Improving the Quality and Policy Relevance of Household-Level Data on Agriculture in Sub-Saharan Africa

A project by the World Bank Development Research Group funded by the Bill and Melinda Gates Foundation

I. Background and Rationale

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 recently released 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 present 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. Over the coming year, the World Bank alone will provide financing of approximately USD 6 billion in support of agricultural and food production around the globe. Similar efforts are underway from other donors and international agencies, including the Bill and Melinda Gates Foundation, to support farmers in the areas of

---

1 For additional information on this project, please contact Gero Carletto (gcarletto@worldbank.org), Katherine Scott (kscott1@worldbank.org), Kathleen Beegle (kbeegle@worldbank.org) or Diane Steele (dsteele@worldbank.org).
irrigation, water management, inputs and technologies, and crop diversification. The need to evaluate the effectiveness and pro-poor nature of these interventions is of paramount importance.

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 (Bryceson, 2002; Reardon, 1997; 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. Improved multi-sectoral data that document those linkages will 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. Of the 44 countries in Sub-Saharan Africa rated by the Food and Agricultural Organization, 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 (Tiffen, 2003; Ngendaumana, 2001). Data are 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).

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. Moreover, 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 (from donors, for example) 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 Offices (NSO), or other line Ministries (such as labor, education and health). 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 rethinking 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.

2 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).
A third important cause of poor data is the lack of analytical capacity in developing countries 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 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.

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 valuation of agricultural output used for home consumption and the measurement of crop yields and plot size. 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. 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.

This proposal outlines an ambitious project that entails the development and implementation of household panel surveys in six Sub-Saharan African countries. The project’s over-arching objective is twofold: to fill gaps in our knowledge about agriculture, and to contribute to improving the quality, relevance and sustainability of
agricultural data systems. The proposed set of surveys are to be integrated into each country’s existing system of household and other surveys, with the aim of building on experience and strengthening the quality and relevance of future efforts. In each country, the proposed data collection effort will consist of a nation-wide, household survey modeled on the Living Standards Measurement Study (LSMS) surveys that will be carried out every three years as a panel. The six countries in this proposal will be selected in order to maximize synergies with other on-going panel efforts in a few countries in the region including the upcoming National Panel Survey of Tanzania – being partially supported by the Gates Foundation – and panel surveys in Ghana, Zambia and Kenya. Given the dire need in rural economies for direct comparisons over time using comparable measures, the proposed program has the potential of transforming our understanding of development in Africa, particularly agriculture and linkages between farm and non-farm activities, and offers a tremendous opportunity to evaluate interventions in agriculture and beyond.

The household panel surveys would focus on agriculture but would use a multi-topic survey instrument as a base. Thus, both non-farm income and agricultural activities would be captured, together with a rich set of multi-sectoral information, for a better understanding of farm production, households and the rural economy. In spite of its focus on agriculture and rural areas, the survey sample will cover both rural and urban areas for a better understanding of geographical mobility. Although the present proposal only covers six and a half years, a 15-20 year time horizon is envisaged to allow a better understanding of some of the key issues at the heart of development and agriculture such as environmental transformation and adaptation, institutional change as well as geographical and socio-economic mobility. The project will link various users (analysts and policy-makers) with producers of data to improve the quality, relevance, and use of the data. This will also pave the way for increasing analytical capacities of survey countries by potentially breaking the vicious cycle of poor analysis undermining the demand for quality data. Finally, the proposed system will promote better linkages with other data sources, such as geo-referenced information and administrative data, and provide an empirical tool for the field validation of other data sources and measurement.
techniques, including “ground-truthing” of aerial photography and remote sensing. This multi-topic and multi-layered approach is essential to generate information to improve evidence-based policy making.

The project will be led by a team from the Living Standards Measurement Study (LSMS) program in the World Bank (WB). The LSMS is widely considered the gold standard for multi-topic household surveys and the LSMS team brings the right mix of field experience, analytical capacity, institutional knowledge and research on methodology to successfully implement the proposed project. For almost three decades, the LSMS team has been behind the global drive for better household survey data. Thanks also to the LSMS program, the number of developing countries with reliable poverty figures has increased dramatically over this period. A large share of the data behind the Global Poverty Monitoring Database (110 countries and more than 550 data points at 2008) and the poverty figures routinely used around the globe have either been produced by the LSMS team or been collected using methodologies and technical advice supported by the LSMS team. The proven track record in spearheading and leading methodological innovation and data collection efforts in the area of poverty and inequality will now be applied to the new challenge presented by deficient, yet essential, statistics on food, agriculture and natural resources in Sub-Saharan Africa.

Lastly, this project will not operate in isolation, but will be conducted in collaboration with on-going efforts to improve data systems in Africa, including those by the World Bank (such as the initiatives for statistical capacity building) and other donors. The World Bank team is currently collaborating closely with the ICRISAT team working on a similar survey system in South Asia under the auspices of the Gates Foundation. The LSMS team is also working closely with other institutions currently involved in panel surveys in Sub-Saharan Africa (e.g. Yale University, Michigan State University and IFPRI, among others). The proposed project will provide a forum for technical discussion towards the harmonization of methodologies and concepts. As importantly, fielding surveys that are linked with existing data collection efforts will permit the costs of this project to be well below those associated with stand-alone surveys, and reduce the
burden on government agencies and respondents, ensuring greater sustainability in the long-run.

This proposal has benefited from extensive consultation with different departments at the World Bank – including the Africa Region, the Agriculture and Rural Department, and the Development Economics Data Group – and from outside contributors, convened for a technical review meeting of this proposal in May, 2008 under the aegis of the Bill and Melinda Gates Foundation. The main recommendations of the technical group have now been incorporated into this proposal.

II. Project Objectives

The over-arching objective of this project is to improve our understanding of agriculture in Sub-Saharan Africa – specifically, its role in poverty reduction and how innovation and efficiency can be fostered in the sector. This goal will be achieved by developing and implementing an innovative model for collecting agricultural data in the region. The focus will be on overcoming some of the main technical and institutional shortcomings of the past and existing systems. The project sets out an ambitious agenda of activities and recognizes that the cost-effectiveness and sustainability of the proposed activities are pivotal to long-run success. To meet the objective, the proposed work is based on four tenets: a multi-sectoral framework, institutional integration, analytical capacity building, and active dissemination.

First, the collection of agricultural data must be integrated into a broad, multi-sectoral framework that goes beyond rural areas. This facilitates the production of data necessary to design effective agricultural policies country-wide and within the broader rural economy.3 Second, the collection of agricultural data must be buttressed by a well-matched institutional setting conducive of collaboration and integration of data sources.

---

3 Note that agriculture is also important in 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).
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 analytical capacity to analyze data as well as capacity to produce timely and relevant public-use data sets. The lack of analytical capacity in developing countries perpetuates low demand for data. In turn, this negatively affects the quality and availability of policy-relevant analysis. Poor dissemination of the available data and results has further aggravated the problem.

III. Project Design and Implementation Plan

This project will invest in the creation of national capacity to develop and implement household surveys and analyze household survey data with a strong emphasis on agriculture. New data will be generated through the design and implementation of panel surveys at the household level over the next six and a half years. In addition to the six countries to be covered under this proposal, work on panel surveys is already on-going or being planned in a few other African countries under the auspices of different institutions, thus enhancing the overall impact of this project on the availability of panel data on agriculture in the continent. The World Bank and the Foundation are already working together in a seventh country, Tanzania, in the design and implementation of the National Panel Survey being carried out by the National Bureau of Statistics. Some other countries are Mozambique (Michigan State University (MSU), World Bank and other donors), Ghana (Yale University), Zambia and Kenya (MSU) and Zimbabwe (World Bank, Duke University and University of North Carolina). Close collaboration with these institutions will be kept during the project, and synergies explored. The capacity to generate and use relevant data will be created through the development of an institutional framework of collaboration and skill building that will be fine-tuned throughout the life of the project and adjusted to specific country needs and existing capacity. Ultimately,
the goal of the project is to produce a model and a set of tools that can be adopted by governments and donors to scale up the effort to other countries in the region.

## III.1. Design Features of Proposed Household Survey System

### III.1.a. 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; this allows 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). More recent examples of large-scale household panels include 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 partial panel was conducted in 2007, following the 2004 national household survey in Malawi (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.
This project calls for a panel survey wherein households would be re-interviewed every three years. It is believed that a three-year spacing would provide an appropriate interval to monitor dynamics of most variables and get a good understanding of changes in behavioral and agronomic patterns.\(^4\) 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, the possibility of carrying out complementary data collection exercises in the interim years may be contemplated if additional funding becomes available\(^5\).

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 is 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. This was the approach in the Vietnam Living Standards Surveys panel rounds (1992/93 and 1997/98).

In addition to potential re-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.

**III.1.b. Multi-topic, multi-purpose**

\(^4\) A similar spacing is being adopted by Yale University in their Ghana panel survey, and by MSU in Mozambique.

\(^5\) These complementary data collection activities could take the form of lighter annual surveys, in-depth village-level studies on a subsample of households or be part of the country’s routine administrative data systems.
The proposed panel survey will be centered on a multi-topic household survey modeled on the LSMS survey of the World Bank. 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. 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 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 survey at times also collects 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 will be gathered through the proposed survey system, and similar

---

6 For details of the LSMS 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.

7 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).
gender-sensitive methodologies will be utilized for a better understanding of women’s role in agriculture and the rural economy.

There are countless examples of how LSMS-type data have been used to inform policy-making and effectively contribute to improving countries’ provision of services. Examples are the use of poverty mapping for the targeting of social assistance programs and public spending in several countries, such as Nicaragua, Guatemala, and Albania, the redesign of Social Investments Funds following evaluations in Nicaragua, and the reform of assistance programs, including food stamps in Jamaica, kerosene subsidies in Ghana, and subsidized food prices in Tunisia (Grosh, 1997).

Since the project aims at operating, as much as possible, within the existing system of surveys, the specific design across countries may vary. 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 will be different across countries in terms of content and depth of each module, each questionnaire is likely to contain information on the following topics in addition to agriculture:

- household demographics
- education
- health and nutrition
- food consumption/expenditures
- non-food expenditures
- dwelling conditions
- assets
- employment
- agriculture and livestock
- internal and international migration
- participation in projects and programs
- non-farm self-employment and other sources of income

Information will be collected to allow gender-disaggregated analysis. For example, information on asset ownership, control of resources and participation in activities and programs will be differentiated along gender-lines. The use of female enumerators and direct interviews to female respondents, will improve the accuracy of the information collected and will allow for the collection of data to design better 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. 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, as part of this project the LSMS team will develop a core module that will serve as a general template for each country. A prototype module is currently being developed as part of the on-going Tanzania work and will provide the basis for further development in the countries under this proposal. The on-going Ghana panel survey by Yale University and a possible 4th round of the Temegoco survey in Kenya by the World Bank will also provide useful input into the design of this agricultural module, particularly with respect to questions on property rights, social networks and gender. The template will be 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.⁸

The core set of broad indicators to be covered in the agricultural module have been identified by the LSMS team through a consultative process with several experts, and during the technical meeting held in May 2008 as part of the project preparation. 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. Moreover, in some specific cases, these measures are used as inputs into Systems of National Accounts and for forecasting (i.e. to predict food shortages). It is envisaged that the Project Steering Committee and members of the Technical Advisory Panel will revise and elaborate further on this list over the coming months. Clearly, flexibility will be 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 and livestock production, storage and sales
- Productivity of main crops, with emphasis on improved measures of:
  - quantification of production
  - plot size
- Land holdings

⁸ 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).
- Access to and use of services, infrastructure and natural resources
  - Agricultural extension services
  - Infrastructure (including roads)
  - Credit (both for agriculture and other purposes)
  - Education and health
  - Market access
  - Access to information
  - Access to natural and common property resources
- Input use and technology adoption
  - Family and hired labor
  - Use of technology and farming implements
  - Seed varieties
  - Fertilizer, pesticides
- Shocks
- Farming practices
  - Mechanization
  - Soil and environmental management
  - Water management
  - Adaptation to climatic changes

A community questionnaire will also be 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.

Beyond the design of the questionnaires, the project will consider three additional aspects of standardization. First, the project will develop a standard data entry application to accompany the core agricultural module. Both recurrent data entry systems as well as Computer Assisted Personal Interview (CAPI) applications will be considered. For example, in Tanzania, on a pilot basis, the LSMS team is working towards the
development of a CAPI platform to be used to administer the agricultural modules in a paperless format, 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. The ultimate objective is to use the experience gained by the LSMS team in Tanzania and in other countries to develop a standardized (but customizable) package empowering client countries to independently implement surveys on UMPCs.

Second, the project will develop 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 and labor 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 module 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. 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.

**III.1.c. 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 will also vary 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 may be phrased and interviews administered, for example with respect to interviewing women. The LSMS team has ample experience in ensuring that this process of customization brings the highest returns both in terms of data relevance and quality. Tailoring in terms of content and design, however, must also take account of international standards and cross-country comparability. The proposed system will incorporate the criteria of country-customization and cross-country comparability in terms of contents as well as international nomenclatures and classifications.

**III.1.d. Gender strategy**
The gender strategy of this project encompasses two main areas: methods of data collection and gender balance of the team engaged in this project.

In regards to methods of data collection, this project recognizes the imperative to look beyond the household as a unit of analysis in order to improve the quality, relevance and sustainability of agricultural data systems. Although the project conceptualizes data collection efforts as part of household panel surveys, in actuality, much of the data are collected at the *individual* level.

As noted in Section III.1.b, information on all household members will be collected concerning their labor, education and health status. This includes time allocated to wage and non-wage activities, such as water collection, firewood collection, farming, and livestock maintenance. Within the agricultural component of the questionnaire, information will be collected on plot ownership, decision-making on plots and on credit access for farming (as well as non-farm activities). The community questionnaires will collect information on gender aspects of community leadership, social networks, customs and norms including those related to inheritance traditions and land access.

Individual-level data collection will address many of the issues raised in the BMGF Gender Checklist. Issues specific to agriculture include women’s access to irrigated land, women’s engagement in cash crop and food crop production, gender differences in the use of new technologies and access to credit, and whether and to what extent agricultural decisions are made by men, women or jointly. Moreover, this project extends gender-disaggregated data to non-farm issues, such as women’s operation of non-farm enterprises, and time allocation of all household members in economic activities. In fact, individual-level data collection within household surveys is the norm in the LSMS program.

The project team will also strive to attain gender balance. This is relevant in terms of data collection level as well as questionnaire development and analysis. It is important that field teams have female enumerators, particularly in regions or countries where the
gendered division of work and responsibilities is strong. Failure to recognize these aspects can compromise data quality. For example, the LSMS team is currently carrying out field research in Afghanistan to rigorously assess the gains in data quality as a result of using female enumerators for female respondents on different types of questions. While this may not be as applicable in most of the countries in which this project will operate, it illustrates the point that the gender composition of field teams can facilitate the collection of better data. In terms of questionnaire development and analysis, having a diverse team involved, not only from a gender perspective, will help to ensure that topics of particular relevance to specific groups are incorporated into the questionnaire design and are analyzed.

The LSMS team has experience in pioneering surveys that have pushed the gender envelope. This includes fielding household surveys with extensive modules on decision making in the household economy. As part of our ongoing research program, we are conducting survey experiments on the consequences of proxy respondents with respect to under-reporting of women’s labor force participation and collecting accurate information on women’s credit and access to finance (often reported only by male household heads). We are also developing a best practices guide for collecting survey data for gender analysis.

III.1.e. Integration into existing or planned government 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 (see point III.2 below) 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 countries in Africa 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 will be aligned closely with countries’ National Strategies for Statistical Development (NSDS), if these exist, and the project will work with countries presently implementing such strategies to ensure that agricultural statistics are integrated appropriately.

It is proposed that, whenever possible, the surveys under the project will be “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.

Such efforts to build on planned or on-going surveys may not always be possible, depending on the features and objectives of the planned survey instruments. Even where feasible, they may come at large costs 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 solution, as implemented in current LSMS surveys, is to split the questionnaire across multiple respondents (the most knowledgeable about each topic) and multiple visits. To ensure the depth and quality of the agricultural data, one can think of administering the agricultural module in its own ‘visit’. This approach was recently applied within the LSMS methodology and consists of including in the core survey a short screening module to identify all farm households in the sample. Based on this identification, a farm module is then administered to this sub-sample of households.

There are various benefits to this approach. First, it allows the agricultural module to be more in-depth than otherwise, while lightening the core questionnaire. Second, the agricultural module can be fielded at a more propitious time for agricultural purposes,
e.g. post-harvest. Third, different interviewers with more knowledge of agricultural issues can be used in this phase of data collection. This approach is being implemented in the Tanzania NPS, where dedicated agricultural enumerators will be embedded in the NPS field team with the explicit task of collecting agricultural data. Based on a short screening module in the core NPS, the agricultural enumerator will be provided with a list of households involved in agricultural and/or livestock production, for an in-depth follow-up interview covering different aspects of agricultural production. The same approach will be followed in those countries planning similar types of surveys, thus making such integration possible.

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. The problem may even be worse concerning cases in which the survey is planned over the course of 12 months, with a subsample visited each month. This is the case for the Tanzania NPS in which one twelfth of the sample is to be visited each month. This may create an additional problem due to different 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.

The project will address this issue in several ways. First, a second interview to the household at a different, more propitious time of the year will be added. For example, information on main transactions and other time-sensitive information related to agricultural production and marketing (e.g. input purchased and application, crop

---

9 Ideally, one could think of identifying when this better time would be based on idiosyncratic information collected from the household during the core survey, e.g. based on household-level cropping patterns and labor activities.
harvesting and sales) will be collected by dispatching enumerators to the households at least once more during the 12-month reference period. The possibility of using existing extension officers to carry out this complementary data collection is being explored in Tanzania and will be considered in this project. A potential advantage of using extension officers is that it would foster inter-institutional collaboration and ownership, while also promoting better integration with the existing administrative data collection systems which extension officers are generally responsible for. Provision is made in the budget for one additional follow-up interview for each agricultural household. Finally, initially on a trial basis, the household will be equipped with basic tools to support the collection of information in between visits and thus enhance the quality of the recall. Examples of such tools are crop cards, diary booklets, and simple measuring devices to facilitate the quantification of certain variables which may be difficult to properly measure through recall in a single, or even double, visit. 10

To fully explore available opportunities for piggybacking and as an input to the survey design, one of the early activities in each country will be to carry out a comprehensive assessment of the existing and planned household-level data collection activities. This will be done by fielding a first mission by a WB staff and, if necessary, the services of a consultant will be hired to contribute to the assessment. Taking advantage of existing surveys is the preferred mode of work under this proposal. However, in some instances, conducting the surveys in parallel to existing efforts may be more viable while in other instances it may be the only option. In light of its worldwide involvement in the design and implementation of household surveys, the LSMS team is best positioned to ensure that these opportunities for piggybacking are fully exploited. The on-going Tanzania work on the NPS is a vivid example of these synergies, as the opportunity of piggybacking for the Gates Foundation’s project arose because of the early involvement of the LSMS team in the design of the NPS. A similar opportunity has now emerged in Mozambique where the LSMS team is working with the Ministry of Agriculture and

---

10 For example, a tape measure could be given to the household at the time of the first visit for the measurement of cattle. Another example, recently adopted in similar types of survey, would be to provide the household with plastic buckets for the measurement of crop production for difficult to measure crops such as vegetables and tubers.
Michigan State University in the design of the 2008 National Agricultural Survey (Trabalho de Inquérito Agrícola). These are only two recent examples of synergies that can be achieved through the LSMS program.

III.1.f. Sample Design

The overall sample design for each country emphasizes two aspects: drawing a valid and nationally representative sample and avoiding excessively large samples which will 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 administrative and agro-climatic sub-divisions of policy relevance, 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 recommended for the proposed survey system. Even though a population-based listing will 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 lion share 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.
The ongoing 2010 Round of Population and Housing Censuses (PHC) is a critical input for the proposed survey system, as it has the potential of providing updated listing of households from which to draw our samples. According to recent FAO guidelines (FAO, 2005), such population listings could be the basis for sample-based agricultural censuses and any follow-up detailed survey on agriculture, provided that the PHC include a limited set of questions for the proper identification of farm households and/or farmers. Based on these guidelines, the sample of the proposed panel surveys could, in principle, be drawn directly from the Agricultural Census listing in turn derived for the PHC, thus guaranteeing full integration between the survey and the censuses.

While potentially appealing as a method, to date few population censuses have included the type of screening questions proposed by the FAO guidelines. Furthermore, some concern could be raised about the accuracy of the screening questions in a large undertaking such as the census, thus potentially leading to under-reporting of farm activities. Finally, even when screening questions are included in the population census, few countries in Sub-Saharan Africa are currently planning to implement an agricultural census. Thus, full integration of the panel system with the population and agricultural census will only be possible in a handful of countries, at least for the time being. The availability of information in the population census and a recent agricultural census will be one of the criteria to be used for country selection. A country where this approach may be exploited is Mozambique. Mozambique carried out a population census in 2007 and questions to identify land and animal ownership were included. The country is currently planning an Agricultural Census to be conducted in either 2009 or 2010, thus making Mozambique an ideal candidate for using this approach for the baseline round of a new panel survey.

---

11 Information on asset ownership is generally considered too sensitive in censuses and thus prone to under-reporting. It must be noted, however, that such information from the population census even if not accurate in terms of quantifying specific variables such as land or livestock ownership could be sufficiently adequate for constructing a dichotomous variable for the identification of households involved in agricultural and/or livestock activities.

12 In those cases in which the population census includes the screening questions and these are deemed reliable, one could think of creating a master sample of both farm and non-farm households and use this master sample for the selection of the panel sample and future sample refreshments. The “farm stratum” of the master sample would be the listing to be used for the Agricultural Census, if and when it is conducted.
In other countries where no agricultural information is available from the census for stratification purposes, a nationally representative sample of both urban and rural households will be 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 adequate.

Depending on the “age” of the population listings, updating will be required and is included in the budget. In view of the large proportion of households involved in farm activities in Sub-Saharan Africa, a standard multi-stage cluster design will be adequate to select a sufficient number of farm households. Screening questions, similar to the ones proposed by FAO for the population census, would be included in the core survey questionnaire to identify all households involved in farm activities. These farm households would then be administered an additional, detailed agricultural module using specialized enumerators.

Countries often require information at highly disaggregated geographical levels. The proposed system, however, will emphasize 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 should be pursued by means of agricultural censuses and administrative data collection. Small area estimation can also be attempted on some variables of interest provided that the necessary complementary data sources are available.

The project envisions a panel survey sample size between 3,000-5,000 households per country; the actual size will mostly depend on the actual number of chosen domains of inference in each country, the relevant policy questions and related power calculations. Assuming that 80 percent of the sampled households are involved in some form of farm activity, 2,400-4,000 farm households will be interviewed in each country out of the total
sample. Oversampling of farm households may also be possible, depending on the expected prevalence and the final sample size requirements. 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 will also be important to ensure greater sustainability of the system beyond the project lifespan. In countries with dire needs for more disaggregated data and with proven track record in complex data collection operations, and if additional funds are available from other sources, it may be possible to expand the sample to pursue representativeness at lower levels of disaggregation. The exact sample size will be 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 may end 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 may have to be larger, thus closer to the proposed upper-bound (5,000 households). The clustering of the sample (via the most common approach to sampling, two-stage sampling) and other design parameters will also enter into sample size and power calculations, suggesting that the proposed figures should be taken as an approximate range. For a more detailed exposition on sampling see Annex G. Finally, in order to maintain the national representativeness of the data, survey samples will be refreshed partially or fully every number of years, depending on the levels of attrition and mobility in each country.

III.1.g. Geo-referencing and improved links to other data sources

Another important feature of the project will be 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. 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 will be used, to the extent possible, for plot measurement. In collaboration with FAO, a background paper summarizing experiences to date and providing methodological guidelines will be produced as part of the project and disseminated at a workshop. 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 great opportunity to ensure that the proposed system fully benefits from contextual overlapping to enable this type of linkages.

The survey system is also to be used for “ground truthing” of aerial photography and remote sensing. In order to develop more precise models for predicting production and forecasting food shortages, there is a need to validate the accuracy of remote sensing

---

13 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.
photographs. Manual measurement of plot size will be carried out during fieldwork, either through GPS or other means; digital pictures of selected plots identified via satellite will also be taken. The technique is rather simple and would only require procuring some basic equipment (digital cameras and tripods) and administering simple training to the enumerators. In addition, soil testing will be carried out on selected plots for the dual purpose of validating remote sensing readings and self-reported measures of soil depletion. The LSMS team will collaborate with the International Food Policy Research Institute (IFPRI), FAO and other CGIAR centers in developing and applying a basic protocol for the ground truthing of remote sensing material and soil quality measures.

III.1.i. Validation of methods

An initiative of this scope and duration can make an important contribution to resolving methodological issues and measurement problems. The ground-truthing of remote sensing mentioned above is one such an example. 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.\textsuperscript{14} Methodological research is needed and the proposed project can provide an ideal platform to field validate some of the more pressing measurement issues related to agricultural data. Validation exercises could easily be conducted within the proposed work with minimal additional effort and cost. In most cases, methodological work will be pursued on a smaller random sample of households. The focus of the LSMS program in the World Bank in the past two years has

\textsuperscript{14} 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.
been on such methodological innovation and validation.\textsuperscript{15} As with the ongoing validation exercises, the LSMS team will engage other researchers within and outside the Bank with specific interests in these issues. Each proposal for validation will have to be approved by the project Technical Advisory Group and the Technical Working Groups that will be set up for each country (see section VI for more details). Additional funding may also come from other sources to carry out some of the methodological work.\textsuperscript{16}

In the course of the project review meeting last May, a number of priority areas and measurement challenges were identified and include:

\begin{itemize}
  \item GPS measure of plots and validation of self-reported estimates;
  \item estimating yield under intercropping;
  \item better quantification of roots, tubers and vegetables harvest;
  \item measuring soil type and quality;
  \item valuation of own production and agricultural income;
  \item interlinked contracts and marketing arrangements;
  \item accounting for nomadic population;
  \item measuring and valuing livestock by-products (e.g. milk)
  \item better accounting of fishery and aquaculture
\end{itemize}

Validation of these and other issues will be carried out on an \textit{ad-hoc} basis based on rigorous protocols to be developed by the LSMS team with collaborating institutions. A minimum of 12 validation exercises will be carried out in the course of the project. For the plot size exercise, for example, the objective would be to assess the accuracy of commonly used self-reported measures \textit{vis a vis} objective measures such as GPS readings. Similar types of validations can be carried out also in the case of soil quality or distance data. Provision is made in the budget for some of this work. Additional funding may be provided through \textit{ad-hoc} alliances with other institutions on specific research questions and validation exercises.

\textsuperscript{16} A possible approach is the one followed by the ICRISAT VLS, in which outside researchers were invited to submit proposals for add-on modules for specific research project and methodological validation exercises.
In addition, the project will support 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)

III. 2 Institutional Framework

One of the biggest challenges of the proposed survey activities will be 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 will provide a useful platform to foster inter-institutional collaboration.

Several relevant fora 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 proposed here 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 will work to establish them. As an active member in most of these country-sponsored initiatives, the World Bank is in a unique position to foster the integration of the proposed system within an existing strategic framework. 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, it will be important to promote collaboration with local and regional institutions like the African Economic Research Consortium (AERC) and the African Association of Agricultural Economists.

III.3. Capacity Building

The ultimate long-term goal of the proposed project is to create a sustainable system for the collection of agricultural and related household-level data. Although full financial independence of the system may not be achieved within the lifespan of the project, within the six and a half years of the project the aim is to build enough awareness and analytical capacity, and enable local counterpart institutions to maintain the system with less external technical assistance.

In this strategy, strengthening the analytical capacity of relevant government agencies in utilizing household-level data will be crucial. The proposed initiative presents a unique opportunity given its long-term horizon. The typical project cycle and one-off survey efforts have proven vastly inadequate to build the necessary in-country capacity to ensure long-term impacts. Furthermore, national statistical offices (NSOs) and line ministries such as agriculture (MoA) are often low on governments’ lists of spending priorities: public spending over the years in both agriculture and statistics has declined.
Building capacity presents a number of challenges. Staff in line ministries and statistical offices often do not have the training needed to design and analyze micro-level data. Also, many of the government agencies in developing countries are characterized by high staff-turnover. Providing the right incentives and building a sufficiently large pool of trained staff are important dimensions to consider.

A successful example of a relevant capacity building effort is the Program for Improving Surveys of Living Conditions in Latin America and the Caribbean (ISLC/MECOVI) supported by the Inter-American Development Bank (IADB), the UN Economic Commission for Latin America and the Caribbean (ECLAC) and the World Bank (http://www.iadb.org/sds/POV/site_19_e.htm). The ISLC/MECOVI incorporated both national and regional technical assistance and training components. The country-specific programs were usually bolstered by the presence of an in-country resident advisor for the initial years of engagement. External technical assistance and training in all aspects of survey design and analysis was provided by the program, in combination with regional workshops and formal training courses on selected topics. The different levels of expertise of each participating country and the sharing of experiences and know-how fostered South-South collaboration and cross-country fertilization. The MECOVI program was also instrumental in promoting an open-access data policy in the region.

Along the same lines, the proposed project envisions hiring a resident advisor for the first two years of the activities in each country. In addition to this Resident Technical Advisor, a combination of short-term technical assistance and training activities on specific aspects of data collection and analysis, combined with strong involvement by staff members from the World Bank, both from headquarters and the Resident Mission in each country, would be adopted. Two regional training workshops, one for Francophone and one for Anglophone Africa, will be organized on topics to be chosen based on consultations with beneficiaries and donors. A minimum of 20 trainees will attend each workshop. The Technical Working Group to be set up for each country will include local researchers and academics who will work closely with the resident advisor and WB staff, and in coordination with the other analytical efforts by the BMGF and other donors, in
generating and making use of the data for joint research projects. An example is a possible collaboration with a Masters Program for Agricultural Economists possibly being supported by the Foundation.

Engaging data producers and data users in a collaborative process extends from the development of survey instruments themselves to the analysis of the resulting data. This is a critical link that will create the possibility of sustainability of the project and promote the development of high quality and policy relevant data. While creating a national capacity for high-level micro data analysis is beyond the scope of this project, the project will support hands-on analytical training through the joint preparation of analytical work by local staff from relevant line ministries in collaboration with international researchers and staff from donor agencies. This link between data production and use is a critical component. It is expected that more in-depth assistance in this area will be provided under the policy component of the BMGF’s program that will be developed in each country.

III.4 Dissemination

A key feature of the proposed system is the public-good nature of the data, which will become available to the broad community of researchers and policy makers within 12 months of completion of data collection. By the end of the 12 month period, the data will be fully cleaned, documented and disseminated through various means, including a specifically designed website to publicize the project and disseminate the data, related documentation and any analytical findings. One alternative to an all-inclusive, one-stop website would be a simpler page with appropriate links to other country-specific websites where the data and documentation will be stored. The LSMS website will also be revised to include all information and links related to this project.

To ensure that data are put in the public domain in a timely manner, it will be important to negotiate and sign open data-access agreements with each country to ensure the data are available to all. The LSMS team has a long track record in ensuring that the data collected are fully documented and disseminated (for more details, visit the LSMS
website at [www.worldbank.org/lsms](http://www.worldbank.org/lsms). In addition, the project will coordinate activities with other data dissemination initiatives such as the Accelerated Data Program (ADP) and the International Household Survey Network (IHSN) of the World Bank and CountryStat of FAO.

As part of the dissemination effort, and to guide the project in identifying key knowledge gaps and areas of methodological interest, the project will also contribute to the preparation of four background documents and review pieces on several topics. Possible topics mirror the areas already identified in the section on methodological validation. For example, the project will work with the Statistics Division of FAO in preparing a document on techniques for improving plot measurements using GPS technology. Furthermore, three regional workshops to report on main findings and disseminate some of the research outputs will be carried out as part of the project. The project and its various findings will also be disseminated through presentations at conferences and workshops e.g. the International Conference of Agricultural Economists. Finally, the project will sponsor two donor meetings in year 1 and in year 4 (following the mid-term evaluation) with the goal of informing the broader donor community of the project and explore financing opportunities to scale up the project both in terms of number of countries and more future panel waves. The scheduled dissemination workshop will also be used as venues to share with other donors some of the main findings of the project and rally their support.

While there will be variation in the data collected to ensure country-relevance and to investigate key issues in different regions, there will still be a core set of questions and indicators that will be comparable across the six countries. To facilitate access to, and use of these data sets, harmonized, or comparable, data sets will be generated from the raw data. These new data sets, fully documented, will be incorporated into the Comparative Living Standards Project, a web interface for online data analysis developed by the LSMS team to calculate a range of indicators, from simple frequencies to on-demand multivariate analysis.
To enhance the usefulness of the data collected, timelines are crucial. The LSMS team has been at the forefront in the design and implementation of field techniques to minimize the lag between fieldwork completion and dissemination of results. The use of smart, concurrent data entry has considerably reduced the time required for preparation of the final, high quality, dataset for analytical purposes. More recently, the team has been experimenting with other techniques, including CAPI, which would farther reduce the delivery time and improve data quality. A CAPI system is being designed for the agricultural module of the Tanzania NPS and will be tested before full implementation.

As described in section III.1.h on Validation of Methods, based on the experience gained in Tanzania and in other countries, a standardized (yet customizable) CAPI platform will be developed.

III.5 Country Selection and Program Development

A total of 6 Sub-Saharan countries will be selected for this project, and a minimum of two panel waves, each consisting of a minimum of two visits to each household, will be carried out within the duration of the project. A phased-in approach is envisioned, with two new countries being added each of the first three years of the project. Countries will be selected based first and foremost on their Government’s interest in, and commitment to, the initiative for the medium- and long-term. Since the final selection of countries will depend on a number of criteria, some of which is impossible to resolve a priori\textsuperscript{17}, a shortlist of countries is proposed, which were identified in the course of the technical review meeting of May 2008 and in consultation with the operations departments of the World Bank. The short list has been developed on the basis of a number of criteria discussed at the May technical review meeting, and include:

1. availability of a recent population census (with agricultural information) and/or an agricultural census;
2. experience with multitopic household surveys and possibility of piggybacking onto existing surveys;

\textsuperscript{17} An example is the government’s commitment, which is difficult to gauge prior to engaging in full discussions on the project.
3. possible government buy-in, as also reflected in the inclusion of the panel survey in the country’s National Statistical Development Strategy;

4. geographical coverage of the project as a whole (Southern, Eastern and Western Africa; Anglophone and Francophone Africa; large and small countries, representation of different agro-ecological zones and cropping systems

The shortlist from which the 6 countries will be selected includes:

1. Mozambique
2. Malawi
3. Uganda
4. Ethiopia
5. Nigeria
6. Mali
7. Senegal
8. Niger
9. Cameroon

Although, Mozambique has a past panel, there is no plan for the continuation of the effort past 2009. Actually, the 2008 TIA survey, which is currently being planned, envisages neither a panel component nor the collection of comprehensive consumption-expenditure data for the construction of a poverty measure. Some countries like Ethiopia (Oxford 15-village study), Niger (ICRISTAT Village Level Study) and Malawi (IFPRI Complementary Panel Survey) have also carried out panel surveys in the past. Other countries, such as Mali, Zambia and Mozambique, are planning targeted data collection efforts to explicitly study specific crops. These efforts, while valuable and important, are not suitable substitutes for the proposed survey program as they neither are nationally representative nor, in most cases, foresee additional rounds in the future. They also often collect only partial information for the measure of poverty. The proposed project, nevertheless, will take these past efforts in due consideration and build upon them.

---

18 Discussions by the LSMS team with the Ministry of Agriculture and MSU have focused on the possibility of re-interviewing in 2008 a sub-sample of the 2005 TIA households and including a consumption module for the measurement of poverty.
The first six countries are the ones that will be prioritized. The six countries that are finally chosen based on the criteria listed above will be complemented by work on panel surveys currently on-going or being planned in a number of African countries, including Tanzania, Ghana, Kenya, Zambia, and Zimbabwe. Close coordination will be kept with the principal investigators of these studies to foster harmonization of methodologies, whenever possible.

The first step in selecting a country will be to review the contents of the national strategies for statistical development and present state of agricultural statistics in the sub-Saharan region. In parallel with this background review, the team will consult with other experts in the field to determine which countries might have interest in the project. The team will field a first mission to countries to discuss the project with the relevant authorities. In countries where there is strong interest, a national counterpart will then be selected and a work program drawn up. Depending on the existing survey infrastructure and activities, this work program could be an increment to what is being done or a substantial investment of time and resources in developing a new survey capacity. The team, with the government counterparts, will fully explore available opportunities for piggybacking. A Technical Working Group for the project in the country would be formed (see point VI below) and would approve the plan. Any funding gaps identified at this time will be closed through government funding or obtaining funding from other donors.

In addition to Tanzania, where the work has already started under a separate grant from the Foundation, work is expected to start in two additional countries in year 1. Preliminary informal discussions have already been held with the Ministry of Agriculture of Mozambique and with the National Statistics Office in Uganda to explore possibilities

---

19 In the Tanzania case, the funding from the Gates Foundation will support about 25 percent of the total funding in the first year. A total of four donors are contributing to the National Panel survey costs, in addition to a substantial contribution by the Government of Tanzania. In other countries, a full survey would need to be developed and implemented. Each country will be different as the goal is to build on existing work where possible but not limit ourselves to the more advanced countries with better survey infrastructure.
of collaboration. Discussions have been held with Michigan State University and IFPRI about possible partnerships in some of the countries in which they have worked in the past and/or currently have resident staff, e.g. Mozambique, Mali and Ethiopia. Also, the possibility of collaborating with the National Statistical Office of Malawi in the implementation of the next Integrated Household Survey scheduled for 2009/10 is being explored.

IV. Potential Risks

One of the major risks associated with the project derives from the low technical and institutional capacity in the partner countries. For this reason, the project envisions strong and prolonged technical assistance activities, including the presence of a Resident Technical Advisor in each country for the initial two years of the survey and the strong involvement of World Bank and other technical advisors in all stages of project design and implementation. A second risk derives from the fact that, in order to reduce costs, decrease the survey workload on often under-staffed national institutions and help foster sustainability, it is preferred that the project be integrated with existing data collection activities planned by the partner institutions. Consequently, the timing and success of the panel survey, in some settings, may depend on the established parameters of the survey to which this component is linked. The option of running a separate survey will be used if it is felt that the cost of integration outweighs its benefits. Developing a stand-alone new survey is, in many cases, more easily done. However, the costs will increase and the sustainability of the data collection effort would be expected to be less than if piggy-backing with an existing survey were possible.

Furthermore, the lack of institutional coordination between the National Statistical office, the Ministry of Agriculture and other relevant line ministries presents another potential hurdle to the successful implementation of the project. The isolation of various government agencies is exactly what this project is aimed at breaking; but there are no “silver bullets” and the success of this effort may vary by country.
In addition, high staff turnover will be a challenge in ensuring continuity and sustainability in the long-run. Technical assistance is only effective when there is a core team to receive it and convert the knowledge received into institutional practices. If staff salaries in the statistical offices of the Ministries of Agriculture are low or institutional prestige is low, staff turnover will be high. Additionally, the act of training staff in these agencies can increase staff turnover as the staff may be able to find better jobs. While there is little that this project can do in terms of salaries and job opportunities, improving data quality and raising the relevance of what is done may increase institutional prestige and/or provide other incentives that may limit turnover.

In countries with limited capacity, there is the risk of overburdening the system by adding a complex survey that will need to be implemented by the government. To minimize this risk, the project aims to build onto existing activities. Even with this focus, there may still be problems, such as crowding out other activities. The TWG for each country will carefully review the existing and planned survey activities as well as the capacity of the country to ensure that the proposed survey will not generate problems for either the NSO or the Ministries of Agriculture.

Finally, government commitment is a critical, yet difficult, component of the project. Unless there is a high level of commitment on the part of the government, the project simply will not function well and will, certainly, have little chance of sustainability. For this reason, it will be important for the project management to have the option to “walk away” from working in a specific country unless the relevant counterpart shows enough commitment to the initiative. Obtaining this commitment may also take more time than envisaged. Clearly, there will always be the risk of political change that will undermine government commitment and/or its ability to maintain commitments. This will have to be monitored closely in each country.

V. Monitoring and Evaluation
The overall monitoring of the project would be done by the Project Steering Committee (PSC) (see point VI below). A mid-term review of the project would be carried out through hiring an independent consultant to assess the activities to date. A meeting of the Technical Working Groups for all countries involved would be held to discuss the Mid-Term Review document and make proposals for improvements, changes and further assessment if needed. A final report would then be prepared and submitted to the Project Steering Committee for its review and approval.

Given the complexity of this project, the inevitable adjustments will surely need to be made during the course of the project, and the work plan, including a revised timeline, will be reviewed and approved by the PSC every year. Any major technical change will be submitted by the PSC to the Technical Advisory Board for comments.

Evaluating a program of this magnitude and scope is difficult, as the expected outcomes are multiple. Ultimately, the success of the program will be manifested through the use of its data for policy making. Another evident milestone of project success will be the level of buy-in by individual countries, as ultimately reflected in the long-term sustainability of the survey system beyond the lifespan of the project. This goal will be achieved by fully integrating the proposed survey system into countries’ National Statistical Development Strategies and by ensuring governments’ financial commitments – even if only partial – to maintain the system beyond the duration of the project.

VI. Organizational Capacity and Project Management Design

The World Bank, and its LSMS team, is best positioned to implement such a multi-pronged strategy given its extensive experience in carrying out large national surveys in developing countries, its strong presence in the targeted countries and its long tradition in working with, and building capacity of, NSOs and relevant line ministries in developing countries in the areas of data collection and policy analysis. The LSMS team regularly provides on-demand technical assistance to developing countries in the design and implementation of household surveys, thus granting countless possibilities for creating synergies between the proposed activities and on-going government-sponsored data
collection programs. An example is the on-going collaboration between the World Bank and the Government of Tanzania in the establishment of a National Panel Survey (NPS) to be fielded starting in 2008. Similar approaches can be used in other countries, as long as similar opportunities arise.

The proposed activities will take place under the LSMS program in the Development Economics Research Group (DECRG) of the World Bank led by Martin Ravallion. The project will be managed by Calogero Carletto (Senior Economist) with full technical cooperation from the other members of the LSMS team: Kinnon Scott (LSMS Manager and Senior Economist), Kathleen Beegle (Senior Economist), Diane Steele (Data Base Manager), Kristen Himelein (consultant), Carlo Azzarri (consultant) and Talip Kilic (consultant). Overall supervision for the project will be under Peter Lanjouw, Research Manager of the Poverty Group in DECRG. The project team will also include two long-term consultants who will be hired for the duration of the project to support the day-to-day operations of the project and support the LSMS team in providing technical assistance to countries and organize the training and dissemination workshops.

The World Bank will contribute financially to the project in a number of ways. First, it will cover staff time of the project manager to contribute to the design of the project and provide technical assistance. Also, additional staff time will be provided by DECRG researchers, as well as by staff from other departments of the World Bank, who will contribute from own resources to the design and implementation of the individual country programs. The World Bank is also contributing in a number of other ways, for example by supporting the preparation of the National Statistical Development Strategies (NSDSs), the construction of meta databases and other dissemination tools through the Accelerated Data Program (ADP) and other complementary initiatives. Finally, as already done for Tanzania, the World Bank staff will support the project by leveraging financial resources and technical inputs from other donors and institutions. Although it is difficult to quantify it in advance, based on past experiences we could expect that several World Bank staff will be involved with own resources in the data collection and analysis of the data. As a way of example, we can use the on-going experience in Tanzania in
which the World Bank has contributed from own resources in one year several weeks of staff time (approximately 10) and travel expenses (approximately $20,000) to assist in the design of the National Panel Survey. Assuming a similar level of effort in the upcoming activities, we could estimate an average in-kind contribution of over $60,000/survey, or over $700,000 in the course of the project.

A Project Steering Committee (PSC) will be established and chaired by Martin Ravallion, Director of DECRG. The PSC will include a representative from the Gates Foundation, two researchers from DECRG, one representative each from the World Bank’s Africa Region and Africa Rural Development (ARD) department, and two outside experts. The PSC will meet once a year to review on-going activities and future plans, and provide guidance to the LSMS team on broad strategic issues. The PSC will have an important advocacy role for the project.

A Technical Advisory Board (TAB) will be established to advise the LSMS team on the main technical decisions, including country selection, technical review of special studies, sampling issues, etc. The TAB will meet at least once a year and will be composed by the members of the LSMS team, three World Bank staff members, and three outside experts from academia and outside research institutions.

A Technical Working Group (TWG) will be set up for each of the selected countries to contribute to the survey design and promote analytical work in the country. The TWG will be composed of one member of the LSMS Team, the country Resident Technical Adviser, at least one other DECRG researcher, one staff from the executing agency in the country (e.g., the NSO or the Ministry of Agriculture), at least two people from the local research and academic community and, when applicable, by a representative of the sub-grantee institution. The TWG will be co-chaired by the two executing institutions i.e. the government counterpart and DECRG. The possibility of inviting to the TAB the principal investigators and government counterparts of panel survey projects in the region beyond the 6 countries (e.g. Ghana, Zambia, Kenya, etc.) will be considered in order to exchange experiences and foster harmonization of methodologies.
In at least one country, the project envisions the establishment of partnerships with sub-grantees for the design and implementation of the country programs. Preliminary discussions have already been held with some of the key players in the area of data collection in Africa, including Michigan State University, the International Food Policy research Institute (IFPRI) and Yale University, and collaboration in some of the selected countries is envisioned.