
# Improving the Availability, Quality and Policy-Relevance of Agricultural Data: The Living Standards Measurement Study – Integrated Surveys on Agriculture

Gero Carletto
Development Research Group
The World Bank

and the LSMS Team

(Kathleen Beegle
Kristen Himelein
Talip Kilic
Siobhan Murray
Mimi Oseni
Kinnon Scott
Diane Steele)

## Abstract

The recent food crisis has sharply highlighted both the importance of sound agricultural policies as well as the weaknesses in agricultural information systems that hinder knowledge generation, innovation and change. Despite the importance of the agricultural sector and its critical role in meeting the MDGs and buttressing governments' poverty reduction and growth strategies, serious weaknesses in agricultural statistics persist throughout sub-Saharan Africa.

There are a number of reasons why the quality and quantity of agricultural data have seldom matched their importance in policy-making. First, poorer countries, for which agriculture is a critical source of livelihoods, often have the poorest data. Second, agricultural data are often collected in institutional isolation, with little coordination across sectors and little analytical value-added beyond the sector. Third, the lack of analytical capacity has created a vicious cycle of poor analysis undermining the demand for high-quality data. Finally, inadequate data and measurement issues have affected the ability of policy analysts and researchers to contribute to the design of innovative and more effective policy.

The Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) program is a new initiative of the Development Research Group of the World Bank, funded by the Bill and Melinda Gates Foundation, aimed at developing and implementing innovative household panel surveys in six Sub-Saharan African countries, with a strong focus on agricultural and rural development. A number of tenets are central to the LSMS-ISA project. First, the collection of agricultural data must be integrated into a broad, multi-sectoral framework that goes well beyond agriculture. Second, the collection of agricultural data must be buttressed by a well-matched institutional setting conducive of collaboration and integration of data sources. Third, national capacity needs to be strengthened to enhance the value of the data generated and bolster the link between data producers and data users. This, in turn, negatively affects the quality and availability of policy-relevant analysis. Poor dissemination of the available data and results has further aggravated the problem.

The paper highlights the main features of the LSMS-ISA project, how it is addressing some of the current weaknesses of agricultural and rural statistics, and summarizes the experience to date in introducing the program in a number of Sub-Saharan countries.

## I. Introduction

Agricultural development is an essential engine of growth, and an effective mechanism for combating poverty and food insecurity, since it often results in greater benefits accruing to the poorest segments of the population (Ligon and Sadoulet, 2007; Ravallion and Chen, 2007). The 2008 World Development Report "*Agriculture for Development*" (World Bank, 2007) argues that agriculture is critical if countries are to achieve the poverty targets set forth by the Millennium Development Goals within the agreed timeframe. In most Sub-Saharan African countries, the vast majority of people suffering from poverty and food insecurity are rural dwellers and other vulnerable groups, including women, who rely heavily on farm activities. Thus, efforts to fight poverty in Africa must focus on rural areas and agriculture, and must be gender-sensitive. Indeed, this is one of the central tenets common to most Poverty Reduction Strategies in the region (Gubert and Robillard, 2002).

The recent food crisis has sharply highlighted both the importance of sound agricultural policies as well as the weaknesses in agricultural information systems that hinder knowledge generation, innovation and change. The distressing events driven by high global food prices, however, have the potential of translating a major crisis into an opportunity, as reflected by a renowned interest and increasing investments in the sector. The need to evaluate the effectiveness and pro-poor nature of these interventions is of paramount importance and better data are needed.

Agriculture does not, however, exist in a vacuum. It forms part of complex household income generating strategies that involve multiple individuals and activities in different sectors (World Bank, 2007; Davis et al, 2007; Foster and Rosenzweig, 2008). Diversification into non-farm activities among smallholders has evolved to be the norm rather than the exception (Reardon, 1997; Bryceson, 2002; Davis et al, 2007). Higher incomes and lower risk exposure can be achieved by enhancing the linkages among the different income sources of the rural poor. The existing institutional setting and functioning of non-farm sectors of most countries in Sub-Saharan Africa, however, is not

conducive to better integration across sectors and stakeholders. Promoting a more holistic approach to rural development will require better coordination among the different players and improved multi-sectoral data that document those linkages must be an important input to this process.

Despite the importance of the agricultural sector and its critical role in meeting the MDGs, serious weaknesses in agricultural statistics persist throughout sub-Saharan Africa. Of the 44 countries in the region rated by the Food and Agricultural Organization of the United Nations (FAO), only two are considered to have high standards in data collection, while standards in 21 countries remain low (FAO, 2008). The scope of coverage and completeness also varies widely (see for example, the four-country case studies by Kelly and Donovan, 2008). Knowledge about agriculture and its impact on welfare and equity is limited by the lack of available, high quality, and consistent data on rural households. In their review of agricultural development, rural non-farm activities, and rural poverty, Foster and Rosenzweig (2008) note that "very few studies permit the direct comparison over time using comparable measures…". Other studies have likewise noted that inconsistencies and other quality issues limit analysis (Ngendaumana, 2001; Tiffen, 2003;). Data are also lacking on intra-household decision-making related to agricultural activities: more gender-disaggregated data on intra-household allocation of resources and participation in agriculture and other income-generating occupations are needed to understand the separate roles played by women and men in productive and reproductive activities. Past investments and technical assistance efforts in the area of agricultural statistics have failed to produce sustainable systems, which continue to suffer from poor quality, lack of relevance and little use in countries' policy dialogues (Binswanger, 2008). Furthermore, the rapidly changing nature of agriculture in less developed countries and the emergence of new issues make the available data and methods obsolete. For example, information on climate change and adaptation and mitigation practices, as well as its impact on poverty, is seldom collected and little is known about methods and best practices on how to collect it.

There are a number of reasons why the quality and quantity of agricultural data have seldom matched their importance in policy-making. First, poorer countries, for which agriculture is a critical source of livelihoods, often have the poorest data in general (African Development Bank et al, 2004). In spite of the clear need for empirical analysis, these countries lack the financial resources to generate survey or administrative data related to agriculture and off-farm activities. Even with sufficient financial resources, they often lack human resources to collect such data in a cost-effective and sustainable manner. Uneven and one-off external support over the years has been inadequate to overcome this situation.

Second, agricultural data are often collected in institutional isolation, with little coordination across sectors and little analytical value-added beyond the sector. In many countries, the data collected by the Ministries of Agriculture (MoA) are not linked or utilized in conjunction with data available from the National Statistical Office (NSO), or other line Ministries (such as Labor, Education and Health).[1] In part, this reflects the long-standing failure to recognize that rural economies are diverse, and that this diversification is found within households. This issue was recognized by the Task Team on Food, Agriculture and Rural Statistics (Paris21, 2002). Key recommendations included re-thinking agricultural surveys, including broadening their scope to include both agricultural and non-agricultural activities, and advocating for an improvement in the coordination of the various agencies responsible for the production of agricultural statistics. This problem of integration figures prominently in the new Global Strategy to Improve Agricultural and Rural Statistics jointly drafted by a group of development partners and recently endorsed by the UN Statistical Commission.

A third important cause of poor data in developing countries is the lack of analytical capacity that has created a vicious cycle of poor analysis undermining the demand for high-quality data. Lags in producing reports with data also present problems. For

[1] In Malawi, for example, there are large differences in the estimates of the number of farm households between the Ministry of Agriculture (3.4 million farm households) and the National Statistics Office (2.47 rural households), which in turn affects the accuracy and effectiveness of planning for the subsidized input program (School of Oriental and African Studies, 2008).

example, the national and regional reports from the 2003 Agricultural Census Sample Survey (ACSS) in Tanzania were only produced in 2006-07 (Kaimu and Muñoz, 2007). Although these problems are common to developing countries around the globe, the problem appears to be more acute in Sub-Saharan African countries. The 2002 Paris21 Taskforce stressed the importance of strengthening the statistical and analytical capacities of the data producers and a number of initiatives, including the Accelerated Data Program (ADP), are aimed at providing support to countries in their data dissemination efforts.

Finally, measurement issues have affected the ability of policy analysts and researchers to contribute to innovative and more effective policy. Many aspects of agriculture are inherently difficult to measure, including the measurement of plot size and crop yields and, valuation of agricultural output used for home consumption. This is exacerbated by the seasonality of farming: while key agricultural decisions are made over a cropping cycle, most surveys are conducted at one point in time, often with lengthy recall periods. There is still substantial work to be done in both developing new survey techniques and validating existing methods.

In summary, understanding agriculture as an integral part of the complex rural economy, as well as the agricultural links that exist between rural and urban economies and populations, is a critical component of any poverty reduction strategy in Sub-Saharan Africa. However, existing data suffer from inconsistent investment, institutional and sectoral isolation, and weaknesses in methods. This, combined with a lack of in-country analytic capacity, has led to serious gaps in knowledge, and has hampered the ability to identify and promote effective innovation and sources of sectoral growth.

The Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) project is a new initiative of the Development Economics Research Group of the World Bank and funded by the Bill and Melinda Gates Foundation. The project centers on the development and implementation of household panel surveys in seven Sub-Saharan African countries, with a strong focus on agricultural and rural development, with project countries being selected in order to maximize synergies with other on-going

panel efforts in a few countries in the region, including panel surveys in Ghana, Zambia and Kenya. Presently, the LSMS-ISA project is working in six countries, including Tanzania, Uganda, Ethiopia, Niger, Nigeria and Malawi.

The project's over-arching objective is twofold: to partly fill in knowledge gaps about agriculture, and to contribute to improving the quality, relevance and sustainability of household-level agricultural data systems. This second goal will be achieved by developing and implementing an innovative model for collecting agricultural data in the region, as well as by conducting methodological validation experiments to improve commonly used measures and indicators.

A number of tenets are central to the LSMS-ISA project. First, the collection of agricultural data must be integrated into a broad, multi-sectoral framework that goes beyond agriculture and rural areas. This facilitates the production of data necessary to design effective agricultural policies country-wide and within the broader rural economy.[2] Second, the collection of agricultural data must be buttressed by a well-matched institutional setting conducive of collaboration and integration of data sources. Currently, the institutional setting in most developing countries is poorly suited to achieving this goal and the lack of a strong rural constituency makes such integration more difficult (Zezza et al., 2007). By promoting a multi-purpose approach to data collection, the project aims at fostering inter-sectoral collaboration and overcoming some of the existing institutional constraints. Third, national capacity needs to be strengthened to enhance the value of the data generated and bolster the link between data producers and data users. This consists of both (1) the ability to produce timely, policy-relevant, high-quality and public-use data sets and (2) the technical capacity to analyze such data. As noted above, the lack of analytical capacity in developing countries perpetuates the low demand for data. This, in turn, negatively affects the quality and availability of policy-relevant analysis. Poor dissemination of the available data and results has further aggravated the problem.

[2] Note that agriculture is also important in urban and peri-urban areas (Lee-Smith and Prain, 2006; UNDP, 1996, and FAO website materials on Urban Harvest Program) and some research shows that linkages between urban populations and their rural, agrarian relatives affect urban welfare (Frayne, 2004).

## II. Design Features of LSMS-ISA

In order to address some of the existing weaknesses, the LSMS-ISA is promoting a number of innovations in the way agricultural data are collected and disseminated.

***Panel***

This project entails the implementation of a set of panel household surveys in Sub-Saharan Africa. Panel, or longitudinal, surveys entail repeat visits to a baseline sample. Panel surveys offer the unique opportunity to follow the socio-economic trajectory of households over-time, allowing for more in-depth analysis than can be undertaken with repeat cross-sectional surveys. Panel surveys also offer a unique opportunity to rigorously evaluate the impact of specific interventions (for more discussion on the analytical advantages of panel data, see Durrant and Menken, 2002, and Rosenzweig, 2003). The Indian Village Level Studies of the International Crop Research Institute for the Semi-Arid Tropics (ICRISTAT) is probably the best known panel survey specifically focused on agricultural households (see Walker and Ryan, 1990). A more recent example of a large-scale household panel is the Indonesia Family Life Survey (Strauss, et al. 2004). While there are some panels focused on agricultural households or specific regions in Sub-Saharan Africa, such as the Tegemeo panel in Kenya (Takashi and Jayne, 2004) or the Kagera Health and Development Survey (Beegle et al. 2004), there are few large-scale, nationally representative panel household surveys which specifically focus on agriculture. A notable example is the Trabalho de Inquérito Agricola (TIA) survey conducted by the Ministry of Agriculture of Mozambique with technical assistance from MSU. In some cases, one-off efforts are undertaken to specifically study pressing topics. For instance, a panel was conducted in Malawi in 2007, following a sub-set of households interviewed as part of the 2004/05 national second integrated household survey (School of Oriental and African Studies, 2008). These data were used to assess the contribution of the fertilizer voucher scheme on maize production, with the contribution found to be lower than was commonly portrayed in the national policy debate.

The LSMS-ISA project calls for panel surveys whereas households would be re-interviewed every three years, and in some cases more frequently. It is believed that a three-year spacing provides an appropriate interval to monitor dynamics of most variables and get a good understanding of changes in behavioral and agronomic patterns.[3] This interval seems also more appropriate in view of the limited local capacity of most countries to analyze a continuous and voluminous flow of data. It is also recognized, however, that given the large temporal instability of agricultural production due to changing agro-climatic conditions, a three-year interval may not be appropriate to fully capture fluctuations in production. For this reason, in some countries, higher frequency data collection is being adopted (e.g. annual in Uganda and Tanzania, and biennial in Ethiopia. Furthermore, smaller complementary data collection exercises in the interim years is being contemplated in some other countries, e.g. in Malawi.

As noted above, panel data are more complex than traditional cross-sectional data, both analytically and in terms of actual data collection. With respect to data collection, the quality of a panel depends on the extent to which attrition can be minimized. As the population ages and moves, some attrition is inevitable, and can result in biases both in terms of basic statistics and in regard to analytical work examining changes in socioeconomic condition (Rosenzweig, 2003). In order to maintain the national representativeness of the data, survey samples will be periodically refreshed partially or fully, depending on the levels of attrition and mobility.

In addition to potential refresher samples, the project will track individuals and households that relocate. The extent to which tracking can be implemented for all individuals regardless of where they move will depend on the nature of mobility in each country and the available budget. In the case of Uganda, for example, the panel survey is being administered to a sub-sample of households that were interviewed as part of the Uganda National Household Survey conducted in 2005/06. The second survey round

[3] A similar spacing is being adopted by Yale University in their Ghana panel survey, and by MSU in Mozambique.

which started in August 2009 – with co-funding from the Government of Uganda, the LSMS-ISA project and the Dutch Government -- will follow the original sample of households, including the ones that have relocated since 2005/06, as well as a random sub-sample of individuals that have split off from original sample households. The decision was taken to follow all split off individuals from two randomly selected households in each Primary Sampling Unit (PSU).

***Multi-topic, Multi-purpose***

The LSMS-ISA surveys is modeled on the integrated household survey design of the Living Standards Measurement Study (LSMS) of the World Bank.[4] As noted above, for almost three decades, the LSMS has been at the forefront of implementing and producing multi-topic, nationally representative household surveys, which are well documented and publicly available.

A number of distinctive features characterize LSMS surveys. First, information is collected across a range of topics related to welfare. Most importantly, the survey instrument is designed to collect comprehensive data for the construction of a full consumption-based welfare measure.[5] This measure provides a critical underpinning to much of the analysis of living standards being produced in developing countries. It is the basis for monitoring the first Millennium Development Goal (MDG) on poverty and generating much of the WB Global Poverty Monitoring Database. There is little doubt that to a large extent the quantum leap in our ability to measure and monitor poverty around the globe over the past three decades may be attributed to the work of the LSMS team and the World Bank. The poverty estimates generated through LSMS-type surveys have also contributed to much of the policy and impact analysis carried out by the World

[4] For details of the LSMS and LSMS-ISA program and surveys, including methodological materials, past and present activities, documentation, questionnaires and data sets, see www.worldbank.org/lsms. For a review of the LSMS methodology and features, also see "Designing Household Survey Questionnaire for Developing Countries; Lessons from 15 Years of the Living Standards Measurement Study", edited by M. Grosh and P. Glewwe, World Bank, 2000. More updates and complete information on the LSMS-ISA project can also be found on the LSMS website.

[5] Consumption expenditure has long been the preferred measure of welfare rather than income for low-income settings with large agriculture and informal sectors like Sub-Saharan African countries, for a number of reasons (see Deaton and Grosh, 2000, and Deaton and Zaidi, 2002, for a review of the issues).

Bank, development institutions and governments around the world. Datasets such as the ones generated by an LSMS, allow distributional and incidence analysis, and are thus a unique tool to assess the poverty impact and targeting efficiency of governments' and donors' programs.

LSMS surveys often also collect individual-level data by gender on control of household resources, decision-making and participation in agriculture and off-farm activities, thus enabling researchers to conduct gender-specific analysis. Furthermore, the use of female enumerators and direct interviews with female respondents are standard applications in LSMS surveys. As explained below, the same type of information is being gathered through the proposed survey, and similar gender-sensitive methodologies will be utilized for a better understanding of women's role in agriculture and the rural economy.

Since the project aims at operating, as much as possible, within the existing system of surveys, the specific design across countries varies somewhat. Although several Sub-Saharan African countries do have existing multi-topic surveys, they have fairly small and deficient agricultural modules. On the other hand, most countries have agricultural sample surveys that contain extensive agricultural information but no corresponding information on, for instance, consumption, health, and non-farm activities of household members. In addition, these agricultural surveys are seldom linked to the national system of household surveys, are often based on non-representative samples, and are virtually never based on longitudinal data. This leads to quality problems, unexplainable differences in estimates from different data sources (see footnote 1), as well as the subsequent under-utilization of the data and its lack of relevance and credibility.

Although the final multi-topic questionnaires may be somewhat different across countries in terms of content and depth of individual modules, each questionnaire is likely to contain information on the following topics in addition to agriculture:

- household demographics
- education
- health and nutrition

- food consumption/expenditures
- food security
- non-food expenditures
- dwelling conditions
- assets
- employment
- non-farm self-employment and other sources of income
- internal and international migration
- participation in projects and programs
- economic shocks and coping strategies

Whenever feasible, information is being collected to allow gender-disaggregated analysis. For example, information on asset ownership, control of resources and participation in programs and productive activities is being differentiated along gender-lines. The use of female enumerators and direct interviews with female respondents is likely to improve the accuracy of the information collected, and will allow for the design of improved female empowerment policy.

Based on the specific policy questions, and the obvious trade-offs, the depth of each module, or the number of modules, will be adjusted accordingly. For instance, although one may typically limit the amount of information collected in the health module to basic morbidity, access to health services, health expenditures and child anthropometric measures, it will be possible in some surveys to include a more extensive module and collect data on full children immunization histories, pre- and post-natal mother behavior and infant feeding practices or even conduct micro-nutrient testing and HIV screening. In the Tanzania case, for example, anthropometric measures of all members of the households, including adults, are taken. In the Uganda case, a small module on reproductive health and contraceptive use is being administered privately to a sample of women of reproductive age. The possibility of rotating modules over time is envisioned, depending on the specific policy questions which are relevant in a given country. For a detailed exposition on the specific objectives and details for each of these modules, the

reader can refer to the LSMS publication *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study* (Grosh and Glewwe, 2000).

With respect to agriculture, a core module has been developed that is being used as a general template for each country. The template is being modified to fit country-specific aspects of agricultural systems. By avoiding a strictly standardized module, the data will be more appropriate to study and address country-relevant policy issues.[6]

The core set of broad indicators covered in the agricultural module have been identified by the LSMS team through a consultative process with several sectoral experts. These key areas are needed both for monitoring trends, but also for analytical work to understand, for example, the factors hindering agricultural efficiency and productivity, and the role of women in agriculture and other productive activities. Clearly, some flexibility is required to be able to address new issues as they arise over time. Information on the following variables will be collected in all countries:

- Basic crop production, storage and sales
    - Productivity of main crops, with emphasis on improved measures of quantification of production; plot size using GPS, and production shocks
- Land holdings
    - Tenure and transactions
    - Type of soil, erosion controls, and irrigation systems
- Input use and technology adoption
    - Family and hired labor
    - Use of technology and farming implements
    - Seed varieties
    - Fertilizer, pesticides/herbicides applications

[6] Some examples of country-specific aspects of agricultural include the land tenure system and inheritance laws, the marketing system (for example, tobacco auction houses in Malawi and coffee cooperatives in Tanzania), the structure of input markets (such as fertilizer distributed as part of a social protection system including free fertilizer distribution or fertilizer vouchers).

- Farming practices
    - Mechanization
    - Soil and environmental management
    - Water management
    - Adaptation to climate change, and mitigation strategies
- Livestock
    - Quantification of current stock, sales and input expenditures
    - Reliance on and access to veterinary practices
    - Quantification of livestock by-product production and sales
- Fishery
    - Quantification of production, sales and input expenditures
    - Ownership and use of fishing implements
- Access to and use of services, infrastructure and natural resources
    - Agricultural extension services
    - Infrastructure (including roads)
    - Credit (both for agriculture and other purposes)
    - Access to markets and information
    - Access to natural and common property resources

A community questionnaire is also being administered in conjunction with the household survey to collect community-level information on topics including access to public services and infrastructure, social networks, institutions, and retail prices. As part of this effort, and in collaboration with the International Fund for Agricultural Development (IFAD), a module on community empowerment and governance is being developed and field tested for inclusion in each of the surveys.

Beyond the design of the questionnaires, the project considers three additional aspects of standardization. First, the project is developing a standard data entry application to accompany the core agricultural module. Both field-based concurrent data entry as well as Computer Assisted Personal Interview (CAPI) applications are being considered. For example in Tanzania, Uganda and Malawi concurrent data entry protocols are in place.

Additionally the LSMS team is working towards the development of a CAPI platform to administer the Uganda National Panel Survey (UNPS) questionnaire in a paperless format during year two, by using Ultra-Mobile Personal Computers (UMPCs). The experience accumulated to date supports the suitability of this hardware for this type of field application, but further field testing of both hardware and software is required. Similar efforts will be pursued in other LSMS-ISA countries, but rolled out gradually based on each country's capacity and readiness. Furthermore, a review of existing applications is being commissioned by the team, in order to guide future choices in terms of both software and hardware.

Second, the project is working on a basic data analysis tool along the line of the ADePT program developed by the Development Economics Research Group (DECRG) at the World Bank (http://econ.worldbank.org/programs/poverty/adept). The ADePT program, based on the statistical software platform STATA, produces a standard set of tables, graphs and maps related to poverty indicators. The program in its current form allows comparisons of poverty statistics across countries. It contains features for sensitivity analysis and data checking to minimize errors. It can also be used to introduce new methods of applied economic analysis to a wider audience. It has also proven to be an excellent training tool for basic poverty analysis. Extensions of ADePT now include separate modules on Labor, Gender and Social Protection. Developing a similar application for the agricultural and livestock modules is likely to enhance involvement of local counterparts and speed up the production of basic agricultural statistics once the data are collected. The existing ADePT modules will also be disseminated and used for basic analyses of other sections of the questionnaire.

Finally, the project will incorporate the new data sets into the Comparative Living Standards Project (CLSP) platform and the Rural Income Generating Activity (RIGA) database. CLSP is a web database developed (currently in Beta version) and maintained by the LSMS team. Comparable data sets will be constructed from the household surveys for each country. The CLSP interface will allow users to access the data using on-line analysis tools. This interface permits those with limited experience with micro-data and

statistical software to use the information as well as provides much of the groundwork for others interested in carrying out comparative and cross-country analysis. Using this tool will ensure the widest possible access to the comparable data generated by the project. The RIGA database is the result of a multi-year collaboration between the FAO and the World Bank, and involves the construction and dissemination of comparable income aggregates.

***Country-Specificity with Core Set of Common Indicators***

Data sets need to respond to country-specific demand while also using international standards to allow for cross-country comparability. One of the features of a typical LSMS survey is balancing customization to country needs and policy questions with cross-country comparability. LSMS surveys are demand-driven and often originate from requests of Governments to meet well-defined data needs. The questionnaires are designed with country-specific policy questions in mind, and in close consultation with local counterparts and stakeholders and, as such, no two LSMS surveys are identical. In order to improve the relevance and sustainability of data collection, the importance of local partnership and ownership during the planning and design stages cannot be emphasized enough. This is the approach followed by the LSMS team: LSMS surveys are typically carried out with government counterparts, often the National Statistical Office (NSO), with substantive involvement from line ministries and local stakeholders.

The extent of customization also varies by module. For example, questions on assets, non-farm enterprises, demographics and consumption may require less customization compared with modules on agriculture or institutions. Cultural norms and practices may also affect how questions are phrased and interviews administered, for example with respect to interviewing women. Tailoring in terms of content and design, however, must also take account of international standards and cross-country comparability. The proposed system incorporates the criteria of country-customization and cross-country comparability in terms of contents as well as international nomenclatures and classifications.

***Integration into Existing or Planned Survey Activities***

Given the limited capacity and human resource base, ideally, any new data collection strategy in Sub-Saharan Africa ought to be integrated, as much as possible, into each country's existing system of surveys. This involves an appropriate institutional framework of coordination as well as physically integrating the proposed activities into on-going or planned surveys. This is a necessary, albeit not sufficient, condition to maximize the relevance, use and sustainability of the surveys and data. Household surveys can be very expensive, particularly in sub-Saharan African countries with high transportation costs and limited human resources, requiring a high-level of technical assistance. Therefore, avoiding duplication and taking advantage of existing surveys and plans is critical. The work proposed here is being aligned closely with countries' National Strategies for Statistical Development (NSDS), if these exist, and the project is working with countries presently implementing such strategies to ensure that agricultural statistics are integrated appropriately.

Whenever feasible, the surveys under the project are being "piggybacked" onto ongoing or scheduled government-sponsored data collection efforts. Such cost-saving designs can improve on the sustainability of the surveys. In addition to the cost-saving benefits, integration into a planned activity will encourage government ownership and endorsement, while also making better use of the limited human resources available in the country. For example, in Malawi, the Malawi Integrated Household Panel Survey (IHPS) has been nested within the Third Integrated Household Survey (IHS3). The IHS3 is a national representative survey carried out by the NSO and administered roughly to 12,000 households throughout 2010. A sub-sample of the 2010 IHS3 households will be revisited in 2013 as part of the IHPS. A similar integrated design is being proposed in Ethiopia, where the panel survey will be nested within the Agricultural Sample Survey and be based on the existing system of resident enumerators. In Nigeria, the survey has been integrated with the General Household Survey (GHS) which the National Bureau of Statistics carries out annually.

Such efforts to build on planned or on-going surveys, however, is not always be possible, depending on the features and objectives of the planned survey instruments. Even where feasible, they may come at a high cost by potentially limiting flexibility in terms of survey design and content. In this respect, one possible complication in fully integrating an extensive agricultural module into an existing survey is the timing of the planned survey (*vis a vis* the agricultural cycle) and/or the planned field organization. A second concern arising from integrating the agricultural module into an existing survey relates to the timing and frequency of the interview. Because of the nature of agricultural production and its heterogeneity in terms of cropping systems and seasonality, information should ideally be collected at the most appropriate times during the cropping season, and at higher frequency than allowed by a traditional survey. Hence, linking the administration of the agricultural module to an on-going data collection effort based on a single visit may have negative consequences both on the quality and quantity of information one may wish to collect. These problems may be even more acute concerning cases in which the data are collected over the course of 12 months, as in Tanzania and Uganda, where equally sized sub-samples are visited each month. Structuring the fieldwork in this way creates differing recall lags for different households in relation to specific event such as planting (input purchase) or harvest (crop sales). The extent of the potential biases due to the timing of survey administration is not clear, and has been analyzed by members of the LSMS team (Beegle, Carletto and Himelein, 2010). To partly, address this issue, the project has introduced a second interview to the household at a different more propitious time of the year has been added, when deemed feasible. For example, information on main transactions and other time-sensitive information related to agricultural production and marketing (e.g. input purchased and application, crop harvesting and sales) is being collected by dispatching enumerators to the households at least once more during the 12-month reference period.

***Nationally Representative Sample***

The overall sample design for each country emphasizes two aspects: drawing a valid and nationally representative sample and avoiding excessively large samples which would compromise the quality of the data and sustainability of the system, as well as the ability to ensure low attrition in rounds following the baseline. Depending on the specific country and its policy-relevant administrative and agro-climatic sub-divisions, the sample will also be representative at sub-national levels i.e. for rural and urban areas, main agro-ecological zones and, whenever feasible, for second-level administrative level such as states/provinces/districts.

The starting point and first methodological issue in sample design is the choice of an appropriate sampling frame from which households are eligible for selection into the survey. In the African context, this is not necessarily a straightforward task. In line with LSMS tradition and with recent guidelines for sampling agricultural surveys (FAO, 2005), a population-based sampling frame is being used. Even though a population-based listing would exclude commercial agricultural holdings, using the household as a unit of observation – instead of the agricultural holding – seems more feasible given the broader focus of the proposed project. Also, in the African context, smallholders still represent the vast majority of agricultural production and their relevance for poverty reduction makes it a more appropriate target of the proposed system. However, depending on the objectives of the survey, and the importance of the commercial sector in a specific country, the need to complement the agricultural data from the proposed household surveys with specialized surveys of commercial farms should be contemplated in some cases.

With the exception of Ethiopia where only rural household will be considered, nationally representative samples of both urban and rural households have been selected directly from the population census listing. Selecting PSUs based on a probability proportional to size (PPS) may not fully capture the composition of the agriculture and livestock sector. However, in view of the multi-purpose objective of the survey, this approach is considered more appropriate. In view of the large proportion of households involved in farm activities in Sub-Saharan Africa, a standard multi-stage cluster design is considered

adequate to select a sufficient number of farm households. In the case of Ethiopia, as already done for the Rural Investment Climate survey, non-farm households will be oversampled from the listing.

Countries often require information at highly disaggregated geographical levels. The LSMS-ISA system, however, emphasizes the estimation of key welfare and agricultural indicators at the national levels and at rather large sub-national sub-divisions, e.g. urban and rural and/or major regions/areas, such as main development regions or agro-ecological zones. More geographically disaggregated estimates may be pursued by means of agricultural censuses and administrative data collection. Small area estimation can also be pursued on some variables of interest provided that the necessary complementary data sources are available.

The project envisions a panel survey sample size of around 3,000 households in Tanzania, Uganda, Malawi and Niger, 5,000 households in Nigeria, and 7,000 households in Ethiopia. The proposed target sample size is based on the LSMS program's extensive experience in conducting multi-topic household surveys. It reflects the balance between having a sample large enough to be representative at national and sub-national levels, and data quality. Smaller samples is also important to ensure greater sustainability of the system beyond the project lifespan.

The exact sample size has been determined in each country based on specific requirements and in consultation with a sampling expert. For instance, if the objective is the estimation of production for the main crop(s) at the national level, the required sample size ended up being small, thus closer to the proposed lower-bound of 3,000 households. However, if the estimation of some commercial crop of interest is being pursued, or representativeness is to be achieved for several agro-ecological zones, cropping systems and lower-level administrative sub-division, then the required sample size is larger, as in Nigeria and Ethiopia. The clustering of the sample (via the most common approach to sampling, two-stage sampling) and other design parameters also enters into sample size and power calculations.

***Geo-referencing and Improved Links to Other Data Sources***

Another important feature of the project is to ensure that new data can be linked to other data sources. To date, the lack of linkages between data sets has plagued household surveys and dramatically reduced the usefulness of the individual data sources. Taking advantage of GPS technology is one such way to improve the 'hand-shaking' of household survey with geo-referenced data sources[7], which will also be crucial for studying climate change mitigation strategies through the LSMS-ISA surveys. The drastic reduction in the cost of GPS equipment and the impressive improvement in their precision and ease of use makes geo-referencing household surveys an inexpensive and practical exercise to carry out routinely during fieldwork. Geo-referencing of households is also extremely useful in panel survey contexts as it helps to track households from one wave to the next. Finally, GPS technology is being used, to the extent possible, for plot measurement. Issues of confidentiality may limit the dissemination of geo-referenced information on an open-access basis but recent developments in techniques may help ameliorate this potential problem.

Data sources can be also linked by other means, including sampling and overlapping in content. An example of the latter is the use of small area estimation techniques for poverty mapping developed at the World Bank (see Elbers, Lanjouw and Lanjouw, 2003, for details on the methodology). The method relies on the availability of a set of common variables between two different data sources, typically a population census and a household survey, to estimate rates of poverty or other indicators at levels of disaggregation not allowed by the household survey because of its sample size. This, and similar methods, demonstrate what proper planning and *ex-ante* coordination across data producers and users can achieve with little or no additional resources. In this respect, the ongoing 2010 Round of Population and Housing Censuses in many countries provides a unique opportunity to ensure that the proposed system fully benefits from contextual overlapping to enable this type of linkages.

[7] Examples of this include linking the national household survey in Malawi with the roads network to map travel times along primary/secondary/tertiary roads to trading centers, and the linking of rainfall data to the rural investment climate survey in Ethiopia to examine the impact of rainfall variation on non-farm and farm activities.

## III. Methodological Validation

An initiative of this scope and duration can make an important contribution to resolving methodological issues and measurement problems. Survey data in general, and income data in particular, are subject to serious measurement problems; agricultural income and production data are no exception. Of particular relevance to the proposed project are the well-known deficiencies in the valuation of agricultural production for home consumption – affecting our measure of welfare – as well as common problems in the measurement of plot size and yields.[8] Methodological research is needed and the project provides an ideal platform to field validate some of the more pressing measurement issues related to agricultural data. Validation exercises are being designed and conducted within the proposed work at only marginal effort and cost. The focus of the LSMS program in the World Bank in the past few years has been on such methodological innovation and validation.[9] Additional funds have also been raised from other sources to carry out some of the methodological work. Examples of such validation work is the collaborations with the World Fish Center for the development and field testing of an improved fishery module and the drafting of guidelines, the collaboration with the International Fund for Agricultural development (IFAD) for the development of a module to measure community governance and empowerment, and the partnership with the World Food Programme (WFP) to design and validate improved modules for the measurement of food security and vulnerability. In addition, in Uganda, the project is validating the use of crop diaries against recall methods, while in Ethiopia methodological work on comparing crop cutting to alternative crop estimation methods may be conducted. In partnership with FAO, a number of field experiments on assessing

[8] It is often difficult to collect information on own production, and more so for frequently or continuously harvested crops like cassava, sweet potatoes and bananas. One example of an innovation in survey methods is the use of a crop card, as described in Ssekiboobo (2007). In Uganda, the National Household Survey (2005/6) included the crop card to complement the 2 rounds of the household survey. A crop card was given to farm households on which households would record harvests. A crop monitor in the village would visit at least once a week to review the card. Nevertheless, the card was not viewed as completely successful. Crop monitors did not visit the households regularly, so problems with incomplete cards, particularly for illiterate households, were not addressed.

[9] See "LSMS IV: Research for Improving Survey Data," LSMS Team, Development Economics Research Group, World Bank, 2006.

the validity of alternative methods to measure agricultural and non-farm income are also being developed. Finally, jointly with the Agricultural and Rural Department at the World Bank, similar work be carried out on livestock data as part of the project. Priority areas for this work are better quantification of livestock production and income, including the valuation of non-tradables and animal by-products, as well as improving the efficacy of household surveys to capture pastoralist communities and their livelihoods.

In addition, the project is supporting the development of a standardized (yet customizable) CAPI package to empower client countries to independently implement surveys on UMPCs. Developing a standardized system which, in turn, can be configured for each country, has a number of advantages, particularly in relation to lowering per unit costs (economies of scale) and fostering standardization of coding systems.

A number of features make a CAPI application on UMPC particularly suitable for the proposed system, including:

a) Interviewing and data entry are brought together so that full validation can take place during the interview;
b) Data from past interviews can be used in the current interview (pre-filling, feed-forward determining follow-up questions). This holds for both panel surveys (with one visit per year) and for surveys which use multiple visits within the production period;
c) The computer can offer background processing of complicated aggregates (such as crop and livestock margins) which can be compared with physical and financial norms as part of a comprehensive data validation system.
d) It allows dynamic multi-language support and the ability to enable/disable questions or whole modules depending on whether they are relevant to the survey;
e) Automated data capture is possible, for example GPS readings, sound recordings and digital photography. Also, for plot-specific questions, the ability to display capture and display a sketch map of the plots for reference purposes while probing.

f) UMPCs offer an optimal trade-off in terms of portability, screen-size, storage and processing power, making them ideal for complex surveys;
g) Data are uploaded daily, via GPRS or CDMA modems, to an FTP server and available immediately to the end-users;
h) Immediate, context-sensitive access to the interviewer manual during the interview;
i) Capability to distribute updates electronically (no need for reprinting questionnaires)

## IV. Institutional Framework

One of the biggest challenges of the program is to identify and work toward creating the necessary institutional framework to ensure relevance, effectiveness and sustainability. The current institutional set ups of most countries still reflect an antiquated paradigm in which the ministries of agriculture have maintained an all-encompassing role in the development and growth agenda of rural areas, and thus have traditionally been in charge of the collection of agricultural data. Although the necessary institutional transformation may go beyond the possible reach of a specific project such as this one, it is important to emphasize its importance and propose a number of measures which can foster and accelerate this institutional transition. In this respect, the multi-topic nature of the proposed data collection provides a useful platform to foster inter-institutional collaboration.

Several relevant forums and initiatives already exist in most developing countries; tying this project to one or more of these initiatives will help raise the profile of agricultural and rural development issues in the broader policy agenda of the countries. The Millennium Development Goals' committees or the Poverty Reduction Strategy Paper Working Groups established and operating in several countries are examples of venues through which awareness and use of household surveys in general, and agricultural data in particular, can be promoted.

Second, avoiding duplication and taking advantage of existing surveys and plans is critical. Sixteen of the Sub-Saharan African countries already have National Strategies for Statistical Development and another sixteen are in the process of developing such strategies (http://www.worldbank.org/data/countrydata/csid.html). The work will be aligned closely with these strategies.

Finally, in several countries, Data User Groups (DUGs) have already been established to foster inter-ministerial collaboration and encourage productive discussion between data producers and data users. Where such groups do not exist, the project is working to establish them. Although the experience to date in advancing the rural development agenda through these venues has produced mixed results, the aim is to step up the efforts and raise awareness of agricultural issues through better data and stronger analytical capacity. Close coordination with agricultural policy initiatives and actions also being supported by the Gates Foundation will be maintained. Finally, as already mentioned, collaborations have been established with the Collaborative Masters Program on Agricultural and Applied Economics (CMAAE), the World Food Program, FAO, IFAD, the World Fish Center and ILRI, among others.

## V. Lessons learned to date

The complexity of a project of this nature makes it an ideal workbench for learning. At different stages of development, the project is currently working in six countries; irrespective of the often sizable differences across countries, the early entrants have already provided a wealth of knowledge which is informing the design of the programs in the other countries. To date, perhaps the main difficulty relates to reconciling two of the features of the projects: integration into an existing system with enough flexibility to make the surveys fully comparable across countries. Integration into existing survey system and, when possible, with scheduled surveys comes at a price, and striking the right balance between country-specificity and comparability over time, on the one hand, and cross-country comparability, on the other, is not always easy. Also, the concept of panel surveys and tracking remains a rather novel idea with statistical offices in the region, often resulting in an underestimation of the required effort and resources. While

the project is emphasizing this aspect of implementation, it may take a few years to be fully valued and embraced by all stakeholders in the countries. Furthermore, building stronger institutional linkages between national statistical office and the relevant line ministries remains a challenge, particularly in some countries. The project's Technical Working Group established in each country has provided a useful forum to elicit inputs from the line ministries and other sector stakeholders, but the political debate surrounding agricultural statistics continues to hinder a more efficient dialogue. Finally, the full integration of advanced data quality control systems based on intelligent, field based data entry while fully embraced by each country, must be perfected, particularly in what pertains the logistic aspects of data management. The CAPI application being developed under the project must ultimately rest on a flawless data management system if one wishes to draw full benefits from the investments.