Reserve Agency
NISH	National Integrated Survey of Households
NHIS	National Health Insurance Scheme
NPopC	National Population Commission
PSU	Primary Sampling Unit
SAE	Small Area Estimation
WB	World Bank

## Table of Contents

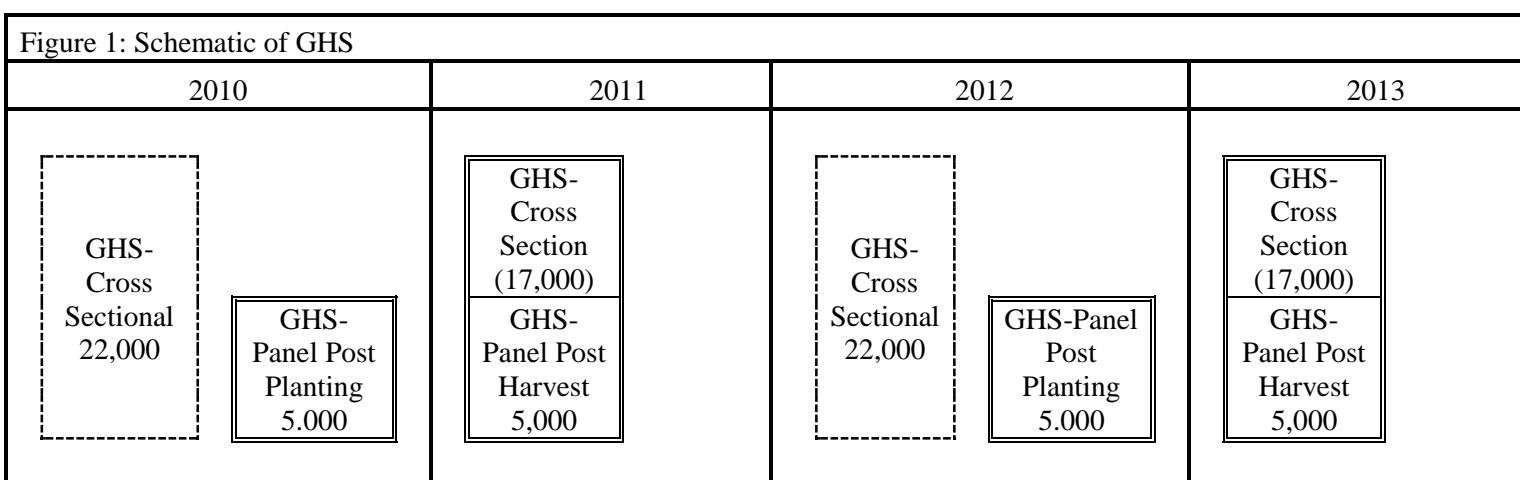
1.0	Introduction.....	1
1.1	Expected Benefits .....	2
1.2	Innovations.....	2
1.3	Coverage and Scope.....	3
2.0	The Survey Instruments .....	4
3.0	Sample Design .....	11
4.0	Training of Field Staff and Data Entry Operators for the Survey .....	13
5.0	Field Work .....	14
5.1	Organization of Fieldwork.....	14
5.2	Fieldwork Monitoring and Evaluation.....	14
6.0	Data Management and Description of Datasets.....	15
6.1	Data Management .....	15
6.11	Data Entry .....	15
6.12	Data Cleaning.....	15
6.13	Data Cleaning Challenges.....	16
6.14	Weighting of Data.....	16
6.2	Description of Datasets .....	17
6.21	Household Data.....	17
6.22	Agriculture Data.....	19
6.23	Community datasets.....	21
6.24	Geospatial variables .....	22
6.25	Consumption aggregate .....	22
7.0	Using the Data .....	26
7.1	File Structure .....	26
7.2	Merging Datasets .....	26
7.21	Household and Agriculture Datasets .....	26
7.22	Post-Planting and Post-Harvest Datasets .....	26
7.23	Community Datasets .....	26
7.3	Network Roster .....	27
7.4	Unit Measure and Conversion Factors.....	27
7.5	Tracking Data.....	30
8.0	Overall Problems and Challenges Faced During the First Wave of the Survey and Recommendations for Wave 2 of the Survey .....	31
8.1	Sampling .....	31
8.2	Field work .....	31

8.21 Pre-filling of Questionnaires .....	31
8.22 Availability of Electricity .....	31
8.23 Flooding .....	32
8.24 GPS Units.....	32
8.25 Data entry .....	32
8.26 Delay in States Responding to Data Queries .....	33
Appendix 1: How to Obtain Copies of the Data.....	34
Appendix 2: Updates to the Data .....	35
Appendix 3: Agriculture Land Conversion Factors.....	37
Appendix 4: Crop Codes.....	38
Appendix 5: Confidential Information, Geospatial Variables .....	39
Appendix 5: Description of Food Photo Files .....	5

## 1.0 Introduction

The purpose of the present document is to provide detailed information on the panel survey component of the revised General Household Survey (GHS-Panel) fielded by the National Bureau of Statistics in 2010-2011. This survey is the first round of a long-term project to collect panel data on households, their characteristics, welfare and their agricultural activities. The survey is the result of a partnership that the NBS has established with the Federal Ministry of Agriculture and Rural Development (FMA&RD), the National Food Reserve Agency (NFRA), the Bill and Melinda Gates Foundation (BMGF) and the World Bank (WB). Under this partnership, a method to collect agricultural and household data in such a way as to allow the study of agriculture's role in household welfare over time was developed. This GHS-Panel Survey responds to the needs of the country, given the dependence of a high percentage of households on agricultural activities in the country, for information on household agricultural activities *along with* other information on the households like human capital, other economic activities, and access to services and resources. The ability to follow the same households over time makes the GHS-Panel a new and powerful tool for studying and understanding the role of agriculture in household welfare over time, as it allows analyses to be made of how households add to their human and physical capital, how education affects earnings, and the role of government policies and programs on poverty, inter alia. The GHS-Panel is the first panel survey to be carried out by NBS.

The GHS survey is a cross-sectional survey of 22,000 households carried out annually throughout the country. Under the work of the partnership, a full revision of the questionnaire was undertaken and, at the same time, a sub-sample of the GHS now forms a panel survey. The panel component (GHS-Panel) applies to 5,000 households of the GHS collecting additional data on multiple agricultural activities and on household consumption. As the focus of this panel component is to improve data from the agricultural sector and link this to other facets of household behavior and characteristics, the GHS-Panel drew heavily on the Harmonized National Living Standards Survey (HNLSS – a multi-topic household survey) and the National Agricultural Sample Survey (NASS – the key agricultural survey) to create a new survey instrument to shed light on the role of agriculture in households' economic wellbeing that can be monitored over time. The first wave of the revised GHS and GHS-Panel was carried out in two visits to the panel households (post-planting visit in August-October 2010 and post-harvest visit in February-April 2011) and one visit to the full cross-section (in parallel with the post-harvest visit to the panel). The GHS-Panel will be carried out every two years while the GHS-Cross Section is usually carried out annually. A schematic of data collection is shown in Figure 1. Note that a separate document details the contents of the GHS-Cross Section. This document provides details on the GHS-Panel only.



This revised and expanded GHS survey (Cross and Panel) forms part of a larger, regional project in Sub-Saharan Africa to improve agricultural statistics. Nigeria is one of the seven countries being supported by the World Bank, through funding from the Bill and Melinda Gates Foundation (BMGF), to strengthen the production of household-level data on agriculture. The other countries are Ethiopia, Uganda, Tanzania, Malawi, Niger, and Mali. This regional project, the Living Standards Measurement Study-Integrated Surveys on Agriculture (LSMS-ISA) has the overarching objective of improving our understanding of agriculture in Sub-Saharan Africa – specifically, its role in household welfare and poverty reduction. The data will also provide insights into how innovation and efficiency can be fostered in the agricultural sector.

The present document is designed to provide an overview of the first wave of data collection under the revised GHS-Panel. The first wave consists of two visits to the household: the post-planting visit occurred directly after the planting season to collect information on preparation of plots, inputs used, labour used for planting, and other issues related to the planting season. The post-harvest visit occurred after the harvest season and collected information on crops harvested, labour used for cultivating and harvesting activities, and other issues related to the harvest cycle.

## 1.1 Expected Benefits

The specific outputs and outcomes of the revised GHS with panel component project are:

- ◆ Development of an innovative model for collecting agricultural data in conjunction with household data;
- ◆ Development of a model of inter-institutional collaboration between NBS, FMA&RD and NFRA, *inter alia*, to ensure the relevance and use of the new GHS;
- ◆ Strengthening the capacity to generate a sustainable system for producing accurate and timely information on agricultural households in Nigeria.
- ◆ Comprehensive analysis of poverty indicators and socio-economic characteristics.

## 1.2 Innovations

The revised GHS with panel component contains several innovative features.

- ◆ Integration of agricultural data at the plot level with household welfare data;
- ◆ Creation of a panel data set that can be used to study poverty dynamics, the role of agriculture in development, and the changes over time in health, education and other labor activities, *inter alia*.
- ◆ Use of small area estimation (SAE) techniques to generate state level poverty data by taking advantage of the integration of the panel households into the GHS cross-section.
- ◆ Collection of information on the network of buyers and sellers of goods with which households interact;
- ◆ Use of GPS units for measuring agricultural land areas;
- ◆ Involvement of multiple actors in government, academia and the donor community in the development of the survey and its contents as well as its implementation and analysis;
- ◆ Use of concurrent data entry in the first wave. In later waves the project will develop and implement a Computer Assisted Personal Interview (CAPI) application for the paperless collection of the GHS-Panel;
- ◆ Use of direct respondents for all sections of the questionnaires where individual level data or specific economic activity data are collected;
- ◆ Creation of publicly available micro data sets for researchers and policy makers;
- ◆ Active dissemination of agriculture statistics.

### 1.3 Coverage and Scope

The revised GHS with the panel component, while having an intensive focus on agriculture, is a national survey. The survey covered all the 36 states and the Federal Capital Territory (FCT), Abuja. Both urban and rural enumeration areas (EAs) were canvassed.

The survey covered a wide range of socio-economic topics, which were collected via three different questionnaires administered to the household and the community. These are the Household Questionnaire, the Agriculture Questionnaire and the Community Questionnaire.

## 2.0 The Survey Instruments

The survey consisted of three questionnaires for each of the visits. The *Household Questionnaire* was administered to all households in the sample. The *Agriculture Questionnaire* was administered to all households engaged in agricultural activities such as crop farming, livestock rearing and other agricultural and related activities. The *Community Questionnaire* was administered to the community to collect information on the socio-economic indicators of the enumeration areas where the sample households reside.<sup>1</sup>

***GHS-Panel Household Questionnaire:*** The Household Questionnaire provides information on demographics; education; health (including anthropometric measurement for children and child immunization); labor and time use; food and non-food expenditure; household nonfarm income-generating activities; food security and shocks; safety nets; housing conditions; assets; information and communication technology; and other sources of household income. Household location is geo-referenced in order to be able to later link the GHS-Panel data to other available geographic data sets.

***GHS-Panel Agriculture Questionnaire:*** The Agriculture Questionnaire solicits information on land ownership and use; farm labor; inputs use; GPS land area measurement and coordinates of household plots; agricultural capital; irrigation; crop harvest and utilization; animal holdings and costs; and household fishing activities. Some information is collected at the crop level to allow for detailed analysis for individual crops.

***GHS-Panel Community Questionnaire:*** The Community Questionnaire solicits information on access to infrastructure; community organizations; resource management; changes in the community; key events; community needs, actions and achievements; and local retail price information.

The Household Questionnaire is slightly different for the two visits. Some information was collected only in the post-planting visit, some only in the post-harvest visit, and some in both visits. See Section 6.21 for more details.

The Agriculture Questionnaire collects different information during each visit, but for the same plots and crops. See Section 6.22 for more details.

The Community Questionnaire collected prices during both visits, and different community level information during the two visits. See Section 6.23 for more details.

The contents of each questionnaire for the GHS-Panel post-planting and GHS-Panel post-harvest are outlined below.

---

<sup>1</sup> The Community Questionnaire does not collect information from communities in the sociological sense. The data cannot be used to represent communities in Nigeria. The data collected at the community level represent information that is common to the households selected for inclusion in the selected sample EAs.



**Table 2.1: GHS-Panel Household Wave 1 Questionnaire – Post Planting Visit**

<b>Section</b>	<b>Topic</b>	<b>Respondent</b>	<b>Description</b>
Cover	Cover	Field staff	Household identifier variables, enumerator, supervisor, and data entry clerk identifiers, date and time of interview and data entry, and observation notes by enumerator regarding the interview
1	Roster	HOUSEHOLD head or spouse.	Roster of individuals living in the household, relationship to the household head, gender, year of birth, age, marital status, spouse identification, parental status, and place of birth
2	Education	Individuals 5 years and above	Educational attainment, school characteristics, and expenditures for the 2009-10 academic year
3	Labour	Individuals 5 years and above	Labor market participation during the last seven days, wage work, and domestic activities within the home
4	Credit and Savings	Individuals 15 years and above	Savings made, loans or credit received, insurance, remittances received by the household during the last twelve months, and conditions of the transaction
5	Household Assets	HOUSEHOLD head	Ownership of assets and value
6	Non-Farm Enterprises	Owner or manager of enterprise	Enterprise ownership, status, labor, value of stock, sales, and business costs
7A	Meals Away From Home	Most knowledgeable person	Naira value of food consumed outside the home during the last seven days
7B	Household Food Expenditure	Person responsible for food purchases	Quantity and value of food consumed within the household during the last seven days
8	Household Non-Food Expenditures	Person responsible for household purchases	Non-food expenditure during the last week/last month/last six months/last 12 months
9	Food Security	HOUSEHOLD head or eligible adult	Food security status of households during the past 7 days/12 months
10	Other Income	HOUSEHOLD head or eligible adult	Others sources of household income since the new year
	Contact Information	HOUSEHOLD head or eligible adult	Contact information

**Table 2.2: GHS-Panel Agriculture Wave 1 Questionnaire – Post Planting Visit**

<b>Section</b>	<b>Topic</b>	<b>Respondent</b>	<b>Description</b>
Cover	Cover	To be completed by field staff. HOUSEHOLD ID must be copied from HOUSEHOLD to Agriculture Questionnaire.	This section contains household location and identification data as well as administrative data as regards administering and managing the questionnaire
11a	Plot Roster	Owner or manager of plot	Information on all plots owned and/or managed by the household. This section includes data on estimated area, GPS measured area, and the GPS measured location of the plot
11b	Land Inventory	Owner or manager of plot	Data on plot acquisition, tenure and use
11c	Input Costs	Owner or manager of plot	Use and cost of pesticide, herbicide, animal labor and use of machinery
11d	Fertilizer Acquisition	Owner or manager of plot	Access to, use and cost of fertilizer
11e	Seed Acquisition	Owner or manager of plot	Data on source, quantity and cost of seeds used on the plot
11f	Planted Field Crops	Owner or manager of plot	Data on crops planted on the plot, amount of crops planted, and expected harvest
11g	Planted Tree Crops	Owner or manager of plot	Details on tree crops
11h	Marketing of Agricultural Surplus	Owner or manager of plot	Marketing of agricultural surplus, quantities sold, value, and information on purchaser
11i	Animal Holdings	Farmer or caretaker of animals	Data on farm animals owned by the household and commercial activity with these animals
11j	Animal Costs	Farmer or caretaker of animals	Livestock farmer/caretaker activities and costs
11k	Agriculture By-Product	Farmer or caretaker of animals	Trading activity in agricultural by-products
11l	Extension	Owner or manager of plot	Access to and utilization of technical support from various sources (government and non-government)
12	Network Roster	Farmer, owner or manager of plot	Roster of places or businesses where the household sells and purchases agricultural produce and/or supplies

**Table 2.3: GHS-Panel Community Wave 1 Questionnaire – Post Planting Visit**

Section	Topic	Respondent	Description
Cover	Cover	To be completed by the field staff	Cover
C1	Respondents Characteristics	Community Focus Group	Respondent characteristics
C2	Food Prices	Market Food Sellers	Food prices
C3	Labor	Community Focus Group	Labor
C4	Land Prices and Credit	Community Focus Group	Land prices and credit

**Table 2.4: GHS-Panel Household Wave 1 Questionnaire – Post Harvest Visit**

Section	Topic	Respondent	Description
Cover	Cover	To be completed by the field staff	Household identifier variables, enumerator, supervisor, and data entry clerk identifiers, date and time of interview and data entry, and observation notes by enumerator regarding the interview
1	Roster	HOUSEHOLD head or spouse.	Roster of individuals living in the household, relationship to the household, gender, year of birth, age, marital status, polygamous marriages, spouse identification, parental status, place of birth, date joined household if new, migration.
2 (A and B)	Education	Individuals 5 years and above	Educational attainment, school characteristics, and expenditures. Section 2a collects information for new members of the household while Sect 2b collects follow-up and current information on original household members for the 2010-11 academic year
3 (A and B)	Labour	Individuals 5 years and older	Section 3a collects data on labour market activity of all household members 5 years and older. This includes employment and earnings information.  Section 3b collects information on employment in one or more industries in the past 6 months
4 (A and B)	Health	All individuals	Section 4a: general health status and utilization and cost of health services for those that need medical care. Data on effect of disabilities on activity and functioning; and anthropometrics.  Section 4b: Child immunization.

Section	Topic	Respondent	Description
5	Information and Communication Technology	All individuals 10 years and above	Access to and use of communication technology, including computers and internet
6	Remittances	All individuals 10 years and above	Remittances received from abroad by household members 10 years and older
7	Household Assets Sale and Acquisition	Most knowledgeable person	Household assets sale and acquisition in the past 6 months
8	Housing	Head of HOUSEHOLD or any knowledgeable adult	Housing, facilities and cost. Access to utilities and costs
9	Non-farm Enterprises and Income Generating Activities	Owner or manager of enterprise	Data on non-farm businesses owned and/or operated by the household. Follow-up data on the businesses from the previous visit and new businesses that were started since the previous visit
10 (A,B and C)	Food Consumption and Expenditure	Female in the HOUSEHOLD responsible for food preparation and/or food purchases	Section 10A: Meals away from home Section 10B: Food expenditures Section 10C: Aggregate food consumption
11	Non-Food Expenditure	Most knowledgeable person or person who is responsible for HOUSEHOLD purchases	Consumption and expenditure on non-food items
12	Food Security	HOUSEHOLD head or eligible adult	Collects information on quantity of food, preferred foods and variety of foods available to household members based on economic reasons. Also collects data on intra-household food security dynamics .
13	Other Household Income	HOUSEHOLD head or eligible adult	Miscellaneous income received by household
14	Safety Nets	HOUSEHOLD head or eligible adult	Household access to and utilization of safety nets
15(A and B)	Economic Shocks and Deaths	HOUSEHOLD head or eligible adult	Section 15a: Data on economic shocks affecting the household Section 15b: Deaths of household members in the past 12 months, including age of deceased and cause of death.

**Table 2.5: GHS-Panel Agriculture Wave 1 Questionnaire – Post Harvest Visit**

<b>Section</b>	<b>Topic</b>	<b>Respondent</b>	<b>Description</b>
Cover	Cover	To be completed by field Staff. HOUSEHOLD ID must be copied from HOUSEHOLD to Agriculture Questionnaire.	Household location and identification data as well as administrative data as regards administering and managing the questionnaire
A1	Land and Dry Season Planting	Farmer, owner or manager of plot	Follow-up on use of land for in post-planting visit and data on any subsequent planting or other use of the plot. Also information collected on new plots (i.e. added since post-planting visit)
A2	Harvest Labour	Farmer, owner or manager of plot	Data on labour that was used for crop harvesting, both from household and hired
A3	Agricultural Production - Harvest of Field and Tree Crops	Farmer, owner or manager of plot	Quantity and value of field crops produced
A4	Agricultural Capital	Farmer, owner or manager of plot	Ownership and value of agricultural machinery and tools owned by the household
A5 (A and B)	Extension Services	Farmer, owner or manager of plot	Access to and utilization of technical support from various sources (government and non-government)
A6	Animal Holdings	Owner or caretaker of animals	Data on farm animals owned by the household and commercial activity with these animals
A7	Animal Costs	Owner or caretaker of animals	Expenditure on livestock
A8	Other Agricultural Income	Farmer or caretaker of animals	Income from sale of agricultural products not captured by previous section under crops and livestock
A9 (A and B)	Fishing, Capital and Revenue	Owner of fishing operations	SectionA9a: Data on fishing activities, includes capture, harvesting and processing  Sectiona9b: Data on boat usage and the use of hired labour
A10	Network Roster	Farmer, owner or manager of plot	Roster of places or businesses where the household sells and purchases agricultural produce and/or supplies

**Table 2.6 GHS-Panel Community Wave 1 Questionnaire – Post-Harvest Visit**

<b>Section</b>	<b>Topic</b>	<b>Respondent</b>	<b>Description</b>
Cover	Cover	To be completed by the field staff	Cover
C1	Respondents Characteristics	Community Focus Group	Respondents Characteristics
C2	Community Infrastructure and Transportation	Community Focus Group	Community Infrastructure and Transportation
C3	Community Organizations	Community Focus Group	Community Organizations
C4	Community Resource Management	Community Focus Group	Community Resource Management
C5	Community Changes	Community Focus Group	Community Changes
C6	Community Key Events	Community Focus Group	Community Key Events
C7	Community Needs, Actions, and Achievements	Community Focus Group	Community Needs, Actions, and Achievements
C8	Food Prices	Market Food Sellers	Food Prices

### 3.0 Sample Design

The sample is designed to be representative at the national level as well as at the zonal (urban and rural) level. The sample size of the GHS-Panel (unlike the full GHS) is not adequate for state-level estimates.

The sample is a two-stage probability sample:

First Stage:

The Primary Sampling Units (PSUs) were the Enumeration Areas (EAs). These were selected based on probability proportional to size (PPS) of the total EAs in each state and Federal Capital Territory (FCT), Abuja and the total households listed in those EAs. A total of 500 EAs were selected using this method.

Second Stage:

The second stage was the selection of households. Households were selected randomly using the systematic selection of ten (10) households per EA. This involved obtaining the total number of households listed in a particular EA, and then calculating a Sampling Interval (S.I.) by dividing the total households listed by ten (10). The next step was to generate a random start 'r' from the table of random numbers which stands as the 1<sup>st</sup> selection. Consecutive selection of households was obtained by adding the sampling interval to the random start.

Determination of the sample size at the household level was based on the experience gained from previous rounds of the GHS, in which 10 households per EA are usually selected and give robust estimates.

In all, 500 clusters/EAs were canvassed and 5,000 households were interviewed. These samples were proportionally selected in the states such that different states had different samples sizes. The distribution of the samples are shown in Table 3.1 below which shows the size of the sample in each state, by geopolitical zone and urban/rural break-out.

Households were not selected using replacement. Thus the final number of household interviewed was slightly less than the 5,000 eligible for interviewing. The final number of households interviewed was 4,986 for a non-response rate of 0.3 percent. A total of 27,533 household members were interviewed. In the second, or post-harvest visit, some household had moved as had individuals, thus the final number of households with data in both points of time (post-planting and post-harvest) is 4,851, with 27,993 household members.

**Table 3.1 Distribution of Final Sample of 500 EAs and 5,000 Households for Panel Survey by State, Urban and Rural Sectors, within Each Zone**

Zone	State	Total		Urban		Rural	
		No. EAs	No. Hhs.	No. EAs	No. Hhs.	No. EAs	No. Hhs.
North-Central Zone	Benue	16	160	2	20	14	140
	Kogi	12	120	4	40	8	80
	Kwara	12	120	6	60	6	60
	Nasarawa	7	70	1	10	6	60
	Niger	18	180	4	40	14	140
	Plateau	11	110	2	20	9	90
	FCT Abuja	4	40	3	30	1	10
North-East Zone	Adamawa	12	120	1	10	11	110
	Bauchi	17	170	3	30	14	140
	Borno	21	210	5	50	16	160
	Gombe	8	80	1	10	7	70
	Taraba	9	90	0	0	9	90
	Yobe	13	130	3	30	10	100
North-West Zone	Jigawa	13	130	2	20	11	110
	Kaduna	12	120	4	40	8	80
	Kano	20	200	3	30	17	170
	Katsina	18	180	3	30	15	150
	Kebbi	10	100	1	10	9	90
	Sokoto	8	80	2	20	6	60
	Zamfara	9	90	2	20	7	70
South-East Zone	Abia	11	110	4	40	7	70
	Anambra	22	220	12	120	10	100
	Ebonyi	14	140	1	10	13	130
	Enugu	14	140	3	30	11	110
	Imo	19	190	2	20	17	170
South-South Zone	Akwa-Ibom	15	150	4	40	11	110
	Bayelsa	7	70	1	10	6	60
	Cross River	13	130	3	30	10	100
	Delta	14	140	4	40	10	100
	Edo	10	100	5	50	5	50
	Rivers	21	210	8	80	13	130
South-West Zone	Ekiti	8	80	6	60	2	20
	Lagos	17	170	16	160	1	10
	Ogun	11	110	7	70	4	40
	Ondo	13	130	6	60	7	70
	Osun	18	180	14	140	4	40
	Oyo	23	230	15	150	8	80



## **4.0 Training of Field Staff and Data Entry Operators for the Survey**

Two levels of training were mounted for the survey. The first level was organized at NBS Headquarters in Abuja and was called the Training of Trainers (TOT). The participants in the TOT became the resource persons for the next level of training. The top management team of the survey participated in the TOT, which lasted for three days. The persons trained in the TOT were then sent to carry out the second level training. Three (3) resource persons were sent to each of the six training centers to perform the training: Minna (Niger State) for North Central (NC), Gombe (Gombe State) for North East (NE), Kaduna (Kaduna State) for North West (NW), Enugu (Enugu State) for South East (SE), Calabar (Cross River State) for South-South (SS) and Ibadan (Oyo State) for South West. Included in the team was one (1) additional resource person per training venue who served as an Information Technology (IT) trainer.

The second level training took six days with four days dedicated to theory and two days to field practice and review. The core training materials for the 2<sup>nd</sup> level training were harmonized and finalized during the TOT. Participants in the training were Zonal Controllers, State Officers, Field Supervisors, Field Interviewers, and Data Entry Operators. Training instructions were given to the field staff by the resource persons from the management team (NBS, FMS&RD, and NFRA) with support from World Bank technical missions. The training consisted of (i) classroom instructions on the questionnaire, concepts and definitions, (ii) interview techniques, and (iii) methods and field practices in performing actual interviews to ensure that field interviewers fully understood the questionnaire. In addition, participants did actual interviews in the field with households that were not scheduled to be part of the actual survey sample. Most of the training instructions are detailed in the interviewer's and supervisor's manuals which are also available.

At the end of the training session, trainees were assessed according to both a test that was administered on the material covered in the training process, and an evaluation by the resource persons. The data entry operators were trained along with the field staff, with supplementary IT training sessions. At the end of the training, field teams were formed of interviewers, supervisors and data entry operators.

## **5.0 Field Work**

### **5.1 Organization of Fieldwork**

Data were collected by teams consisting of a supervisor, between 2 and 4 interviewers, and a data entry operator. The number of teams varied from state to state depending on the sample size or number of EAs selected. The teams moved in a roving manner and data collection lasted for between 20 – 30 days for each of the post-planting and post-harvest visits. Additional details on the structure of the visits are available in Section 6.

### **5.2 Fieldwork Monitoring and Evaluation**

As an additional aid to ensure the good quality of the data collected, extensive monitoring was done on the fieldwork. Monitoring and evaluation guidelines and formats for fieldwork were developed. One (1) monitor was assigned to 1 – 2 states and all the states and FCT, Abuja were covered. There were three levels of monitoring and evaluation, with the first and the third levels carried out by NBS state officers and zonal controllers while the second level was carried out by the technical team, which included individuals from the National Bureau of Statistics (NBS), the Federal Ministry of Agriculture and Rural Development (FMA&RD), the National Food Reserve Agency (NFRA) headquarter staff, and World Bank officials and consultants.

The monitors made sure that proper compliance with the procedures as contained in the manual were followed, effected necessary corrections and tackled problems that arose. The monitoring exercise was arranged such that the first level took place at the commencement of the fieldwork, and the third level no later than a week before the end of the data collection exercise. In-between these two periods of time, the technical team visited all the states of the federation and FCT, Abuja. While NBS state officers monitored in their state, the zonal controllers monitored in at least two (2) states (the zonal headquarters state and one other state of the same zone). The 1<sup>st</sup> and 3<sup>rd</sup> rounds of the monitoring exercise lasted for eight (8) days while the 2<sup>nd</sup> round by the technical team lasted for seven (7) days. Monitoring instruments were developed and discussed during both training of trainers and zonal training.

## **6.0 Data Management and Description of Datasets**

### **6.1 Data Management**

#### **6.11 Data Entry**

This survey used concurrent data entry approach. In this method, the fieldwork and data entry were handled by each team assigned to the state. Each team consisted of a field supervisor, 2-4 interviewers and a data entry operator. Immediately after the data were collected in the field by the interviewers and supervisors (the supervisors administered the community questionnaires and collected data on prices), the questionnaires were handed over to the supervisor to be checked and documented. At the end of each day of fieldwork, the questionnaires were then passed to the data entry operator for entry. After the questionnaires were entered, the data entry operator generated an error report which reported issues including out of range values and inconsistencies in the data. The supervisor then checked the report, determined what should be corrected, and decided if the field team needed to revisit the household to obtain additional information. The benefits of this method are that it allows one to:

- ◆ Capture errors that might have been overlooked by a visual inspection only,
- ◆ Identify errors early during the field work so that if any correction required a revisit to the household, it could be done while the team was still in the EA

The CSPro software was used to design the specialized data entry program that was used for the data entry of the questionnaires.

#### **6.12 Data Cleaning**

The data cleaning process was done in a number of stages. The first step was to ensure proper quality control during the fieldwork. This was achieved in part by using the concurrent data entry system which was, as explained above, designed to highlight many of the errors that occurred during the fieldwork. Errors that are caught at the fieldwork stage are corrected based on re-visits to the household on the instruction of the supervisor. The data that had gone through this first stage of cleaning was then sent from the state to the head office of NBS where a second stage of data cleaning was undertaken.

During the second stage the data were examined for out of range values and outliers. The data were also examined for missing information for required variables, sections, questionnaires and EAs. Any problems found were then reported back to the state where the correction was then made. This was an ongoing process until all data were delivered to the head office.

After all the data were received by the head office, there was an overall review of the data to identify outliers and other errors on the complete set of data. Identified problems were reported to the state. There the questionnaires were checked and where necessary the relevant households were revisited and a report sent back to the head office with the corrections.

The final stage of the cleaning process was to ensure that the household- and individual-level data sets were correctly merged across all sections of the Household Questionnaire. Special care was taken to see that the households included in the data matched with the selected sample and where there were differences these were properly assessed and documented. The agriculture data were also checked to ensure that the plots identified in the main sections merged with the plot information identified in the other sections. This was also done for crop-by-plot information as well.

### 6.13 Data Cleaning Challenges

The cleaning process at the head office was impeded by the fact that the questionnaires were not immediately available for inspection when problems were identified in the data. The questionnaires were retained by the state in case there was the need for household revisits. So whenever problems were identified at the head office, the state office had to be contacted in order to determine if the suspect data were the same as the information on the questionnaire, and to ensure that changes were captured in both places. This was a very cumbersome and time consuming process since communication was difficult and in many instances the response was not timely. However, this is a necessary process to ensure that the households can be re-visited to provide the correct information to avoid having to make imputations. Also, this process allows the state officers to understand the key issues that arose during fieldwork and will serve to enhance further rounds of data collection. It will be important, nonetheless, to find a mechanism to facilitate this process in the next round of data collection and cleaning.

A second challenge in data management and cleaning was the difficulty faced by state offices in sending the data from the state to the head office. There were difficulties in accessing internet facilities in many of the EAs and surrounding areas where the field teams were active. The consequence of this was that the data were not sent to the head office until the teams returned to state capitals where, due to the distance, it was difficult to return to the EAs for household revisits when requested by the head office. This issue will need to be addressed for future rounds of the survey.

### 6.14 Weighting of Data

When a sample of households is selected for a survey, these households represent the entire population of the country. To accurately use the datasets, the data must be weighted to reflect the distribution of the full population in the country. A population weight was calculated for the panel households. This weight variable (*wght*) has been included in the household dataset: Section A (*secta\_plantingw1* for post-planting and *secta\_harvestw1* for post-harvest). When applied, this weight will raise the sample households and individuals to national values adjusting for population concentrations in various areas.

## 6.2 Description of Datasets

The first wave of the GHS-Panel was administered in two visits: post-planting (Aug-Oct 2010) and post-harvest (Feb-Apr 2011). During each visit two questionnaires were administered to the household respondents (Household Questionnaire and Agricultural Questionnaire) and a third questionnaire was administered at the level of the enumeration area (Community Questionnaire). The datasets are organized by visit and then by questionnaire. The household dataset corresponds to the Household Questionnaire, the agricultural dataset corresponds to the Agriculture Questionnaire, and the community dataset corresponds to the Community Questionnaire.

The data file naming scheme is a combination of the prefix '*sect*', followed by section number, and then followed by suffix '*plantingw1*' for post-planting data and '*harvestw1*' for post-harvest data. For example, the data set that corresponds with section 1 of the Household Questionnaire in the post-planting data folder is called '*sect1\_plantingw1*'. The exception to this rule are sections where the files are broken down even further due to different reference period or different levels of recording the data. An example is section 8 of the household post-planting questionnaire on nonfood expenditure where the section is split into 5 files with each file corresponding with the reference period collected in the section. In this case, the name of the corresponding files will be '*sect81\_plantingw1*', '*sect82\_plantingw1*', etc.

### 6.21 Household Data

In the Household Questionnaire, some of the modules were administered in both the post planting and post-harvest visit and others were only administered during one of the two visits. This should be taken into account when using the datasets.

Group 1: These modules are administered in both visits but the module in the post-harvest version is a follow up to the post-planting module. For example, for the Roster and Education modules, additional information is gathered during the post-harvest only for individuals who had joined the household since the first, or post-planting, visit. (These people are referred to as 'new household members'.) For assets and enterprises, the module attempts to update the information from the first visit.

- Roster
- Education
- Household Assets
- Nonfarm Enterprises

Group 2: These modules are administered in both visits as standalone modules. For these topics we have complete information at two points in time during the year of the survey.

- Labor
- Meals Away From Home
- Food Consumption and Expenditure
- Nonfood Expenditure
- Food Security
- Other Household Income

Group 3: These modules only appear in either the post-planting or the post-harvest visit

- Post-planting only
  - Credit and Savings
- Post-harvest only
  - Health and Child Immunization
  - Information and Communication Technology
  - Remittances
  - Housing
  - Aggregate Food Consumption
  - Safety Nets
  - Economic Shocks and Deaths

Tables 6.1a and 6.1b show the sections of the Household Questionnaire and the datasets that correspond to these. There are 16 data files in the post-planting household data folder, which include all the modules in the questionnaire.

**Table 6.1a: Post-planting household datasets**

Section	Section Name	Dataset Filename
Cover	Cover	secta_plantingw1
1	Roster	sect1_plantingw1
2	Education	sect2_plantingw1
3	Labour	sect3_plantingw1
4	Credit and Savings	sect4_plantingw1
5	Household Assets	sect5_plantingw1
		sect5b_plantingw1
6	Non-farm Enterprises	sect6_plantingw1
7A	Meals Away From Home	sect7a_plantingw1
7B	Household Food Expenditure	sect7b_plantingw1
8	Household Non-Food Expenditures	sect81_plantingw1
		sect82_plantingw1
		sect83_plantingw1
		sect84_plantingw1
		sect85_plantingw1
9	Food Security	sect9_plantingw1
10	Other Income	sect10_plantingw1

There are 26 data files in the post-harvest household data folder, which include all the modules in the questionnaire.

**Table 6.1b: Post-harvest household datasets**

Section	Section Name	Dataset Filename
Cover	Cover	secta_harvestw1
1	Roster	sect1_harvestw1
2A	Education- New Member	sect2a_harvestw1
2B	Education – Original Household Members	sect2b_harvestw1
3A	Labour	sect3a_harvestw1
3B	Labour Activity	sect3b_harvestw1
4A	Health	sect4a_harvestw1
4B	Child Immunization	sect4b_harvestw1
5	Information and Communication Technology	sect5_harvestw1
6	Remittances	sect6_harvestw1
7	Household Assets Sales and Acquisition	sect7_harvestw1
8	Housing	sect8_harvestw1
9	Non-farm Enterprises and Income Generating Activity	sect9_harvestw1
10A	Meals Away From Home	sect10a_harvestw1
10B	Food Consumption and Expenditures	sect10b_harvestw1
10C	Aggregate Food Consumption	sect10c_harvestw1
11	Non-food Expenditures	sect11a_harvestw1
		sect11b_harvestw1
		sect11c_harvestw1
		sect11d_harvestw1
		sect11e_harvestw1
12	Food Security	sect12_harvestw1
13	Other Household Income	sect13_harvestw1
14	Social Safety Nets	sect14_harvestw1
15A	Economic Shocks	sect15a_harvestw1
15B	Deaths	sect15b_harvestw1
		sect15b1_harvestw1

## 6.22 Agriculture Data

It should be noted that in the Agriculture Questionnaire, the **plot roster and land inventory information collected during the post-planting visit is updated during the post-harvest visit in the Land and Dry Season Planting section to include additional plots households may have acquired or old plots they have disposed of since the first, post-planting visit.** Information on inputs to agricultural production was collected in the post-planting visit only. The crop codes used in the Agriculture Questionnaire are presented in Appendix 3. As with the Household Questionnaire, some modules were administered in both visits. For these modules, during the post-harvest visit, information was gathered on the activities since the post-planting interview.

There are 15 data files in the post-planting agriculture data folder corresponding to the modules in the questionnaire.

**Table 6.2a: Post-planting Agriculture datasets**

Section	Section Name	Dataset Filename
11a	Plot Roster	sect11a_plantingw1
		sect11a1_plantingw1
11b	Land Inventory	sect11b_plantingw1
11c	Input Costs	sect11c_plantingw1
11d	Fertilizer Acquisition	sect11d_plantingw1
11e	Seed Acquisition	sect11e_plantingw1
11f	Planted Field Crops	sect11f_plantingw1
11g	Planted Tree Crops	sect11g_plantingw1
11h	Marketing of Agricultural Surplus	sect11h_plantingw1
11i	Animal Holdings	sect11i_plantingw1
11j	Animal Costs	sect11j_plantingw1
11k	Agriculture By-product	sect11k_plantingw1
11l	Extension	sect11l1_plantingw1
		sect11l2_plantingw1
12	Network Roster	sect12_plantingw1

There are 15 data files in the post-harvest agriculture data folder corresponding to the modules in the questionnaire.

**Table 6.2b: Post-harvest Agriculture datasets**

Section	Section Name	Dataset Filename
A1	Land and Dry Season Planting	sectaa_harvestw1
		secta1_harvestw1
A2	Harvest Labour	secta2_harvestw1
A3	Agricultural Production – Harvest of Field and Tree Crops	secta3_harvestw1
A4	Agricultural Capital 1	secta41_harvestw1
	Agricultural Capital 2	secta42_harvestw1
A5	Extension Services 1	secta5a_harvestw1
	Extension Services 2	secta5b_harvestw1
A6	Animal Holdings	secta6_harvestw1
A7	Animal Costs	secta7_harvestw1
A8	Other Agricultural Income	secta8_harvestw1
A9a	Fishing	secta9a1_harvestw1
		secta9a2_harvestw1
A9b	Fishing Capital and Revenue	secta9b1_harvestw1
		secta9b2_harvestw1
A10	Network Roster	secta10_harvestw1



## 6.23 Community datasets

Tables 6.3a and 6.3b show the sections of the community questionnaire and their corresponding data sets. There are 8 files in the post-planting data folder and 9 data files in the post-harvest community data folder corresponding to the modules in the questionnaire.

**Table 6.3a: Post-planting Community datasets**

Section	Section Name	Dataset Filename
Cover	Cover	sectc_plantingw1
C1	Respondent Characteristics	sectc1_plantingw1
C2	Food Prices	sectc2_plantingw1
C3	Labor	sectc3a_plantingw1 sectc3b_plantingw1 sectc3c_plantingw1
C4	Land Prices and Credit	sectc4a_plantingw1 sectc4b_plantingw1

**Table 6.3b: Post-harvest Community datasets**

Section	Section Name	Dataset Filename
Cover	Community identification	sectc_harvestw1
C1	Respondent Characteristics	sectc1_harvestw1
C2	Community Infrastructure and Transportation	sectc2_harvestw1
C3	Community Organizations	sectc3_harvestw1
C4	Community Resource Managements	sectc4_harvestw1
C5	Community Changes	sectc5_harvestw1
C6	Community Key Events	sectc6_harvestw1
C7	Community Needs, Actions and Achievements	sectc7_harvestw1
C8	Food Prices	sectc8_harvestw1

Note that, for purposes of maintaining the confidentiality of the data, all names and addresses have been removed from the datasets. Additionally, the GPS coordinates have also been removed as these could be used to locate households and plots with accuracy. Various approaches to use of the GPS data are available: the user who is interested in these data is requested to contact NBS directly to discuss how these can be used. (See Appendix 1 for information on how to contact NBS.)

## 6.24 Geospatial variables

To increase the use of the GHS-Panel data, a set of geospatial variables has been provided by using the georeferenced plot and household locations in conjunction with various geospatial databases that were available to the survey team. More information is available in Appendix 4 on how these variables are constructed and linked to the GHS-Panel data. The table in Appendix 4 provides the name, type, source, reference period, resolution, description, and source of each geospatial variable included.

## 6.25 Consumption aggregate

The consumption aggregate is constructed using Stata executable programs (do-files), which are available on request.

As noted above, the survey was implemented in two visits – the first visit occurred in the post planting season and the second in the post-harvest season. Consumption and/or expenditure information was collected in both visits. The steps used to calculate the consumption aggregate are applied uniformly across both the 2010-11 and 2012-13 surveys. The consumption aggregate includes (a) food expenditures (including meals eaten outside the home); (b) non-monetary food consumption resulting from consumption of home production, gifts and in-kind payments; (c) education expenditures; (d) health expenditures; and (e) housing expenditures, including imputed rent.

A consumption aggregate file was created from the data for each visit. These files have been combined to create an overall annual consumption file which is being distributed with the data. The individual visit files are available upon request.

### Food

The food component of the consumption aggregate includes food consumption from all possible sources: food eaten away from home (such as from kiosks or on the street), food purchases, food from own-production, and food received as gifts or in-kind income. The recall period for all of the food sources was last the 7 days.

In the case where food consumption was reported in non-standard units, such as cigarette cup or olodo, units are converted using files that contain conversion units for most units reported by respondents.

For those non-standard units which are not included in the conversion file, total consumption of the food item (in kg) was imputed from the Local Government Area (LGA) median value if more than 20 observations are found in the LGA. If LGA values could not be used, the median value as calculated at the state level was used – again if more than 20 observations are found in the state. Finally, the national median was used if the lower level units could not be used.

Median prices per kilo for each item within each of the areas (LGA, state, zone) are calculated as long as at least 50 observations are seen at the area level. Using these median regional prices, the total expenditure for each item was obtained. From this value, group-wise expenditure for items is obtained. Outliers are replaced with the national median expenditure per item.<sup>2</sup>

Total household food expenditure (fdtexp) is obtained from the total expenditure on purchased foods (fdtotby) and the total value of own production consumed by household members (fdtotpr).

Households reporting 0 food expenditure are dropped from the sample.

## Non-food

There are different recall periods used in the collection of the non-food data depending on how often the purchases are likely to be made: 1, week, 1 month, 6 months and 12 months.<sup>3</sup>

Weekly expenditures are recorded for non-food items which are purchased frequently. Median expenditures per item are calculated for the month of the interview and the zone. Outliers, observations that are 3 standard deviations away from the mean, are replaced with the median.

A similar method was used to calculate monthly, bi-annual and annual expenditures as was used for weekly expenditures. Median expenditures per item are calculated and outliers are dealt with on a per capita basis and replaced by the median.

Total household non-food expenditure (nfdtexp) is obtained from the sum of expenditures on non-food items (hlftexp nfdftexp nfditexp nfdtexp hhtexp).

## Education

Education expenditures are self-reported values given in the education module. Reported expenditures on different scholastic items are included: tuition, book purchases, extra-curricular activities, food and boarding, transport, fees, insurance, and other expenditures related to education. If the respondent was unable to provide information for the individual expenditure items, total household expenditure on education was recorded. This value is used in all cases where it is not missing. If the reported total expenditure on education is missing, the sum of the individual components is used.

## Health

Health expenditures are collected primarily in the post-harvest visit. There was no health module in the post-planting visit, although some health expenditures are collected in the non-food expenditure module. The expenditures include consultation fees, medication, hospitalization, transport to hospital, health insurance, therapeutic equipment, and expenditures on health not mentioned elsewhere.

---

<sup>2</sup> Outliers are defined as values 3 standard deviations away from the median.

<sup>3</sup> See the questionnaires for details.

## Housing

Housing expenditures include payments for electricity, landline telephones, cell phones, drinking water, garbage disposal, and estimated rent. Values for phone and electricity expenditures are reported for different time spans (monthly, weekly, annually) and are adjusted accordingly. Outliers are dealt with on a zone basis. Values are imputed with the zone median if the value is 3 standard deviations away from the zone mean.

Households that actually pay rent are used to estimate an hedonic model. Because most rural households have no rent values, estimated rent was based on mostly urban households. Even using urban households, less than 20 percent of households reported actual rents. The model used the natural log of rent paid in Naira as the dependent variable and various characteristics of the housing such as composition of the walls and floors; amenities such as connection to water, type of cooking fuel used, type of lighting used; and location (rural-urban, regions, or ecological zones) to obtain coefficients. Subsequently predicted rent values are obtained for households that did not report rent but had reported dwelling characteristics.

In the final analyses, households whose actual rent are outliers are replaced by the estimates from the model.

## Totals

### *Non-food expenditure*

Non-food expenditure totals are obtained from the sum of the total expenditure on education (edtex), expenditure on frequent non-food items (nfdftexp), and expenditure on infrequently consumed items (nfditexp).

### *Food expenditures*

Total food expenditure is obtained by adding the total food expenditure originating from purchases and value of consumption of own goods.

### *Household expenditure*

Total household expenditure is calculated by adding total food and non-food expenditures. Per capita expenditure is obtained by dividing total expenditure by the number of members in the household.

### *Regional Deflators*

All households in the 40th percentile are used to create a food basket. Using this population, the share of expenditure on each food item is estimated. Consumption data are then deflated spatially using zone level price deflators. Deflators are built using zone level baskets and monthly priced using item level prices at the state levels.

### *Aggregated Expenditures*

The aggregated file for the GHS-Panel2010 is called cons\_agg\_w1 (w1 indicating “wave 1” for the first wave of the panel survey), and the aggregated file for the GHS-Panel2012 is called cons\_agg\_w2 (w2 indicating “wave 2” for the second wave of the panel survey).

Since households are interviewed twice for each survey, post planting (between August and October) and post-harvest (between February and April), consumption information is averaged over the two visits to proxy for the household’s annual expenditure pattern. Thus the aggregated variables are only relevant for households present in both visits during each wave and only those households are included in the aggregate files.

In order to construct the aggregated expenditure per each wave, the regionally deflated expenditures are adjusted to 2010 PPP prices. Using these adjusted values the expenditure for each of the two waves is estimated as the mean expenditure between the post-planting and post-harvest visits.

The decision on which figures to use are left to the end user, and thus values for each visit as well as the wave aggregate values are present in each dataset. Within each file the relevant variables are labeled accordingly, the variables ending in PP come from post-planting figures, PH comes from post-harvest, and w1 or w2 indicates the mean value between the two visits made in that wave.

## 7.0 Using the Data

### 7.1 File Structure

The data should always be used in conjunction with the questionnaire and the interviewer's instruction manual. Where there are no issues of confidentiality all the variables from the questionnaire have been included in the data sets. In some cases there is an additional variable which contains the "other specify" information that was written in the questionnaire. So, for example, if there is a variable with two parts question 5a and question 5b, a third variable, question 5c, might be added which would contain the other "specify information".

### 7.2 Merging Datasets

#### 7.21 Household and Agriculture Datasets

All household and agriculture datasets in both the post-planting and post-harvest files contain a variable (*hhid*), which is a unique identifier for the household. This variable is used as the unique key variable in the merging of all household type datasets. In some of the other types of datasets, additional key variables may be required in the merging process. In the case of individual type files, the variable that uniquely identifies the individual in the household is *indiv*. So in order to merge any two individual type files, both the variables *hhid* and *indiv* would be used. In the agriculture datasets, plot files are merged using *hhid* and *plotid* while crop files are merged using *hhid*, *plotid* and *cropid*.

#### 7.22 Post-Planting and Post-Harvest Datasets

Post-planting and post-harvest files can be merged using the methodology explained above. That is, the *hhid* is the same for a specific household in the post-planting and post-harvest visit. It should be noted that there was **some attrition of households between the post-planting and post-harvest visits so some households in the post-planting files will not have a match in the post-harvest data sets.** Note also that people may have left the households or joined them in the time between the two visits. Thus the number of people will vary between visits.

#### 7.23 Community Datasets

The community questionnaire is administered at the EA level so the location variables *lga* and *ea* are unique for each community questionnaire. Merging of community files within the round or with community files from the other round or with any of the household or agriculture files from either round should be done using the *lga* and *ea variables*, in that order.

Location variables: *zone*, *state*, *lga*, *sector*, *ea* and *ric* have not been included in all the datasets. Instead, these variables have been included in the questionnaire cover datasets, i.e.

secta\_harvestw1, secta\_plantingw1, sectc\_harvestw1 etc., and from there they can be merged into any of the other datasets using the key variables as explained above.

## 7.3 Network Roster

A network roster is included in both the post-planting and post-harvest agriculture questionnaires. The network roster keeps a record of the list of places (businesses, markets, persons etc.) with which the household engages in agricultural trading activities. Each place is assigned the network code of the line in which it is in that section. Each place is recorded only once so we have for example, network codes N1, N2 etc. which is just a serialization of the places. This is similar to the household roster where an individual acquires the individual code of the line in which the person's name is written.

After the information has been entered in the network roster, the network code can be used in any section of the Agriculture Questionnaire where a place of trading is requested. The network roster contains information on the type of place and its location.

## 7.4 Unit Measure and Conversion Factors

This sub-section describes the unit of measurement for food items and conversion factors employed in the data collection. There were some issues encountered in Wave 1 post-planting visit regarding the units reported in the food consumption module and local, non-standardized units for quantities. Some of these issues were addressed in Wave 1 post-harvest visit. The food consumption module is in Section 7B in post-planting questionnaire and Section 10B in post-harvest questionnaire.

The post-planting questionnaire (visit 1) included "pieces" and "other, specify" as possible units of measure for food (see Figure 7.1 below).

Figure 7.1: Question 2 - Post-planting Questionnaire Section 7B

I T E M  C O D E	1 Within the <u>past 7 days</u> , did the members of this household eat/drink any of this [ITEM] within the household?	2. How much in total did your household consume of this [ITEM] in the <u>past 7 days</u> ?		3. How much did the household purchase of this [ITEM] during the <u>past 7 days</u> ?		4. How much did your household spend on this [ITEM] during the past 7 days?		5. How much of this [ITEM] came from own production during the past 7 days?		6. How much of this [ITEM] came from gifts and other sources during the past 7 days?	
	PLEASE ONLY LIST ITEMS CONSUMED WITHIN THE HOUSEHOLD AND EXCLUDE FOOD CONSUMED OUTSIDE	KILOGRAMS.....1 GRAMS.....2 LITRE.....3 MILLILITRE.....4 PIECES.....5 OTHER (SPECIFY).....6		IF NONE WRITE 0 FOR QUANTITY AND LEAVE UNIT BLANK ► Q5  KILOGRAMS.....1 GRAMS.....2 LITRE.....3 MILLILITRE.....4 PIECES.....5 OTHER (SPECIFY).....6		THIS QUESTION REFERS TO THE QUANTITY IN QUESTION 3		IF NONE WRITE 0 FOR QUANTITY AND LEAVE UNIT BLANK  KILOGRAMS.....1 GRAMS.....2 LITRE.....3 MILLILITRE.....4 PIECES.....5 OTHER (SPECIFY).....6		EXCLUDE FOOD TAKEN OUTSIDE THE HOUSEHOLD  IF NONE, WRITE 0 FOR QUANTITY AND LEAVE UNIT BLANK  KILOGRAMS.....1 GRAMS.....2 LITRE.....3 MILLILITRE.....4 PIECES.....5 OTHER (SPECIFY).....6	
	ASK THIS QUESTION FOR ALL ITEMS, BEFORE MOVING ON TO THE NEXT QUESTIONS FOR ITEMS WITH YES  YES...1 NO...2 (► NEXT ITEM)										
		QUANTITY	UNIT	QUANTITY	UNIT	NAIRA		QUANTITY	UNIT	QUANTITY	UNIT

The "other specify" units used in reporting quantities of purchases, home production, and gifts were predominantly:

- Milk Cup
- Bottle
- Tin
- Sachet
- Pack
- Mudu
- Wrap
- Loaf
- Basin
- Bunch
- Bundle
- Paint Rubber
- Tuber
- Heap

They posed a significant problem when it came to arriving at a value, especially for home production and gift items since no estimated value was collected for reported quantities produced or received. Values of quantities produced or received in non-standard units could have been established if the household reported the purchase of item in these units, but in many cases the household did not report purchase of the items, or where purchased, the quantities were not in the specified units. In cases where the items were purchased in kilograms or liters, the absence of conversion factors to non-standard, or other specified, units also made it difficult to assign values.

It was also noted that other local units such as congo, derika and tiya did not appear as "other specified" units so it is assumed that there was some attempt at conversion to standards units by the interviewers or supervisors in the field.

The "pieces" unit was also problematic because there was no reliable way to convert these quantities to standard units. More specifically, when an item quantity was recorded as a "piece" there was no way of knowing whether this was a small, medium or large piece or how this piece would translate into a standard weight for purposes of valuing home consumption and gifts.

In order to address these challenges, a number of changes were made in the post-harvest questionnaire (visit 2):

- "pieces" and "other specify" were removed as units of measure.
- the only units that were on the questionnaire were kilograms, grams, liters, and milliliters.
- photographs were taken of commonly consumed food items with the measured standard weight or volume indicated.

The photographs were of items that were frequently reported in "pieces" or in local units in the post-planting survey. The interviewers were instructed to first ask respondents to report the quantity in standard units. If this was not possible they were instructed to show the respondent the



photographs so that they could identify the item, depicted by size or quantity, that they purchased or consumed. Each picture had an indicated weight or volume in standard units, and the interviewer recorded the weight (or volume) or multiples of measurement of what was shown, depending on the quantity purchased or consumed. The pictures are included as an external attachment to this basic information document called “*GHSW1\_foodphotos*”. A table describing the photo files is included in appendix 5.

The photographs were not an exhaustive set so a solution had to be provided in cases where photographs were not available. In cases where the quantity could not be obtained in standard units and the pictures were not available for the particular item, a third solution was provided. This solution was the crop non-standard unit conversion sheet that had local units and their equivalent weight in standard units which was used in the Agriculture Questionnaire. This sheet is shown in Figure 7.2. The interviewers were instructed to use the crop conversion sheet in cases where the household consumed from own production in quantities such as basins, baskets etc. The crop conversion sheet could also be used for purchases if in large quantities such as sacks or bundles etc.

However, it was subsequently found that this conversion sheet was not suitable for purchases and consumption of food by households. For example, a small basket of grains sold in the market was unlikely to weigh 15 kg. The weights are also incorrect for medium, big, and extra-large baskets as it relates to the purchase and consumption by the household. There were also problems with the basin in terms of size and weight. It should also be noted that different products would have different weights even if the volumes are equal and this conversion sheet does not cater for that. The conversion sheet was useful, however, when items such as rice or beans were purchased in sacks or bags.

Figure 7.2: Crop non-standard unit conversion sheet for Agriculture Questionnaire  
Units of measure for production (Col. 8 and 9)

Unit Weight code	Type of Measure	Weight	Unit Weight code	Type of Measure	Weight	Unit Weight code	Type of Measure	Weight
	<b>Sack/Bag</b>			<b>Bunch of Plantain/FFB</b>			<b>Wheel Barrow</b>	
11	Small	20 kg	41	Small	5 kg	71	Small	60 kg
12	Medium	50 kg	42	Medium	8 kg	72	Medium	85 kg
13	Big/Large	100 kg	43	Big	15 kg	73	Big/Large	110 kg
14	Extra Large	120 kg				74	Extra Large	150 kg
	<b>Basket</b>			<b>Tuber of Yam</b>			<b>Pick-up Van</b>	
21	Small	15 kg	51	Small	3 kg	81	Small	1,500 kg
22	Medium	30 kg	52	Medium	5 kg	82	Medium	2,000 kg
23	Big	50 kg	53	Big/Large	8 kg	83	Big	2,500 kg
24	Extra Large	75 kg		<b>Bundle of Millet, G/Corn, Sugarcane, Vegetable, etc.</b>			<b>Jerry can, Keg, Rubber of Palm oil</b>	
	<b>Basin</b>		61	Small	15 kg	91	Small	10 Lt.
31	Small	10 kg	62	Medium	25 kg	92	Medium	20 Lt.
32	Medium	25 kg	63	Big	40 kg	93	Big	25 Lt.
33	Big/Large	40 kg				94	Large	50 Lt.
34	Extra Large	75 kg				95	Drum	200 Lt.

## 7.5 Tracking Data

A number of households moved to new locations after responding in the post-planting visit of the survey, but before the post-harvest visit was conducted a few months later. As a result, the households that moved were not administered the post-harvest set of questionnaires. In order to reduce attrition of households, a tracking plan was formulated to track the relocated households to their new location and administer the post-harvest questionnaires. In addition, the plan was to reintegrate these tracked households into the set of panel households for future visits of the survey. The tracking exercise was also used to test the use of CAPI in administering panel LSMS-ISA type surveys in Nigeria. So, apart from designing the tracking plan, the CAPI application with a modified post-harvest questionnaire was also developed, using BLAISE CAPI software.

The tracking exercise was conducted between October and November 2011. A total of 111 complete households were identified as having moved permanently to other locations in Nigeria. Of these, 61 were tracked to their new location and the modified post-harvest questionnaire was administered. After the modified post-harvest survey was administered, the data for the tracked households were appended to the data of households that responded in the main post-harvest visit. A variable *trackingobs* has been added to the data files in the post harvest folder to identify the tracked households. The values of the variable are '1' for tracked households and '0' for all others. This variable can be used to identify the households and the household identification variable (hhid) can then be used in all other sections to identify the households in those sections that have been tracked.

The sections of the post-harvest questionnaire that were collected for tracked households and appended to the post-harvest datasets are:

1. Cover
2. Household Roster
3. Education
4. Labor
5. Health
6. Household Assets
7. Housing
8. Nonfarm enterprises

## 8.0 Overall Problems and Challenges Faced During the First Wave of the Survey and Recommendations for Wave 2 of the Survey

Designing and implementing a complex survey such as the GHS-Panel presents various challenges. In this section we outline some key issues that arose, lessons learned and make recommendations for the next Wave of the survey.

### 8.1 Sampling

During the post-planting round, each interviewer was provided a list of addresses of the households in the sample that they were responsible to interview. There were some problems of EA replacement, in cases of inaccessibility to the EA that were not done according to plan. One cause of this may have been the fact that the EAs listing had been done with a substantial gap between it and the actual field work.

- ◆ A re-listing exercise was done at the end of the post-harvest round and weights from this relisting have been included in the post-planting and post-harvest data.

There were also issues on the agreement of geographical codes between. The statistical system of the country is decentralized to some extent, with the central office of the NBS working with six zonal offices and 36 state (plus FCT) offices. It was discovered during the field work that there are different codes in use in the states and headquarters for LGAs and EAs.

It is recommended that for Wave 2, work be done to harmonize all geographic codes before another round of field work is done. This will benefit all surveys, not just the GHS.

### 8.2 Field work

#### 8.21 Pre-filling of Questionnaires

Prior to the fielding of the post-harvest survey, interviewers were required to pre-fill questionnaires with select data from the post-planting questionnaires. The pre-filled information was used to ask follow-up questions to those asked in the post-planting survey.

The pre-filling proved problematic in many instances. The main problem was that there were many cases where data was prefilled into the wrong column of the post-harvest questionnaire and in some cases in the wrong row. This resulted in data being assigned into the wrong variable or in a mismatch of cases between the post-planting and post-harvest data

#### 8.22 Availability of Electricity

Electricity was required by the data entry operator to operate the laptop computer and printer when in the field. This problem was anticipated so inverters were purchased and made available to each data entry operator. The intention was that these inverters would be connected to the cigarette lighter socket or battery of the vehicle that was providing transportation for the team. By doing so

electrical power would be generated to run the laptop and printer. Unfortunately, in most cases, no vehicle was available since the team was simply transported and left at the location. In addition some of the locations could only be accessed by motorcycles, which did not have the facility to operate the inverters.

In order to alleviate this problem, special provision had to be made for the purchase or renting of portable generators. This proved to be a satisfactory solution to the problem.

### 8.23 Flooding

During the post-planting visit many roads leading to the rural EAs were in very poor condition and were flooded during the rainy season. In rural areas, movement was difficult as only motorcycles were able to access some of the rural EAs. Using these motorcycles turned out to be more expensive than the taxis or vans that normally travelled those routes. In addition, in the flooded areas it was not possible to collect any information about the farms since these farms were under water.

Special provision should be made for transportation for the post-planting round of the next wave of the survey and it should be expected that there will be challenges in accessing some farm land in flood prone areas.

### 8.24 GPS Units

In some states, a number of the allocated GPS units malfunctioned and consequently reduced the number of farms that it was anticipated would have been covered in a given period of time. These GPS units were eventually replaced but unfortunately not before some delays were incurred.

In addition, only one GPS unit was allocated to each field team. Given the number of plots of land and the needs of 2-4 interviewers to measure all of this, it was determined that one GPS unit per team is not sufficient to collect the amount of data required.

### 8.25 Data entry

Overall the data entry in the field served as a useful quality control mechanism and improved the quality of the data. There were, however, a number of problems that were encountered during the post-planting round that were addressed before the fielding of the post-harvest visit. This was done by providing two extra days of training and practice for the data entry persons and supervisors only. The problems addressed included:

- ◆ Some of the data entry operators found the new system challenging. Specifically managing the laptop computers and managing issues with the data entry program.
- ◆ There were challenges in sending data via internet to NBS headquarters, in part due to poor connections but also due to inexperience of data entry operators with the use of the internet.

- ◆ The concurrent data entry system requires that the data from each questionnaire be entered electronically and a list of errors, inconsistencies and missing data produced. This list is then used by the supervisor to determine if a household needs to be re-visited and the interviewer is responsible for correcting any errors on the list. There were some problems in the effective managing of data problems while the teams were in the field such as printing and the correct reading of error messages.

## 8.26 Delay in States Responding to Data Queries

In both the post-planting and post-harvest rounds, the questionnaires were entered in the state where the survey was conducted and the questionnaires remained in that state. When data problems were identified the state was requested to check the questionnaire and/or revisit the household and correct errors or obtain additional data where the need arose.

The response from most states took an excessively long time and in some cases when the information was returned the query was not properly addressed and had to be resent. This introduced excessive delays in the data cleaning process.

## Appendix 1: How to Obtain Copies of the Data

The data are available through the NBS web site:

<http://www.nigerianstat.gov.ng/>

or through the LSMS-ISA website:

<http://www.worldbank.org/lsms-isa>

Users do not need to obtain the permission of the NBS to receive a copy of the data, but will be asked to fill in a data access agreement. In this agreement, users agree to: (a) cite the National Bureau of Statistics as the collector of the data in all reports, publications and presentations; (b) provide copies of all reports publications and presentation to the National Bureau of Statistics (see address below) and the Poverty and Inequality Division of the World Bank (see address below); and (c) not pass the data to any third parties for any reasons.

Leo Sanni  
Statistical Information Officer  
Plot 762, Independence Avenue,  
Central Business District,  
FCT, Abuja  
Nigeria  
[www.nigerianstat.gov.ng](http://www.nigerianstat.gov.ng)  
Phone: +2348033865388  
Email: [leosanni@nigerianstat.gov.ng](mailto:leosanni@nigerianstat.gov.ng)

LSMS Database Manager  
Poverty and Inequality Division  
The World Bank  
1818 H Street, NW  
MSN MC3-306  
Washington, DC 20433  
[www.worldbank.org/lsms-isa](http://www.worldbank.org/lsms-isa)  
Email: [lsms@worldbank.org](mailto:lsms@worldbank.org)

## Appendix 2: Updates to the Data

In June 2014, updates were made to the 2010-2011 GHS-Panel data. Starting at that time, all data downloaded from the web site included the following updates. If the data that you are using does not include these updates, you should download a newer version of the data by going to the LSMS web site (see Appendix 1).

### Post Harvest- Household:

#### *Section 1: HOUSEHOLD ROSTER*

- Description: Missing information on where members had moved within Nigeria corrected. There were several cases where the LGA name (s1q36a) was given but a corresponding LGA code (s1q36b) was missing or did not match the name. There were similar cases for State name (s1q36c) and code (s1q36d). LGA and State codes were added or corrected according to the name specified. Consistency between LGA and State was also checked, but was not always able to be corrected.
- Relevant file:
  - sect1\_harvestw1
- Variables affected:
  - s1q36a
  - s1q36b
  - s1q36c
  - s1q36d

#### *Section 7: HOUSEHOLD ASSET SALE AND ACQUISITION*

- Description: In the previous version, duplicate item observations were identified using an additional identifier “item\_seq”. However, further investigation revealed that many of these observations were in fact duplicates. Therefore, relevant duplicate observations were dropped. Also, several additional “other” item codes were added so that each observation in the data set can be identified using hhid and item\_cd.
- Relevant file:
  - sect7\_harvestw1
- Variables affected:
  - item\_cd
  - item\_seq (removed)

#### *Section 10b: HH FOOD EXPENDITURES*

- Description: This section was completely recleaned starting from the raw data file due to several irregularities within the original public version of the data.
- Relevant file:
  - sect10b\_harvestw1
- Variables affected (ALL):
  - s10bq1
  - s10bq2a
  - s10bq2b
  - s10bq3a

- s10bq3b
- s10bq4
- s10bq5a
- s10bq5b
- s10bq6a
- s10bq6b

**Miscellaneous:**

*Food photo aids:*

- Description: There were three photos missing from the food item photo aids uploaded to the website.
- Photos added:
  - Guinea corn (i10guineacorn1\_5k)
  - Millet (i11millet1\_7kg)
  - Maize (i12maize1\_48kg)



## Appendix 3: Agriculture Land Conversion Factors

The table below shows the conversion factors used to convert self-reported land areas (for agricultural land area of crops planted and harvested) into hectares.

### General Conversion Factors to Hectares

Zone	Unit	Conversion Factor
All	Plots	0.0667
All	Acres	0.4
All	Hectares	1
All	Sq Meters	0.0001

### Zone Specific Conversion Factors to Hectares

Zone	Conversion Factor		
	Heaps	Ridges	Stands
1	0.00012	0.0027	0.00006
2	0.00016	0.004	0.00016
3	0.00011	0.00494	0.00004
4	0.00019	0.0023	0.00004
5	0.00021	0.0023	0.00013
6	0.00012	0.00001	0.00041

Note: All conversion is to Hectares

## Appendix 4: Crop Codes

CROP	CODE	CROP	CODE	CROP	CODE
BEANS/COWPEA	1010	GINGER	2100	COCOA	3040
CASSAVA OLD	1020	GINGER PEELED	2101	COCOA POD	3041
COCOYAM	1040	GINGER SPLIT	2102	COCOA BEANS	3042
COTTON	1050	OTHER SPICES/VANILA	2103	COCONUT	3050
SEED COTTON	1051	GUM ARABIC	2110	COFFE	3060
COTTON LINT	1052	OKRO	2120	COFFE ARABICA	3061
COTTON SEED	1053	ONION	2130	COFFEE ROBUSTER	3062
GROUND NUT/PEANUTS	1060	PEPPER	2140	DATE PALM	3070
UNSHELLED GROUND NUTS	1061	SWEET PEPPER	2141	GRAPE FRUIT	3080
SHELLED GROUND NUTS	1062	SMALL PEPPER	2142	GUAVA	3090
GUINEA COUN/SORGHUM	1070	ATARE	2143	JUTE	3100
MAIZE	1080	PIGEON PEA	2150	KOLANUT	3110
UNSHELLED MAIZE(COB)	1081	PINEAPPLE	2160	KOLANUT UNSHELLED	3111
SHELLED MAIZE(GRAIN)	1082	PLANTAIN	2170	KOLANUT SHELLED	3112
POP CORN MAIZE	1083	POTATO	2180	BITTER KOLA	3113
MELON	1090	SWEET POTATO	2181	LEMON	3120
UNSHELLED MELON	1091	PUMPKIN	2190	LIME	3130
SHELLED MELON	1092	PUMPKIN LEAVE	2191	LOCUST BEAN	3140
WATER MELON	1093	PUMPKIN FRUIT	2192	MANDARIN/TANGERINE	3150
MILLET/MAIWA	1100	PUMPKIN SEED	2193	MANGO	3160
RICE	1110	GREEN VEGETABLE	2194	ORANGE	3170
UNSHELLED RICE(PADDY)	1111	DRY LEAVES(KUKA)	2195	OIL PALM TREE	3180
SHELLED RICE(MILLED)	1112	RIZGA	2200	FRESH FRUIT BUNCH	3181
YAM	1120	SHEA NUTS	2210	FRESH NUT	3182
WHITE YAM	1121	SOYA BEANS	2220	PALM OIL	3183
YELLOW YAM	1122	SUGAR CANE	2230	PALM KERNEL	3184
WATER YAM	1123	TEA	2240	AGBONO(ORO SEED)	3190
THREE LEAVE YAM	1124	TOBACCO	2250	OIL BEAN	3200
ACHA	2010	TOMATO	2260	PAWPAW	3210
BAMBARA NUT	2020	WALNUT	2270	PEAR	3220
BANANA	2030	WHEAT	2280	AVOCADO PEAR	3221
BEENI-SEED/SESAME	2040	ZOBO	2290	RUBBER	3230
CARROT	2050	ZOBO SEED	2291	RUBBER LUMP	3231
CUCUMBER	2060	APPLE	3010	RUBBER SHEET	3232
CABBAGE	2070	CASHEW	3020	CHERRY(AGBALUMO)	3240
LETUS	2071	CASHEW FRUIT	3021	ERU	3250
GARDEN EGG	2080	CASHEW NUT	3022	IYERE	3260
GARLIC	2090	CHILLI	3030		

## Appendix 5: Confidential Information, Geospatial Variables

The GHS-Panel collects confidential information on respondents. The confidential variables pertain to (i) names of the respondents to the household and community questionnaires, (ii) village and constituency names, (iii) descriptions of household dwelling and agricultural plot locations, (iv) phone numbers of household members and their reference contacts, (v) GPS-based household and agricultural plot locations, (vi) names of the children of the head/spouse living elsewhere, (vii) names of the deceased household members, (viii) names of individuals listed in the network roster, and (ix) names of field staff. To maintain the confidentiality of our respondents, certain parts of the GHS-Panel database have not been made publicly available.

To increase the use of the GHS-Panel data, a set of geospatial variables has been provided by using the georeferenced plot and household locations in conjunction with various geospatial databases that were available to the survey team. The table below provides the name, type, source, reference period, resolution, description, and source of each variable. There are two data sets with these georeferenced data.

All geospatial variables have been produced using the unmodified GPS data. These include extensive measures of distance, climatology, soil and terrain and other environmental factors. 37 Time-series on rainfall and vegetation have also been used to describe the survey agricultural season relative to normal conditions. These variables are intended to provide some understanding of how geophysical characteristics vary at the landscape level.

### *NGA\_PlotGeovariables\_Y1*

The household plot-level file, *NGA\_PlotGeovariables\_Y1*, contains four geospatial variables measuring plot distance to household, slope of plot, elevation of plot and Plot Potential Wetness Index. The observations are uniquely identified by the combination of **hhid** **plotid**. The observations included in this file are plots that are owned and/or cultivated by the household and that have been visited for GPS-based land area measurement.

### *NGA\_HouseholdGeovariables\_Y1*

The household-level file, *NGA\_HouseholdGeovariables\_Y1*, contains a range of variables measuring (on the basis of the household dwelling) distance to main points, climatology, landscape typology, soil and terrain, and crop season parameters. The observations are uniquely identified by **hhid**. This file also contains the GPS coordinate offsets provided to satisfy the user interest in geo-referenced location while preserving the confidentiality of sample household and communities. It provides the average of household GPS coordinates within each EA and applied a random offset within a specified range to the average EA value (following the Measure DHS approach).

More specifically, the coordinate modification strategy relies on random offset of cluster center-point coordinates (or average of household GPS locations by EA in GHS-Panel) within a specified range determined by an urban/rural classification. For urban areas a range of 0-2 km is used. In rural areas, where communities are more dispersed and risk of disclosure may be higher, a range

of 0-5 km offset is used. An additional 0-10 km offset for 1% of rural clusters effectively increases the known range for all rural points to 10 km while introducing only a small amount of noise. Offset points are constrained at the state level, so that they still fall within the correct state for spatial joins, or point-in-polygon overlays. The result is a set of coordinates, representative at the EA level, that fall within known limits of accuracy. Users should take into account the offset range when considering different types of spatial analysis or queries with the data. Analysis of the spatial relationships between locations in close proximity would not be reliable. However, spatial queries using medium or low resolution datasets should be minimally affected by the offsets.

Theme	Source	Dataset Title	Variable Name	Variable Type	Reference Period	Resolution	Description	Web
Distance	LSMS-ISA	Plot Distance to Household	dist_household	Continuous	N/A	N/A	Plot distance to household	
	AICD & NRNS	Household Distance to Main Road	dist_road	Continuous	1998	N/A	Household distance to nearest trunk road (Nigerian Road Network Survey 'Federal' roads, with some 'State' links)	
	AfriPop	Household Distance to Towns	dist_popcenter	Continuous	2006	N/A	Household distance to nearest town of >20,000 based on 2006 Census and AfriPop gridding algorithm	<a href="http://www.afripop.org/">http://www.afripop.org/</a>
	USAID FEWSNET	Household Distance to Key Market Centers	dist_market	Continuous	N/A	N/A	Household distance to nearest major market (FEWSNET key market centers)	
	GoogleEarth and other map sources	Household Distance to Border Posts	dist_borderpost	Continuous	N/A	N/A	Household distance to nearest border post on main road	
	Wikipedia and other map sources	Household Distance to State Capital	dist_admctr	Continuous	N/A	N/A	Household distance to the capital of the State of residence	
	National Bureau of Statistics	Household Distance to LGA Headquarters	dist_admctr2	Continuous	N/A	N/A	Household distance to the headquarters of the LGA of residence	
Climatology	UC Berkeley	WorldClim Bioclimatic Variables	af_bio_1	Continuous	1960-1990	0.008333 dd	Average annual temperature calculated from monthly climatology, multiplied by 10 (°C)	<a href="http://www.worldclim.org/bioclimate">http://www.worldclim.org/bioclimate</a>
	UC Berkeley	WorldClim Bioclimatic Variables	af_bio_8	Continuous	1960-1990	0.008333 dd	Average temperature of the wettest quarter, from monthly climatology, multiplied by 10. (°C)	<a href="http://www.worldclim.org/bioclimate">http://www.worldclim.org/bioclimate</a>
	UC Berkeley	WorldClim Bioclimatic Variables	af_bio_12	Continuous	1960-1990	0.008333 dd	Total annual precipitation, from monthly climatology (mm)	<a href="http://www.worldclim.org/bioclimate">http://www.worldclim.org/bioclimate</a>
	UC Berkeley	WorldClim Bioclimatic Variables	af_bio_13	Continuous	1960-1990	0.008333 dd	Precipitation of wettest month, from monthly climatology (mm)	<a href="http://www.worldclim.org/bioclimate">http://www.worldclim.org/bioclimate</a>

	UC Berkeley	WorldClim Bioclimatic Variables	af_bio_16	Continuous	1960-1990	0.008333 dd	Precipitation of wettest quarter, from monthly climatology (mm)	<a href="http://www.worldclim.org/bioclimate">http://www.worldclim.org/bioclimate</a>
Landscape Typology	ESA and UC Louvain	GlobCover v 2.3	fsrad3_lcmaj	Categorical	2009	0.002778 dd	Majority landcover class within approximately 1km buffer	<a href="http://ionial.esrin.esa.int/">http://ionial.esrin.esa.int/</a>
	ESA and UC Louvain	GlobCover v 2.3	fsrad3_agpct	Continuous	2009	0.002778 dd	Percent under agriculture within approx 1 km buffer	<a href="http://ionial.esrin.esa.int/">http://ionial.esrin.esa.int/</a>
	IFPRI	IFPRI standardized AEZ based on elevation, climatology	ssa_aez09	Categorical		0.008333 dd	Agro-ecological zones created using WorldClim climate data and 0.0833dd resolution LGP data from IIASA.	<a href="http://harvestchoice.org/production/biophysical/agroecology">http://harvestchoice.org/production/biophysical/agroecology</a>
Soil & Terrain	NASA	SRTM 90m	srtm_nga	Continuous		0.000833 dd	Elevation (m)	<a href="ftp://xftp.jrc.it/pub/srtmV4/arcasci/">ftp://xftp.jrc.it/pub/srtmV4/arcasci/</a>
	USGS	Slope (percent)	slopepct_nga	Continuous		0.008333 dd	Derived from 90m SRTM, aggregated to 1km block	<a href="http://pubs.usgs.gov/of/2007/1188/">http://pubs.usgs.gov/of/2007/1188/</a> , data provided USGS upon request
	AfSIS	Topographic Wetness Index	twi_nga	Continuous		0.000833 dd	Downloaded from AfSIS website. Derived from modified 90m SRTM. Local upslope contributing area and slope are combined to determine the potential wetness index: $WI = \ln (A_s / \tan(b))$ where $A_s$ is flow accumulation or effective drainage area and b is slope gradient.	<a href="http://www.ciesin.columbia.edu/afsis/bafsis_fullmap.htm#">http://www.ciesin.columbia.edu/afsis/bafsis_fullmap.htm#</a>
	LSMS-ISA	Terrain Roughness	srtm_nga_5_15	Categorical		0.000833 dd	Derived from 90m SRTM using 15 Meybeck relief classes and 5x5 pixel neighborhood	
	FAO	Harmonized World Soil Database	SQ1	Categorical		0.083333 dd	Nutrient availability	<a href="http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/">http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/</a>
	FAO	Harmonized World Soil Database	SQ2	Categorical		0.083333 dd	Nutrient retention capacity	<a href="http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/">http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/</a>
	FAO	Harmonized World Soil Database	SQ3	Categorical		0.083333 dd	Rooting conditions	<a href="http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/">http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/</a>

Crop Season Parameters	FAO	Harmonized World Soil Database	SQ4	Categorical		0.083333 dd	Oxygen availability to roots	<a href="http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/">http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/</a>
	FAO	Harmonized World Soil Database	SQ5	Categorical		0.083333 dd	Excess salts	<a href="http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/">http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/</a>
	FAO	Harmonized World Soil Database	SQ6	Categorical		0.083333 dd	Toxicity	<a href="http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/">http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/</a>
	FAO	Harmonized World Soil Database	SQ7	Categorical		0.083333 dd	Workability (constraining field management)	<a href="http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/">http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/</a>
	NOAA CPC	Rainfall Estimates (RFE)	anntot_avg	Continuous	2001-2010	0.1 dd	Average 12-month total rainfall (mm) for Jan-Dec	<a href="ftp://ftp.cpc.ncep.noaa.gov/feeds/newalgo_est_dekad/">ftp://ftp.cpc.ncep.noaa.gov/feeds/newalgo_est_dekad/</a>
	NOAA CPC	Rainfall Estimates (RFE)	wetQ_avg	Continuous	2001-2010	0.1 dd	Average total rainfall in wettest quarter (mm) within 12-month periods from Jan-Dec	<a href="ftp://ftp.cpc.ncep.noaa.gov/feeds/newalgo_est_dekad/">ftp://ftp.cpc.ncep.noaa.gov/feeds/newalgo_est_dekad/</a>
	NOAA CPC	Rainfall Estimates (RFE)	wetQ_avgstart	Continuous	2001-2010	0.1 dd	Average start of wettest quarter in dekads 1-36, where first dekad of Jan =1	<a href="ftp://ftp.cpc.ncep.noaa.gov/feeds/newalgo_est_dekad/">ftp://ftp.cpc.ncep.noaa.gov/feeds/newalgo_est_dekad/</a>
	NOAA CPC	Rainfall Estimates (RFE)	h2010_tot	Continuous	2010	0.1 dd	12-month total rainfall (mm) in Jan-Dec, starting January 2010	<a href="ftp://ftp.cpc.ncep.noaa.gov/feeds/newalgo_est_dekad/">ftp://ftp.cpc.ncep.noaa.gov/feeds/newalgo_est_dekad/</a>
	NOAA CPC	Rainfall Estimates (RFE)	h2010_wetQ	Continuous	2010	0.1 dd	Total rainfall in wettest quarter (mm) within 12-month periods starting January 2010	<a href="ftp://ftp.cpc.ncep.noaa.gov/feeds/newalgo_est_dekad/">ftp://ftp.cpc.ncep.noaa.gov/feeds/newalgo_est_dekad/</a>
	NOAA CPC	Rainfall Estimates (RFE)	h2010_wetQstart	Continuous	2010	0.1 dd	Start of wettest quarter in dekads 1-36, where first dekad of January 2010 =1	<a href="ftp://ftp.cpc.ncep.noaa.gov/feeds/newalgo_est_dekad/">ftp://ftp.cpc.ncep.noaa.gov/feeds/newalgo_est_dekad/</a>
	BU	MOD12Q2 Land Cover Dynamics (PHENOLOGY)	eviarea_avg	Continuous	2001-2010	0.004176 dd	Average total change in greenness (integral of daily EVI values) within growing season, averaged by state	<a href="ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005">ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005</a>
		MOD12Q2 Land Cover Dynamics (PHENOLOGY)	evimax_avg	Continuous	2001-2010	0.004176 dd	Average EVI value at peak of greenness, averaged by state	<a href="ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005">ftp://e4ftl01.cr.usgs.gov/MOTA/MCD12Q2.005</a>

	BU	MOD12Q2 Land Cover Dynamics (PHENOLOGY)	grn_avg	Continuous	2001-2010	0.004176 dd	Average timing of onset of greenness increase in day of year 1-356, within early growing season (first peak of phenological cycle), averaged by state	<a href="ftp://e4ftl01.cr.usgs.gov/MO/TA/MCD12Q2.005">ftp://e4ftl01.cr.usgs.gov/MO/TA/MCD12Q2.005</a>
	BU	MOD12Q2 Land Cover Dynamics (PHENOLOGY)	sen_avg	Continuous	2001-2010	0.004176 dd	Average timing of onset of greenness decrease in day of year 1-356, within growing season, averaged by state	<a href="ftp://e4ftl01.cr.usgs.gov/MO/TA/MCD12Q2.005">ftp://e4ftl01.cr.usgs.gov/MO/TA/MCD12Q2.005</a>
	BU	MOD12Q2 Land Cover Dynamics (PHENOLOGY)	h2010_eviarea	Continuous	2010	0.004176 dd	Total change in greenness (integral of daily EVI values) within growing season of 2010, averaged by state	<a href="ftp://e4ftl01.cr.usgs.gov/MO/TA/MCD12Q2.005">ftp://e4ftl01.cr.usgs.gov/MO/TA/MCD12Q2.005</a>
		MOD12Q2 Land Cover Dynamics (PHENOLOGY)	h2010_evimax	Continuous	2010	0.004176 dd	EVI value at peak of greenness within growing season of 2010, averaged by state	<a href="ftp://e4ftl01.cr.usgs.gov/MO/TA/MCD12Q2.005">ftp://e4ftl01.cr.usgs.gov/MO/TA/MCD12Q2.005</a>
	BU	MOD12Q2 Land Cover Dynamics (PHENOLOGY)	h2010_grn	Continuous	2010	0.004176 dd	Onset of greenness increase in day of year 1-356, within growing season of 2010, averaged by state	<a href="ftp://e4ftl01.cr.usgs.gov/MO/TA/MCD12Q2.005">ftp://e4ftl01.cr.usgs.gov/MO/TA/MCD12Q2.005</a>
	BU	MOD12Q2 Land Cover Dynamics (PHENOLOGY)	h2010_sen	Continuous	2010	0.004176 dd	Onset of greenness decrease in day of year 1-356, within growing season of 2010, averaged by state	<a href="ftp://e4ftl01.cr.usgs.gov/MO/TA/MCD12Q2.005">ftp://e4ftl01.cr.usgs.gov/MO/TA/MCD12Q2.005</a>



## Appendix 5: Description of Food Photo Files

File name	Item		Item if other	Specification in the JPG:		
	Code	Name		KGS/liters	Non standard unit	
i10guineacorn1_5kg	10	Guinea Corn		1.5	One Mudu	
i11millet1_7kg	11	Millet		1.7	One Mudu	
i12maize1_48kg	12	Maize		1.48	One Mudu	
	13;					
i13i14rice1kg	14	Rice - local ; Rice - Imported		1	One Mudu	
i17yamflour1_43kg	17	Yam Flour		1.43		
i31yamroots	31	Yam Roots		2.3 3.2 5.2		
i32gariwhite1_25kg	32	Gari White		1.25	One Mudu	
i34cocoyam1_4kg	34	Cocoyam		1.4		
i34cocoyam1_73kg	34	Cocoyam		1.73		
i34cocoyam_69kg	34	Cocoyam		0.69		
i36sweetpotatoes	36	Sweet Potatoes		1 2.2 3.6	Heap	
i42whitebeans1_35kg	42	White beans		1.35	One Mudu	
i43groundnuts1_5kg	43	Groundnut		1.5	One Mudu	
i44othMelonseeds_2kg	44	Other nuts/seeds/pulses	Melon Seeds	0.20	Small Derica	
i44othMelonseeds_9kg	44	Other nuts/seeds/pulses	Melon Seeds	0.9	Large Deeica	
i44othMelonseeds_35kg	44	Other nuts/seeds/pulses	Melon Seeds	0.35	Medium Size	
i44othRedbeans1_65	44	Other nuts/seeds/pulses	Red Beans	1.65	One Mudu	
i60bananas1_3kg	60	Bananas		1.30		
i60bananas1_55kg	60	Bananas		1.55		
i60bananas3_6kg	60	Bananas		3.60		
i61oranges	61	Oranges		1.2 2.60 4.3		
i64pineapples	64	Pineapples		1.90 2.2 2.50		
i66othPawpaw	66	Other fruits	Pawpaw	2.3 3.00 3.4		

File name	Item		Item if other	Specification in the JPG:			
	Code	Name		KGS/liters			Non standard unit
i70tomatoes	70	Tomatoes		1.00	2.4	4.50	
i73gardeneggs	73	Garden Eggs		0.4	0.65	0.9	
i78othCarrot_1kg	78	Other vegetables (fresh or canned)	Carrot	0.10			
i78othCarrot_05kg	78	Other vegetables (fresh or canned)	Carrot	0.05			
i78othCarrot_17kg	78	Other vegetables (fresh or canned)	Carrot	0.17			
i78othGreenpepper1_5kg	78	Other vegetables (fresh or canned)	Green pepper	1.5			
i78othGreenpepper1kg	78	Other vegetables (fresh or canned)	Green pepper	1.00			
i78othGreenpepper_53kg	78	Other vegetables (fresh or canned)	Green pepper	0.53			
i78othRedpepper	78	Other vegetables (fresh or canned)	Red pepper	0.80	1.5	4.00	
i83agrieggs2kg	83	Agricultural Eggs		2			Crate
i83agrieggs_1kg	83	Agricultural Eggs		0.10			6 pieces
i83agrieggs_89kg	83	Agricultural Eggs		0.89			12 pieces
i103driedfish.jpg	103	Dry Fish		0.30	1.1	1.90	Small, medium, large