DRAFT - NOT FOR QUOTATION OR CIRCULATION

# A Note on Information Systems in Mozambique

Channing Arndt, Professor
University of Copenhagen

December 2009

Updated May 2010

## 1 Introduction

Mozambique is a developing country, and a poor one at that. The most recent UN Human Development Index ranks Mozambique 172nd out of 182 countries. Indeed, Mozambique ranks in the lower half of the 24 countries defined by the UN as having low levels of human development. A defining feature of poor countries is institutional weakness. Institutional weaknesses almost always extend into information systems. Poor countries as a rule encounter difficulties collecting, analyzing and disseminating adequate information as a base for decision-making.

This fact needs to be borne firmly in mind. Compared to nearly all countries characterized by the UN as having low levels of human development and many countries characterized as having medium levels of human development, Mozambican information systems fair reasonably well.[1] The National Institute of Statistics (INE) can, with justifiable pride, point to a series of achievements including two censuses, numerous surveys, and serious national accounts. Most recently, the Mixed Indicator Cluster Survey (MICS) has provided valuable information and is considered to be of high quality.

The purpose of this note is to provide realistic, concrete and constructive ideas for making information systems in Mozambique better. As the preceding paragraph indicates, it is not a criticism of INE. Indeed, the gradual accretion of institutional capability within INE makes this note possible. At the same time, it is also fair to say that all is not well. Mozambique is in the process of creating plans for the next five years on an undesirably thin information base. Official agricultural statistics were unreliable for an extended period (this has been addressed though the current source for statistics has difficulties with timeliness), enterprise monitoring is weak, and households survey information is largely (but not completely) outdated. This constellation of shortcomings has unavoidably negative implications for national accounts. Overall, the relative information vacuum makes it difficult to plan and conduct productive policy debate, particularly with respect to development strategy.

[1] This includes countries like Ethiopia that are classified as having low levels of development and receive large volumes of foreign assistance.

The following section considers issues of information systems in four major areas: agricultural statistics, enterprise surveys, household surveys, and national accounts. Each subsection provides a diagnosis of issues and concrete suggestions. The final section of this note summarizes and concludes.

## 2 Information Systems

### *2.1 Agricultural Statistics*

While the lack of timely information on poverty and wellbeing is, at the moment, the most discussed aspect of the information system, for an extended period the most glaring overall weakness in the entire system probably revolved around agricultural statistics. There are two sources for agricultural statistics. On one end, there is a database of statistical information obtained through sample surveys called TIA (*Trabalho de Inquérito Agrícola*) On the other end, forecasts of crop production, estimated using a different methodology, are generated regularly through the Early Warning System *(Aviso Prévio)*. These sources show huge discrepancies in their estimates and trends. For example, in 2005, cereal production was estimated at 1,900,000 tons by *Aviso Prévio* compared to 1,137,000 tons by TIA, a 67% difference. This is a disconcertingly large range for what is one of the most important annual statistics in the entire country.

An evaluation of the two systems conducted in 2007 makes it clear that the TIA produces more reliable information. The final report states that TIA "uses a universally accepted methodology" while the Early Warning System suffers from "irregular" data collection due to a lack of funding and trained staff.[2] In addition, "the documentation of the [Early Warning] system does not render it sufficiently transparent for users to have full confidence in the data." Despite these misgivings, which were well known, the Early Warning system remained the official source of information until only recently.[3]

INE has taken a substantial step towards improving official agricultural statistics by switching to TIA as the official source for agricultural production information. Nevertheless, a great deal remains to be done for at least three reasons. First, TIA is insufficiently timely and highly unlikely to become so. Second, the TIA structurally cannot provide early warning of low levels of production and hence elevated food insecurity. Third, the TIA, while much better than the Early Warning System as currently practiced, has quality problems of its own. In particular, the same review concludes that "the quality of [TIA] data is compromised by the length of the questionnaire and the dependency on recall."[4] This dependence on recall is necessary for

[2] Discussion paper 64E available at http://www.mpd.gov.mz/gest/publicat.htm.
[3] It merits mentioning that, in 2005, the TIA and other corollary data point to a strong drought in that year. Therefore, the best evidence indicates that the Early Warning System completely missed a strong drought.
[4]To date, the TIA has been dogged by a perplexing failure of dissemination. Commitment to the TIA both within and outside of the Ministry of Agriculture has been weakened because so little public information has actually emerged. With the elevation of TIA to the status of official data source, significant efforts are now in process to properly publish and widely disseminate TIA data and principal.

determination of rural income; however, it is not necessary for determining levels of agricultural production on an annual basis.

The evaluation report provides useful technical suggestions for the way forward. In particular, the Early Warning System and the TIA should be combined so that the best elements of both systems can be exploited. TIA operates with a viable sample frame that could be exploited both for the purposes of early warning and estimation of final crop production. For these purposes, crop cuttings, as practiced by the Early Warning System, are likely to be superior than recall. In collecting crop cuttings, there is ample scope for exploiting existing technologies that permit enumerators to locate predetermined areas for taking crop cuttings with very high accuracy. At the same time, the TIA, as currently practiced, is probably the best available way to estimate rural income. Accurate estimates of crop production via crop cuttings taken from farms within the same sample frame would strengthen the TIA as a vehicle for estimating rural income. It is not necessary to estimate rural household income every year, but the TIA should be a part of a regular household survey program, a point to which I will return.

### *2.2 Enterprise Surveys*

There have been a number of attempts to establish systems for regular monitoring of private sector activities, constraints, and performance. Current efforts are spread across a variety of actors including the Statistics Institute (INE), the Confederation of Business Associations (CTA), the Ministry of Commerce and Industry (MCI), the Ministry of Planning and Development (MPD) and work by the World Bank designed to feed their "Doing Business" reports and the Regional Program for Enterprise Development (RPED) as well as to evaluate specific programs such as the infrastructure corridors. These initiatives have recently produced a relative abundance of survey data, which has been the subject of useful analysis over the course of 2008-09.

Nevertheless, the programs remain diffuse and ad hoc. Frequently, larger businesses are surveyed multiple times with over-lapping questionnaires. Survey fatigue combined with a lack of dissemination leads to low response rates and a lack of public information. This lack of information stimulates ad hoc efforts to collect information from the private sector which increases survey fatigue. In short, a vicious circle of survey fatigue, low response, lack of information, and ad hoc attempts to collect more information has been in place for some time. Instead of the effectively piecemeal system currently in place, a convincing, reliable, efficient and sustainable program for monitoring and evaluation of private sector performance is required.

This sounds daunting, but the outlines of such a program are reasonably clear. It involves:

1. Granting INE and MPD access to tax data from firms as reported to the Tax Authority (ATM) on an annual basis. In many countries, this data replaces the need for conducting financial surveys of large formal sector firms. At the moment, INE surveys large firms in the weeks immediately following the due date of tax forms using a questionnaire that purposefully mimics the tax form in order to encourage response. Despite these efforts, response rates are low. Access to tax data by INE

would simultaneously increase information while allowing INE to focus scarce data collection resources elsewhere.

2. Business environment and perception surveys. These surveys only need to be carried out once every few years. The World Bank has been carrying out these surveys on a global basis on an approximate three year time frame. Now is the time to ensure that the World Bank program feeds and supports a national enterprise monitoring program.

3. Regular (annual or maybe every other year) monitoring of the informal sector and small and medium enterprises whose reports to the tax authority are either non-existent or lacking in detail.

4. Links with ongoing monitoring efforts undertaken by the Ministry of Commerce and Industry and the CTA.

5. Commitments from donors involved in private sector development to support the national monitoring program instead of developing and financing a series of one-off monitoring exercises.

As indicated above, enterprise monitoring involves a number of players. INE has the role of the ultimate responsible entity. However, in the initial organizing phase, the coordinating role of MPD should be capitalized upon with DNEAP playing a leading role. It would be ideal to include an intelligent enterprise monitoring program into the monitoring and evaluation component of the new five year plan.

### *2.3 Household Surveys and Poverty Monitoring*

With respect to household surveys and poverty monitoring, the principal problem is frequency and timeliness. Since 1994, a reasonably large number of household surveys have been undertaken. These surveys are widely viewed as being of acceptable quality (with ample room for improvement), and they have been analyzed in considerable detail. For each survey, at least one basic descriptive report has been produced. For most of the surveys, a large number of more detailed analyses have been conducted focusing on specific issues. The descriptive reports and more detailed analyses are actively disseminated and widely read. The problem is that these efforts, while relatively good on the African continent, do not meet demand. Timeliness is also a problem. Within about six months from now, it is reasonable to expect that results from MICS, IOF, TIA, and the 2007 Census will have been analyzed, published, compared and publicly debated. Unfortunately, that information base will be available more than six months after the new five year plan (PQG) goes to Parliament. The current information base, meaning the one on which the PQG must be based, is thin.

And, we have been here before. In terms of information availability, the current situation is remarkably similar to the situation that prevailed in 2002 and into early 2003. At that time, Mozambique was six years from the conclusion of field work for the household budget (IAF)

and Demographic and Health Surveys (DHS). A Core Welfare Indicators Survey (QUIBB) had been conducted in 2000. National accounts showed strong GDP growth from 1996-2001 including strong growth in agriculture. The situation in terms of survey information available in mid 2009 was in large measure the same with the dates changed. In particular, about six years had passed since the most recent IAF and DHS surveys. The latest QUIBB had been conducted four years (as opposed to two years) before, but this was compensated by a complete agricultural income survey (TIA) in 2005. In addition, national accounts point in the same direction. In particular, published growth rates in GDP for the period 2003-2008 are quite comparable with the corresponding 1996-2001 period.

As in 2002, there is massive over-reliance on the next major poverty oriented survey to appear, which will be the household budget survey (renamed IOF), for the determination of poverty evolution. This heavy reliance on a household budget survey conducted once every six years is problematic for a number of reasons including:

1) Poverty is a multi-dimensional phenomenon. The household budget surveys examine some, but certainly not all, important dimensions of poverty.

2) Focusing on the consumption dimension, the available evidence (especially TIA 2005 which contains a panel dimension) indicates that true consumption poverty rates are fairly volatile through time especially when disaggregated (e.g., provincial consumption poverty rates). Further, we attempt to observe this fairly volatile number through a lens that is distorted by sample error (which we can estimate formally) and non-sample error (which cannot be estimated formally). While the national level indicators are likely to be somewhat more stable, the potential for drawing inappropriate conclusions at the national level remains. At the sub-national levels, such as the provinces, where the degree of underlying variability is higher and our confidence in the measure is lower, this potential is large. This is particularly true when the household budget survey is only carried out once every six years.

3) Strong shocks may make comparison difficult. For example, the fieldwork period for IOF 2008-09 was characterized by a combination of drought and flooding that reduced food production. At the same time, world prices for food had risen to levels not seen for decades. Government and its international partners will have to make decisions on the basis of a single and potentially high very weak information base for at least the time it takes to organize, launch, and analyze another survey (an additional two years at least).

4) A six year interval is too long to develop a viable panel dimension (e.g., interviews with the same households at two or more points in time). Without a panel dimension, our ability to understand vulnerability, chronic poverty and transitory poverty is severely compromised. The available evidence, particularly TIA 2005, suggests that these dimensions are very important for policy formulation.

As highlighted in the introduction, the monitoring program in Mozambique is fairly active by low income country standards. The critiques cited above could apply to many other countries. Why should Mozambique have a particularly active program? There are three major reasons.

First, Mozambique has a far more dynamic economy than the average. Current information is required to monitor and adapt to ongoing dynamic changes. Second, Mozambique obtains much greater levels of foreign assistance than the average. This assistance is targeted at poverty reduction. Mozambique's international partners clearly have strong demand for information on poverty evolution. Third, Mozambique does not aspire to be average. An active monitoring program is required for doing well over the long run.

In sum, while surveys are costly, so is ignorance. In terms of household surveys, the program should strive for:

a) Increased frequency of collection of core poverty analysis data. I favor, as the core program, a three year rotation with the following surveys:

   - Year T: IOF.
   - Year T+1: DHS or similar (such as MICS).[5]
   - Year T+2: Full TIA in rural areas and a labor force survey in urban areas.
   - Repeat.

   Under this plan, significant dimensions of poverty will be investigated every year. The program allows for panel dimensions to be developed across and among the various surveys. Finally, after three rotations (nine years), the household survey program would be delayed by one year in order to complete the next population census.

   The devils are in the details and it is not entirely straightforward to operationalize this core program. A specific suggestion of a program extending out 20 years that follows this basic model is included in the Appendix. An important element of any plan is to bring forward the next census from 2017 to 2016. This allows sufficient time for the census to be conducted and analyzed prior to the 2019 election and subsequent PQG. This is necessary because the census is the basis for the sample frame for all subsequent surveys. So, if there is to be a household survey conducted prior to the 2019 election using a reasonably up to date sample frame, then the census must be conducted beforehand. If the census is conducted in 2017, then the timeliness problems that we are experiencing now are highly likely to recur in 2019.

b) Enhanced availability of raw data appropriately anonymized. Raw data for the DHS surveys are already made available following an explicit procedure. A procedure for making available data from other surveys to researchers both within and outside of Mozambique should be developed.

c) Full reproduceability of key results ideally from the publicly available raw data. The ability to reproduce results from raw data is a key tenet of science. Interested researchers should be able to request the code that generates published statistics from the raw data.

Had such a program been implemented following IAF 2002-03 and DHS 2003, we would be in a vastly better informed position than we are now. There are good reasons why this was not done.

[5] There is no reason to run both DHS and MICS surveys.

In particular, the massive demands of the 2007 Census combined with the institutional capacity of INE supported arguments for a more limited program. Nevertheless, at this stage, the core program discussed above fits well within the capacities of INE *if properly supported and budgeted*. It is time to put in place a more active poverty monitoring program.

### *2.4 National Accounts*

We all quote statistics on GDP growth and other statistics emanating from national accounts. These statistics are, without doubt, important. But, national accounts can only be as accurate as the available database permits. National accountants in Mozambique have had to contend with agricultural statistics from two sources that differ drastically in levels as well as trends, low response rates on enterprise surveys, and low periodicity of relevant household surveys (among other difficulties). Under the circumstances, they cope very well.

But, National Accounts can do much more than cope. Because they deal with the whole economy, national accountants are armed with a variety of accounting identities that require reconciliation across the full set of available data sources. For example, maize production and imports (supply) have to equal use (demand) where stock accumulation is a component of use.[6] Reconciliation of information across data sources within a coherent framework is a powerful tool, perhaps the most powerful tool available, for diagnosing problems in the statistical system. And, reconciliation of information across data sources happens in national accounts as a matter of course.

Often national accounts are regarded as essentially the last step in the statistical system where the available information is combined to produce a coherent picture of the economy. However, it should also be viewed as the first step where existing weaknesses, and the most likely sources of those weaknesses, are identified. National accountants should be engaged in these two steps, both of which are complex and idiosyncratic. National accounts departments should be budgeted, staffed, and technically supported accordingly.

## 3 Conclusions and Implications

In summary, the main recommendations proposed here can be summarized in a few points.

1) Build on existing momentum to improve annual agricultural production statistics.

2) Develop and implement over the next few years a functional plan for enterprise monitoring.

3) Build on existing strengths in household survey data collection and analysis to develop a system that provides a more consistent flow of information on poverty and living standards including panel dimensions. The program in the Appendix provides a potential model.

[6] Note that this is pure accounting. It merely asserts that maize neither magically appears nor magically disappears.

4) Elevate the profile and role of national accounts.

These recommendations will require somewhat greater allocations of budgetary resources to central data collection and analysis entities. However, the marginal budget allocations are not large, particularly when compared to the costs of operating with insufficient information (as we are now).

At least as relevant, a credible and reliable central data collection program along the lines described will reduce demand for the numerous duplicative and ad hoc data collection efforts that are underway. Most obviously, one credible system for agricultural statistics is better than two conflicting systems and not likely to cost very much more in total. Other, potentially less obvious, examples abound. For instance, the SETSAN has conducted since 2004 at least two national or nearly national level surveys on various nutrition and child anthropometric indicators. Reports from these surveys describe the data quality from these surveys as low, and dissemination of results has been less careful than would be desirable. For example, in 2007, the front page headline in *Notícias* practically screamed that child malnutrition was increasing based on results from a SETSAN survey published on the SETSAN website. However, a careful analysis found no evidence that child malnutrition was increasing. Rather, the point estimates indicated a decline in child malnutrition though data quality concerns prevented drawing robust conclusions. In short, substantial misinformation was achieved at high cost.[7] As a second example, the World Bank has funded an effort to develop a panel of households from IAF 2002-03 in a few provinces. This effort was expensive and has yielded, as yet, no published information. Finally, the surfeit of ad hoc enterprise surveys and associated survey fatigue has already been mentioned.

In sum, while allocations to central data collection programs will go up under the program described, it is reasonable to hope that total allocations to data collection overall will increase rather little (or even decline) due to a reduction in duplicate and/or ad hoc surveys. At the same time, information quantity, timeliness, and quality should be expected to increase markedly.

[7] I pick on the SETSAN mainly because I know the case fairly well not because they are in any way special. Fundamentally, the point is not to criticize SETSAN. The point is that SETSAN has plenty to do aside from organizing, conducting and analyzing national level surveys of health and nutrition. A more regular DHS or MICS survey, with both the report and the basic data made available, would be a boon for the SETSAN and would allow them to focus on their specific role.

## Appendix: A Suggested Household Survey Program

| | | Survey | | |
|---|---|---|---|---|
| Event | Year | In field | Out of field | Analysis complete |
| | 2010 | TIA | TIA | |
| | 2011 | LF/IOF | LF | TIA |
| | 2012 | MICS | IOF/MICs | LF |
| | 2013 | TIA | TIA | IOF/MICs |
| PQG | 2014 | LF/IOF | LF | TIA |
| | 2015 | MICS | IOF/MICs | LF |
| Census | 2016 | Census | Census | IOF/MICs |
| | 2017 | IOF | | Census |
| | 2018 | TIA | IOF/TIA | |
| PQG | 2019 | LF/MICS | LF/MICs | IOF/TIA |
| | 2020 | IOF | | LF/MICs |
| | 2021 | TIA | IOF/TIA | |
| | 2022 | LF/MICS | LF/MICs | IOF/TIA |
| | 2023 | IOF | | LF/MICs |
| PQG | 2024 | TIA | IOF/TIA | |
| | 2025 | LF/MICS | LF/MICs | IOF/TIA |
| Census | 2026 | Census | Census | LF/MICs |
| | 2027 | IOF | | Census |
| | 2028 | TIA | IOF/TIA | |
| PQG | 2029 | LF/MICS | LF/MICs | IOF/TIA |

NOTES: IOF is the household budget survey, TIA is the rural income survey, MICS (or DHS) is the demographic and health survey, and LF is the urban labor force survey. IOF is in the field for one year from July to June. TIA is typically in the field from September to November. If MICS were conducted from September to November (roughly) and LF from March to May, then the program above would never have two surveys in the field simultaneously.

The purpose of the above table is to suggest a program rather than the program. Modifications are certainly possible.