



IMPROVING THE METHODOLOGY FOR USING ADMINISTRATIVE DATA IN AN AGRICULTURAL STATISTICS SYSTEM

*Reviewing the Relevant
Literature and Studies on the
Quality and Use of
Administrative Sources for
Agricultural Data*

August 2015

Working Paper No. 1

Global Strategy Working Papers

Global Strategy Working Papers present intermediary research outputs (e.g. literature reviews, gap analyses etc.) that contribute to the development of Technical Reports.

Technical Reports may contain high-level technical content and consolidate intermediary research products. They are reviewed by the Scientific Advisory Committee (SAC) and by peers prior to publication.

As the review process of Technical Reports may take several months, Working Papers are intended to share research results that are in high demand and should be made available at an earlier date and stage. They are reviewed by the Global Office and may undergo additional peer review before or during dedicated expert meetings.

The opinions expressed and the arguments employed herein do not necessarily reflect the official views of Global Strategy, but represent the author's view at this intermediate stage. The publication of this document has been authorized by the Global Office. Comments are welcome and may be sent to gsars@fao.org.

Improving the Methodology for Using Administrative Data in
an Agricultural Statistics System

Working Paper 1:
Reviewing the Relevant
Literature and Studies on
the Quality and Use of
Administrative Sources for
Agricultural Data

By
School of Statistics and Planning
College of Business and Management Sciences
Makerere University
Uganda

and

The Center for Survey Statistics and Methodology
Iowa State University
United States of America

Executive Summary

The Global Strategy to Improve Agriculture and Rural Statistics was adopted by the United Nations Statistical Commission in 2010 with the objective to improve statistics in agriculture, livestock, aquaculture, small-scale fisheries and forestry production in developing countries and ensure that they are maintained over time. One of the key components of the Global Action Plan is its research plan, which has as one of its priorities “improving the methodology for using administrative data in agricultural statistics” (World Bank, FAO and United Nations 2010).

The research involves eight tasks, outlined in sub-section 1.1.2. The present report is for Task 1, which essentially entails conducting a review of relevant literature and studies on the quality and use of administrative sources for producing agricultural data and proposing a conceptual framework for using administrative data in agricultural statistics.

The working definition of administrative data for this research is “information collected primarily for administrative (not statistical) purposes by governmental departments and other organizations usually during the delivery of a service or for the purposes of registration, record keeping or documentation of a transaction”.¹

The benefits of the administrative data identified in the review are that they: tend to be cheaper to use; reduce the response burden and also improve public image; can be collected more frequently, enabling them to be more timely and up-to-date; and can cover the whole target population. In developing countries, administrative data are sometimes the only data available, especially for lower-level administrative units.

Sources of administrative data applicable to agricultural statistics identified in the review included: regular returns/reports by agricultural field/extension staff (for various agricultural items, including crops and livestock); tax data; land ownership records; information on government subsidies; import/export data; agricultural production and inputs manufacturers and distributors; farm registers

¹ See www.adls.ac.uk/adls-resources/guidance/introduction/.

and other registration or licensing systems; records on agro-tourism; farmers' associations; private businesses data; meteorological data; and traceability data, such as traceability livestock data.

Administrative data can be used in all countries in the following areas: sampling frame construction and sample design; administrative records used to cover data gaps from surveys and censuses; forecasting; planning; and provision of small area estimates and administrative uses, including informing policy and decision-making.

The research showed that while developed countries tended to use scientific approaches for adjusting or improving administrative data before use, developing countries were more likely to use simpler subjective methods, such as expert opinion, screen surveys and eye estimation. The scientific approaches in developed countries included use of integrating factors in which the linkages established involved: (1) incorporating a common set of core items; (2) consistent application of common and compatible concepts, definitions and classifications; and (3) possibly use of a common field organization. As this framework should provide continuity, it will facilitate the accumulation of experiences, which can then be referred to enhance the quality of data in subsequent rounds of data collection.

The use of administrative data for agricultural statistics in developing countries is quite high, especially for cash/commercial crops, crop forecasting/early warning, livestock and poultry, inputs and trade data, but is limited for estimating production of food crops, resulting in serious implications for estimates of food security. Currently, most agricultural administrative data are produced by public/government institutions. The private sector, farmers' and traders' organizations and research organizations could be tapped to also produce such data. In developing countries, the agricultural administrative data are mostly used for decision-making, while in developed countries; they are mostly used to improve statistics. Much of the administrative data is collected and compiled without employing standard statistical procedures or personnel with training in statistical methods. A good portion of the data from this source remains in raw form and are not turned into usable information. Apart from exports and imports and agricultural price information, which are often published widely, agricultural administrative data are not widely disseminated for use. In addition, there is untimely and incomplete information flow from the lower reporting levels.

In developing countries, documentation of administrative data collection and processing methodologies are also limited. Assessments of the quality of administrative data, which are carried out subjectively in most developing countries, tend not to be detailed enough to cover the different quality dimensions. Agricultural statistics can be strengthened in developing countries by using new geospatial tools, such as satellite imagery and the Global Positioning System (GPS). As a result of decentralization at the lower levels, there is delayed information flow, limited capacity, and data are only collected on an intermittent basis. This subsequently constrains programme monitoring by the local governments because of inadequate understanding of their own data. Operational constraints make it difficult for extension staff to go around their areas of jurisdiction to collect data. They have several other responsibilities and are not normally legally under the statistics authorities. Agricultural returns written by agricultural extension staff are based on data collection forms that lack standardization and can lead to reporting errors and inconsistencies.

In general, suggestions for improving usability are centred on improving cross checks and corrections made through survey data. An efficient and cost effective agricultural statistics system would involve the use of limited surveys based on strong statistical principles to provide validation and corrective factors for the data generated on a regular basis through administrative accounts (sample audit system). The main factors governing the usability of administrative data for statistical purposes are the legal framework underpinning the data collection activity and the political importance attached to the government interventions concerned and governance.

Limitations of administrative data in developing countries include: changes in administrative processes, which result in inconsistencies in estimates across time and reduced data availability; lack of standardized definitions of variables, units and identifiers, which can make synthesis of multiple sources a challenge; under-coverage as, in some cases, not all members of the target population participate in the administrative process; reporting errors stemming from incentives underlying administrative processes; computational demands as use of administrative data often involves processing and cleaning massive volumes of data; and the challenge of maintaining confidentiality even though the use of administrative data may involve confidential information. A solution going forward would be to review and revise the legal framework to cover administrative sources.

Table of Contents

Executive Summary	i
Acronyms and Abbreviations	vii
1 Description of the Work and Methodology	1
1.1. Introduction.....	1
1.1.1. Background.....	1
1.1.2 Overview of objectives and approaches	1
1.2. Administrative data.....	4
1.2.1. Definition.....	4
1.2.2. Administrative data sources.....	5
1.2.3. What is a register?.....	6
1.3. Benefits of using administrative data in the production of statistics	8
1.3.1. Coping with budget cuts in national statistical offices budget cuts... ..	8
1.3.2. Roles of administrative data in sampling and estimation	8
1.4. Common problems with administrative data and solutions.....	11
2 Proposed framework for improved systems of administrative agricultural statistics.....	21
2.1. Analytical framework adapted from the agricultural market information systems in sub-saharan africa.....	21
2.1.1. Structural issues	22
2.1.2. Conduct issues	23
2.1.3. Performance	24
2.2 The health matrix network framework.	24
2.3. Complementary arrangement for an administrative data system, surveys and censuses.....	28
2.4. Review of the quality of administrative data	28
2.5. Core data items to be reviewed.....	31
2.6. Review of data use	33
3 Specific cases of using administrative data.....	34
3.1. Developed countries.....	34
3.1.1 Administrative data in an integrated national statistical system.....	35

3.1.2. Uses of administrative data in developed countries.....	37
3.1.2.1. Direct tabulation of statistical registers.....	37
3.1.2.2. Reduction of respondent burden and costs of data collection..	38
3.1.2.3. Improving efficiency of survey based estimators	38
3.1.2.4. Improving frame construction and sampling design	39
3.1.3. Methodological issues with administrative data	40
3.1.3.1. Quality control: measurement and coverage error in developed countries	40
3.1.3.2. Record linkage from multiple administrative sources	42
3.1.3.3. Software for probabilistic record linkage in developed countries	42
3.1.4. Discussion	44
3.2. Developing countries	44
3.2.1. Background.....	45
3.2.2. Main uses of administrative data for statistical purposes in developing countries	46
3.2.2.1 Administrative Data for Sampling Frame Construction and Sample Design	46
3.2.2.2. Crop Forecasting	47
3.2.2.3 Integration of Administrative Data in Agricultural Surveys.....	48
3.2.2.4 Administrative Data for Planning	48
3.2.2.5. Small Area Estimates	49
3.2.2.6. Non-probabilistic Procedures of Integrating Use of Administrative Data Sources.....	49
4 Issues in the Use of Administrative Data in Developing Countries.....	53
4.1. Background.....	53
4.2 Methodology for obtaining the data.....	54
4.3 Decentralization and levels of information management	55
4.4 Operational constraints	55
4.5 Difference between administrative purposes of the data and statistical requirements.....	56
4.6. Crop forecasts	57
4.7. Livestock.....	58
4.8. Import and export.....	58
4.9. Fisheries	59

4.10. Improving usability of administrative data 59

4.11. Linkages as integrating factors 62

4.12. Potential sources of administrative data 63

5 Conclusions..... 64

References 66

Appendix I 79

Appendix II..... 80

Acronyms and Abbreviations

ADLS	Administrative Data Liaison Service
ADSAS	Administrative Data Systems for Agricultural Statistics
AfDB	African Development Bank
CSO	central statistics office
ECE	Economic Commission for Europe
ESS	European Social Survey
FAO	Food and Agriculture Organization of the United Nations
GDP	gross domestic product
GPS	Global Positioning System
IACS	Integrated Administrative and Control System
ICAS	International Conference on Agricultural Statistics
ICT	information and communications technology
INE	<i>Instituto Nacional de Estadística</i>
ISAD	Statistical Methodology Project on Integration of Survey and Administrative Data
ISTAT	Italian Statistical Institute
LGMD	Local Government Management Database
MDG	Millennium Development Goals
MINAG	<i>Ministério de Agricultura</i>
NGO	non-governmental organization
NSDS	National Strategy for the Development of Statistics
NSI	national statistical institute
NSO	national statistics office
PARIS21	Partnership in Statistics for Development in the 21st Century
PDA	personal digital assistants
PRSP	poverty reduction strategy paper
SADC	Southern Africa Development Community
SAP	<i>Système d'Alerte Précoce</i> /Early Warning system
SMS	short message service
SWOT	strengths, weaknesses, opportunities and threats
UBOS	Uganda Bureau of Statistics
USAID	United States Agency for International Development

Description of the Work and Methodology

1.1. INTRODUCTION

1.1.1. BACKGROUND

The Global Strategy to Improve Agriculture and Rural Statistics was adopted by the United Nations Statistical Commission in 2010 with the objective to improve statistics in agriculture, livestock, aquaculture, small-scale fisheries and forestry production in developing countries and ensure that they are maintained over time. Its main goal is to build statistical capacity in developing countries for key basic food and agricultural statistics (World Bank, FAO and United Nations 2010).

One of the key components of the Global Action Plan is its Research Plan, which is aimed at developing cost-effective methods to serve as the basis for preparing technical guidelines, handbooks and training material to be used by consultants, country statisticians and training centres. One of the key priorities of the Research Plan, which was to be implemented in 2014, is to improve the methodology for using administrative data in agricultural statistics (World Bank, FAO and United Nations 2010).

1.1.2 OVERVIEW OF OBJECTIVES AND APPROACHES

The aim of the research is to develop strategies and methodologies for improving the collection and management of data from administrative sources and of their use in an integrated agricultural statistics system in developing countries. This entails investigating cost-effective approaches and methods for the production of regular and geographically disaggregated reliable agricultural data, including the combination of surveys and administrative data. The expected primary products of this research will be (a) a technical report that

includes a country-tested and validated methodology to improve and make available administrative data for producing agricultural statistics in developing countries and (b) a proposed strategy on how to use administrative data in cost-effective agricultural statistics systems. The technical report and proposed strategies will develop sound methodology for improving and using administrative sources for agricultural statistics in developing countries, taking into consideration the existing approaches regarding administrative information systems in the different countries, such as differences in approaches for collecting and using administrative data between on one side, French-Portuguese-Spanish-speaking countries and on the other side, English-speaking countries. New potential sources of administrative data will also be examined.

More specifically, we propose a four-step approach.

(a) A thorough review of the quality and use of administrative data to improve agricultural statistics in developed countries. This review will enable us to formulate recommendations for improving the quality of the sources of administrative data, for example, the quality of the registers from which the data are derived or the methodology and institutional framework for obtaining the data. One mechanism for verifying the quality of administrative data used in developed countries is to conduct a survey to check the administrative data (Carfagna and Carfagna 2010). In addition, combining multiple sources of administrative data in a common register system can increase coverage and improve the consistency of estimates based on different sources (Wallgren and Wallgren 2010). The review will include the following four areas:

- Direct tabulation of statistical registers;
- The quality and use of administrative records to reduce cost of data collection;
- The quality and use of administrative records to improve estimation;
- The quality and use of administrative records for frame construction and sampling design.

(b) An analysis of the country assessment surveys and other documentation to identify methodological issues in using administrative data in developing countries, with a particular emphasis on record linkage and data quality control, including controls for both measurement errors and coverage errors.

(c) Based on the findings in (a) and (b) and experiences in developed countries, such as the work of the Center for Survey Statistics and Methodology of Iowa State University with the National Agriculture Statistic Service of the United

States Department of Agriculture, developing a general methodology for integrating administrative data and other sources of auxiliary information with survey data as part of an integrated national statistical system.

(d) Work with the Food and Agriculture Organization of the United Nations (FAO) to select two developing countries for in-country testing to validate and improve the methodology developed in (c), and produce final guidelines for developing countries to integrate administrative data into agricultural statistics.

The research under the project is to be carried out in eight tasks. Task 1 involves a review of relevant literature and studies on the quality and use of administrative sources for producing agricultural data. In Task 1, a conceptual framework is proposed for assessing the quality and use of administrative data in agricultural statistics. The examination of the quality of administrative sources involves a review of the data systems in place and a review of the data that are collected, stored and disseminated for use by the national statistics systems. The review begins with literature available from FAO, the United Nations Statistics Division and developed countries (especially Scandinavian countries, the United States of America, Canada and Australia) followed by reviews of literature from developing countries and all other available sources. The review also identifies systems with potential for providing data, which can then be enhanced for use. The paper concludes with a general summary of issues related to the role of administrative data in the production of official statistics.

Task 2 involves a review of country experiences and ongoing research on the role of administrative data in developed countries. Task 3 entails a review of current practices in the use of administrative data sources in an agricultural statistics system in developing countries. Task 4 involves identification and an analysis of gaps and ways to improve the quality and use of administrative data sources for agricultural statistics. Task 5 involves drafting a strategy and methodology for a cost-effective agricultural statistics system combining sample surveys and administrative data sources. A pilot to test the proposed system will be conducted in two developing countries as Task 6. The system, strategy and methodology will be revised following the pilot as Task 7 and finally, Task 8 is the compilation of a methodological report.

1.2. ADMINISTRATIVE DATA

The discussion in this section is centred on the definition of administrative data and sources, especially sources of agricultural administrative data.

1.2.1. DEFINITION

Several authors have traditionally defined administrative data sources as collections of data held by other parts of government, collected and used for purposes of administering taxes, benefits or services. Perhaps, the most comprehensive of the traditional definitions was set out by Gordon Brackstone of Statistics Canada (Brackstone 1987). In his 1987 paper, he identified the following four distinguishing features of administrative data:

- a) The agent that supplies the data to the statistical agency and the unit to which the data relates to are different (in contrast to most statistical surveys);
- b) The data were originally collected for a definite non-statistical purpose that might affect the treatment of the source unit;
- c) Complete coverage of the target population is the objective;
- d) Control of the methods by which the administrative data are collected and processed rests with the administrative agency.

United Nations (2011) reports on an internal Eurostat task force which in 1996-1997 examined ways to better coordinate work relating to the use of administrative sources across different domains of statistics. It explained that the task force used a simple typology of data sources to consider how administrative sources should be defined. First, all data sources were divided into primary sources (data collected for statistical purposes) and secondary sources (all other data). A traditional or “narrow” definition of administrative sources comprises just public sector non-statistical sources, whereas a wider definition would also include private sector sources. Thus, under the narrow definition, administrative sources are a subset of secondary sources, while under the wider definition these terms are synonyms. There is a growing number of reasons for favouring the wider definition, including, among them: (a) increasing privatization of government functions; (b) register-based population statistics (United Nations 2011), growth of private sector data and “value-added resellers”; and (c) user interest in new types of data (Brackstone 1987).

Pronab (undated) provides an alternative way of defining administrative data as being distinct from statistical data when the specific identity of the respondent or data source is central to the use of the data. Statistical data, on the other hand, would be those in which such identity is in itself of no consequence. In other words, the distinction between statistical and administrative data is based on the focus on the general as against the particular.

The definition used in the United Nations Statistical Handbook (United Nations 2011) is:

“Administrative sources are data holdings containing information that is not primarily collected for statistical purposes.”

The Administrative Data Liaison Service of the United Kingdom of Great Britain and Northern Ireland defines administrative data as “information collected primarily for administrative (not research) purposes. This type of data is collected by government departments and other organizations for the purposes of registration, transaction and record keeping, usually during the delivery of a service”.

For the purpose of this research, administrative data is defined as “information collected primarily for administrative (not statistical) purposes by government departments and other organizations usually during the delivery of a service or for the purposes of registration, record keeping or documentation of a transaction.

1.2.2. ADMINISTRATIVE DATA SOURCES

As implied in the previous paragraphs, the potential range of administrative sources that can be used for statistical purposes is large and growing. Many sources of administrative data can be used to produce agricultural statistics, in particular. These include the regular returns or reports by agricultural field/extension staff (for various agricultural items, including crops and livestock), tax data, land ownership records, information on government subsidies, import/export data, associations of agricultural producers and inputs manufacturers and distributors, administrative farm registers and other registration or licensing systems, records on agro-tourism, farmers' associations, private businesses data, meteorological data and traceability data. In the case of agriculture, countries are increasing the “traceability” of products as a way to monitor quality. An example is traceability of livestock, which is widely used because of requirements of importing countries. For livestock, traceability allows having precise information for each bovine in a country

herd, for example, in Uruguay, 12 million bovines are now traceable (Presidencia undated). Traceability processes generate large volumes of data that are accurate, updated, complete and reliable and can easily be accessed from a database. This kind of data is reliable and can easily be a substitute for or complement data from sample surveys or censuses.

Other General Sources

In addition to the sources specifically for agricultural data, there are several other sources of administrative data. Examples are given in the following list provided by United Nations (2011). The list is not meant to be exhaustive, but instead aims to show the range and types of potential data sources as the final step towards arriving at an operational definition of administrative sources.

- Tax data
 - Personal income tax; value added tax (VAT); business/profits tax; property taxes; and import/export duties
- Social security data
 - Contributions; benefits and pensions
- Health/education records
- Registration systems for persons/businesses/property/vehicles
- Identity cards /passports/driving licenses
- Electoral registers
- Local council registers
- Building permits
- Licensing systems, e.g. television, sale of restricted goods
- Published business accounts
- Internal accounting data held by businesses
- Private businesses with data holdings:
 - Credit agencies; business analysts; utility companies; telephone directories and retailers with store cards

It, however, appears that developed countries rarely use agricultural field/extension staff regular returns or reports as a source of data; instead, they rely on registers.

1.2.3. WHAT IS A REGISTER?

Typically, a register is a structured list of units that contains many attributes for each of those units and has a regular updating mechanism. It can be obtained from a census, a survey or administrative records. Many administrative data files can be considered to be registers, but the results of one-off data collections

are not. The agricultural census is sometimes used as a first step to build a farm register. In such cases, the directory from the census of agriculture with all the auxiliary information is a farm register. This should be differentiated from a register built from purely administrative sources.

A register is defined in “Register-based statistics in the Nordic countries” as a systematic collection of unit-level data organized in such a way that updating is possible (Nordbotten 2008). Updating is the processing of compiling identifiable information with the purpose of establishing, bringing up to date, correcting or extending the register, namely keeping track of changes in the data describing the units and their attributes. As a rule, a register contains information on a complete group of units, a target population, such as persons, buildings and firms. These units are defined by a precise set of rules (for instance, resident population in a country), and the attributes are updated in line with changes undergone by the units. The definition of registers given here refers exclusively to the content of the data collections we are dealing with. Information in registers can be stored in a variety of ways; a key requirement is that each unit in the register can always be uniquely identified. This is best achieved by using a system of identification codes (keys), but identification is also possible without such a code if sufficient information on the units is available, such as for a person name, their address or date of birth.

Ideally, for statistical use, registers should be as comprehensive and accurate as possible, but even incomplete registers can be used for statistical purposes. Basically a register may be a list of all units in the population and nothing more. In practice, most registers also comprise additional attributes for each unit (data). The term data register is used to separate registers with data from the mere lists of units. Here, the term register is used in both senses. “Administrative registers” are registers primarily used in administrative information systems. That means that the registers are used for the production of goods and services in public or private institutions or companies, or that the information is a result of such production. However, in some broader definitions (Wallgren and Wallgren 2007) censuses are considered to be a source of data for statistical registers because they are comprehensive and make updating of unit-level information possible. The term “statistical register”, in its broad definition, describes a register created by processing data from either censuses conducted by the statistical office or from administrative registers generated by authorities and organizations outside the statistical office. This is the definition that will be used in this study.

1.3. BENEFITS OF USING ADMINISTRATIVE DATA IN THE PRODUCTION OF STATISTICS

Several benefits have been identified for using administrative sources to generate statistics. These are discussed in this section.

1.3.1. COPING WITH BUDGET CUTS IN NATIONAL STATISTICAL OFFICES BUDGET CUTS

According to principle 5 of the Fundamental Principles of Official Statistics, adopted by the Statistical Commission in 1994, “*data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on respondents*” (United Nations 2011).

Most national statistics offices (NSOs) seem to have experienced budget cuts, without receiving any priorities or signals on how to reduce the statistical programme or maintain the quality of the statistics. These budget cuts have occurred despite growing interest in statistics on new subjects and on new combinations of variables. In many countries, there has been rising demand for statistics on small population groups or areas, especially with the decentralization policy more frequent than the ten-year based population censuses. At the same time, NSOs have a duty, according to the fundamental principles of official statistics, to produce official statistics with the highest possible quality.

Doing more with less resources will only be possible by shifting the statistics production paradigm from traditional methods of data collection to more efficient use of the already available data records or administrative sources that can be used for statistical purposes.

1.3.2. ROLES OF ADMINISTRATIVE DATA IN SAMPLING AND ESTIMATION

There are many potential benefits to using administrative sources in official statistics, either to complement or replace statistical sources. Of course, it is not all good news, along with the benefits, there are also usually a range of problems to be overcome.

Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness, costs and the burden on

respondents. The presentation below very closely follows Chapter 2 of the United Nations handbook of practices and principles of using administrative data, (United Nations, 2011).

(a) Cost

Statistical surveys are an expensive way to collect data. Questionnaires have to be developed, special staff have to be recruited and trained, samples have to be designed, which may even require the creation of a specific sampling frame, respondents have to be contacted, responses have to be processed and verified and then analysed and results have to be disseminated. Traditional censuses are even more cumbersome because they are conducted on a much larger scale.

Although the set-up costs of using administrative sources to produce statistical outputs can easily be as high as the set-up costs for a statistical survey, the running costs are usually significantly lower.

(b) Response Burden

Using data from administrative sources helps to reduce the response burden on data suppliers. This is a strong political consideration in many countries, particularly if the respondents are businesses. Policies to encourage business development and growth often include reducing regulatory burdens. In these circumstances, statistical surveys are often seen as an easy target for cuts.

Thus, if policy-makers and respondents are united in calling for reductions in the statistical response burden, it is extremely difficult for NSOs to resist this pressure, and the reuse of data collected by others is the logical solution.

(c) Frequency

Related to the reductions in cost and response burden, a further advantage of the use of administrative sources is that they may in some cases allow statistics to be produced more frequently, with no extra response burden and little extra cost. The main constraint to the frequency of statistics produced from administrative data is usually the frequency with which the administrative source is updated. Thus, it would be difficult to produce monthly statistics from administrative data updated once per year unless those data were updated on a rolling basis with no seasonal bias (or at least sufficient information to remove any seasonal biases). Administrative sources that are not based on a particular time period, such as those that record events, such as birth, death or granting of licences for various activities, offer considerable flexibility. This is because, as

long as the date of the event is recorded accurately, statistics can be produced for any given period or periodicity.

(d) Estimates for small areas

Administrative sources often give complete, or almost complete, coverage of the administrative population. The use of administrative sources eliminates sampling errors, removes (or significantly reduces) non-response, and provides more accurate and detailed estimates for various subpopulations, such as respondents in small geographic areas or with other specific characteristics. This is critical in cases in which data are required for the smaller administrative units or subgroups especially in the light of the decentralization policies in several countries.

(e) Timeliness

The use of administrative sources may increase the timeliness of statistical outputs by enabling access to more up-to-date information concerning certain variables. This is because statistical surveys generally take time to plan, to design and pilot forms, to, among other things, analyse the population and optimize the sample. This is particularly the case for annual or ad hoc data collections. Therefore, access to a suitable administrative source can be a more efficient solution. It should be noted, however, that there are also likely to be cases in which the use of administrative sources leads to a reduction in timeliness, particularly regarding short-term indicators. One area in which administrative sources can have a particularly positive impact on timeliness is in the management of statistical registers and survey frames. Administrative information on changes to the target population, such as births and deaths of people or the opening and closing of businesses often provides more information than data collected through statistical surveys, simply because of the coverage benefits mentioned above.

(f) Public Image

Public opinion relating to the sharing of data, particularly between different government departments, varies considerably from country to country. In cases in which public opinion generally accepts, or is in favour of data sharing, the increased use of existing data sources can help to enhance the prestige of a national statistical institution by making it more efficient and cost-effective. Although the public is often apprehensive about data sharing, there are also contradictory pressures to improve the efficiency of government, particularly if this results in lower taxes or more funding in voter-popular areas, such as health

or education. Political slogans such as “joined-up government” are often appealing to the public, and can help to counter fears of loss of privacy.

Thus, the extent to which improvements to public image can be seen as an advantage of using administrative sources depends heavily on how that use is presented to and perceived by the public.

1.4. COMMON PROBLEMS WITH ADMINISTRATIVE DATA AND SOLUTIONS

Many authors have identified a number of problems with administrative data. However, over time, solutions have been found for some of the problems. Second, from a statistical point of view, administrative data also have some disadvantages. For example, the methods used, sources, collection and processing of administrative data and the institutional framework are beyond the control of a national statistics office. It is the data source keeper who manages these aspects, and not the national statistics office. The same is true for the units and variables an administrative data source contains. These are defined by administrative rules and may, therefore, not be identical to those required by NSOs (Wallgren and Wallgren 2007). The disadvantages are predominantly the result of the fact that, in most cases, NSOs use an administrative data source for a purpose different than the one for which the data was originally collected. As a result of this difference, the “statistical” usability of a data source needs to be thoroughly studied by NSOs prior to being used. The issues of quality and use of administrative data that need to be addressed are outlined below, similar to the discussion on the role of administrative data in sampling and estimation, this presentation very closely follows United Nations (2011), namely Chapter 4.

(a) Public Opinion

Public opinion against sharing data around government presents a challenge. It is very difficult to reduce such concerns, but possible approaches could include the publication of clear limits and rules regarding the use of data, ensuring that people and businesses understand that sensitive data used or collected for statistical purposes will not be fed back to other parts of government, particularly tax and benefits agencies.

This is in line with the United Nations “Fundamental Principles of Official Statistics”, in which Principle 5 (“Data for statistical purposes may be drawn from all types of sources, be they statistical surveys or administrative records. Statistical agencies are to choose the source with regard to quality, timeliness,

costs and the burden on respondents”) encourages the use of administrative data and taken together with Principle 6 (“Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes”), establishes the principle of the one-way flow of data.

Other ways to help overcome hostile public opinion include the publication of analyses of the costs and benefits, both to government and to respondents, of the use of different sources. It may also be possible to claim that microdata are more secure when administrative sources are used as no questionnaires are sent by post, data are not held on paper or electronically by interviewers, and fewer clerical staff are needed for the statistical production process, reducing the number of people that have access to sensitive data.

(b) Public Profile

Direct contact with the public through surveys helps to raise the profile of a statistical organization. The use of administrative data can reduce that contact and hence decrease public awareness of the work of the statistical organization. If this becomes an issue, the most obvious solution is to improve the “marketing” of the statistical organization and its data outputs, which may require that a small proportion of the savings from using administrative sources be transferred to the marketing budget. Perhaps the most effective way of promoting the activities and outputs of a national statistical organization, particularly in the medium to long term, is to ensure greater involvement with education institutions, business groups and other target customers. User groups are also particularly important in this respect, and should be actively encouraged.

(c) Management of Changes

Public sector administrative sources are generally set up to collect taxes or monitor government policies, making them susceptible to political changes. If a policy changes, administrative sources may be affected in terms of coverage, definitions or thresholds or possibly even abolished completely. For example, in Uganda, the institutional framework for providing agricultural extension services has changed from being centralized to decentralized and back to being a centralized system. This has affected the collection of administrative data (UBOS 2007). Changes to the computer systems used to store and process administrative data may also have an impact on the supply of data for statistical purposes. Private sector sources are also not immune from those sorts of

changes, though for them, the changes are mostly driven by changing market factors. Such changes may happen suddenly, with little warning. Particularly high-risk times tend to be immediately after a change of government, a change of minister, or a change in legislation. The risk of a policy shift leading to a structural change in administrative data is discussed in United Nations (2011).

Changes in government programmes have affected the administrative data of several statistical offices in developed countries. Swedish tax legislation resulted in the aggregation of farm income and other forms of income on tax returns, limiting the uses of tax data for producing statistics on farmer income (Cafagna and Carfagna 2010). As a different example, a consequence of a change in policy governing distribution of agricultural subsidies in Europe was that farmers were no longer required to report area planted with certain specialty crops. This policy change degraded the quality of administrative data on area planted in minor crops. In some cases, a change in policy can have a positive effect on the resulting administrative data. Brackstone (1987) reports an example in which a change in tax policy led to an increase in the number of people required to file a tax form, improving the coverage of the administrative data.

Another good example occurred in Slovenia some years ago when the supply of administrative data on employment was halted for a while following a change of minister, leaving the statistical office with serious problems in its efforts to produce employment statistics. In the case of Uganda, strong support to the Community Information System programme declined following changes in the ministers for finance. Procedures, backed up by legislation, have since been implemented to minimize the likelihood and impact of this type of change. Reliance on a particular source carries a certain degree of risk. These risks can be managed to some extent by legal or contractual provisions. The best way in practice to avoid such problems tends to be through regular contact with those responsible for the administrative source in order to ensure that they are aware of the statistical requirements, and to try to influence and get early warning of any possible changes. This seems to have worked very well in the case of customs data to generate export/import data in several countries where close working relationships have been established between the central statistics office and customs authorities. These often include drawing up memoranda of understanding.

(d) Units

One major problem often encountered when using administrative sources is that the units used in those sources do not correspond directly to the definition of the required statistical units. The process of converting from administrative units, such as legal units, tax units and claimants to statistical units, such as enterprises, people and households can be quite difficult conceptually, and often involves some form of modelling. In business statistics, this process is known as **profiling**, and typically is a function of statistical business registers.

Eurostat has published guidelines for this process in Chapter 19 of its Business Registers Recommendations Manual, in which it defines **profiling** as “a method to analyse the legal, operational and accounting structure of an enterprise group at national and world level, in order to establish the statistical units within that group, their links, and the most efficient structures for the collection of statistical data” (Eurostat 2015).

Although profiling gives a better understanding of complex unit structures, it is expensive and time consuming, and requires trained staff, making it impractical to conduct detailed clerical profiling for all business units in an economy. Instead, this method should focus on cases that provide the most benefit. Profiling can be seen as a trade-off or compromise between three factors:

- Quantity of business structures profiled;
- Quality or depth of the profiling activity;
- Available resources (determined both by cost and suitability of staff).

Although clerical profiling is not practical for all units in a large population, some form of automated, rules-based profiling may be. Standard rules based on attributes or the nature of links between units can help to overcome differences between administrative and statistical units in many areas of statistics. For example, statistical households can be derived based on relationships between the individuals living in a building. This approach is used successfully within the register-based population census methodology applied in Nordic countries. For example, in 2011, the first register-based census was carried out in Norway in which data from the Register of Ground Properties, Addresses and Buildings (GAB Register) was used. Data from census forms were used to establish links between resident persons and their dwellings in the Central Population Register (CPR) and to update information on dwellings in the Ground properties, Addresses and Buildings Register (Andersen and Utne, 2007; Statistics Norway, 2011).

(e) Definitions of Variables and Differences in the Definitions of Units

There are also likely to be differences in the definitions of variables between administrative and statistical systems. The data in administrative sources have generally been collected for a specific administrative purpose, and the needs and priorities relating to that purpose are likely to be different to those of the statistical system. For example, turnover for VAT purposes may not include turnover related to the sales of VAT exempt goods and services, whereas the statistical system is likely to require total turnover.

Another common example is the definition of unemployment. The standard statistical definition is:

“The "unemployed" comprise all persons above a specified age who during the reference period were:

(a) "Without work", i.e. were not in paid employment or self-employment;

(b) "Currently available for work", i.e. were available for paid employment or self-employment

during the reference period;

(c) "Seeking work", i.e. had taken specific steps in a specified recent period to seek paid employment or self-employment”.

However, definitions of unemployment in administrative sources are more often based on the number of people claiming unemployment benefits, or registered as looking for work. Some people who are out of work may not register as unemployed, if they expect to find work quickly, and in some cultures, there may be a social stigma attached to claiming unemployment benefits. On the other hand, some people claiming unemployment benefits may not be available for work or actively seeking work and thus should not be counted as statistically unemployed. Some countries have very limited unemployment benefit systems so the register captures only a small part of the unemployed.

(f) Classification Systems

As is the case for variables, the classification systems used within administrative sources may be different to those used in the statistical world. Even if they are the same, they may be applied differently depending on the primary purpose of the administrative source, perhaps focusing on specific attributes of the unit. For example, an administrative source concerned with licensing, health and safety or environmental protection may be more directed at the economic activities of a business that are of most concern to that source,

rather than the main economic activity of a business, which is required for statistical purposes.

In other cases, classifications in administrative sources may not be applied at the level of detail required for statistical purposes or the classification may simply not be a priority variable for the administrative source, resulting in quality deficiencies. In cases in which classification systems or versions are different, the usual solution is to construct conversion matrices to map the codes in the administrative classification to those in the statistical classification. Such mappings may be one to one, many to one, one to many or many to many. In the latter two cases, some sort of probabilistic allocation may be required.

(g) Timeliness

There are three separate issues relating to timeliness that affect the usefulness of administrative data for statistical purposes:

- Administrative data may not be available in time to meet statistical needs;
- Administrative data may relate to a period that does not coincide with the statistical reference period;
- Administrative data may be measured over a period, while the statistical requirement is for a specific point in time (or vice-versa).

Considering the first issue, there will generally be a lag between an event happening in the real world, and it being recorded by an administrative source. This is then followed by a further lag before the data are made available to the national statistical organization.

(h) Inconsistency between Sources

Inconsistencies between sources are a specific problem. It occurs when multiple sources are used. Data from one source may appear to contradict those from another. This may be because of different definitions or classifications, differences in timing, or simply an error in one source and can happen when comparing administrative data with statistical data, or when comparing two administrative (or two statistical) sources. To resolve such conflicts, it is necessary to establish priority rules by deciding which source is most reliable for a particular variable. Once a priority order of sources has been determined for a variable, it should then be possible to ensure that data from a high priority source are not overwritten by a lower priority source. This process is much

easier if source codes are stored alongside variables for which several sources are available.

(i) Missing Data

The problem of missing data is not unique to administrative sources. It can also be due to total or partial non-response to statistical surveys, or even to the removal of data values during the editing process. However, with administrative sources, the issues can sometimes be different, particularly as the problem of missing data can often be more systematic.

The main reasons for this are that a particular variable may not be collected at all by the administrative source, or it may only be collected for certain categories of units to which there is a specific administrative requirement. The variable may also simply be a low priority for administrative purposes, so the owners of that source do not see the missing data as a problem.

Some of the standard solutions for dealing with non-response in statistical surveys can also be used to solve the problem of missing data in administrative sources. Various imputation methods, such as deductive, “hot-deck” imputation or “cold-deck” imputation are often suitable when the problem only affects some of the units. Another solution is advocacy for proper administrative records for statistical purposes to improve the basic data quality.

(j) The Dilemma: Preservation of Confidentiality while using Administrative Registers

One of the main problems in using information coming from administrative registers is confidentiality. All national statistical systems are based on the secrecy of individual information. Statistical laws and acts usually have a section devoted to secrecy of statistical information. Confidentiality is seen as a guarantee for both the informant (the data will not be used against him) and the administration (confidentiality helps ensure that the data are more accurate because the informant has no fear of being harmed by the administration). Information contained in administrative registers cannot be protected by this principle. Therefore, NSOs face some restrictions from using administrative registers.

How then do we preserve confidentiality while accessing and using the administrative data for statistical purposes? One of the solutions to this challenge would be for the offices responsible for generating administrative data to work with NSOs to come up with protocols for data anonymization.

Anonymization makes possible accessibility to data while at the same time protects the entities that provide the data. The practice in Nordic countries of jointly developing data collection protocols coupled with the development of a legal framework covering confidentiality of respondents providing administrative data may be a feasible approach (Nordbotten 2008).

(k) Resistance to Change

One of the main barriers to the more effective use of administrative sources in official statistics, and one of the least recognized, can come from within the organization. Statisticians may resist the use of administrative data because they do not trust data that they have not collected themselves. They often focus on the negative quality aspects of administrative data and have an overoptimistic view of the quality of survey data, often based on the largely untested assumption that survey responses actually comply with statistical norms.

The solution is clearly through better education of statisticians regarding the possibilities offered by administrative sources, encouraging them to take a wider view of all the dimensions of quality and focus on the impact on data suppliers and users. In this context, it is important to determine the real relative quality of survey and administrative data.

(l) Administrative Data Processing

According to the United Nations (2011), even if the incoming data are perfect, their quality can still be affected by the different processes they go through before being used for statistical outputs. Ideally processing should improve quality, but unfortunately this is not always the case. The report gives the following examples of how data processing can affect quality:

- Data matching and linking – too many false matches will lead to errors in the data; too many false non-matches will lead to duplication, which will overstate the size of the population of interest, and possibly introduce bias.
- Outlier detection and treatment – using outlier detection methods to detect errors can help to improve the quality of the data, and generally the more extreme the outlier, the more likely it is to be an error. However, overzealous treatment of outliers will result in genuine data values being altered and can lead to important trends in the data being missed.

- Quality of data editing – as for outlier detection and treatment, data editing should improve quality, but if not done carefully it can introduce error and bias.
- Quality of imputation – Using imputation missing values or records can help to improve coverage, but again the methods used must be scrutinized carefully to avoid the introduction of bias.

One very important principle that should always be followed, particularly when processing data from administrative sources, is to keep a copy of the raw data (and any associated metadata) for reference purposes, if necessary.²

Metadata are vital for informing both producers and users about data quality and should be present at all the stages. Incoming data should be accompanied by sufficient metadata to fully understand them and to ensure that values are correctly allocated to the relevant variables. Metadata are at the heart of the management of the interpretability indicator. The Integrated Metadata Base (IMDB) is the single source of metadata information describing surveys and programmes for Statistics Canada. The quality of IMDB information has to be monitored regularly to ensure completeness and accuracy. Statistical agencies should publish good metadata as this will show openness and transparency and build trust with data users (Dion 2007).

Detailed documentation on the concepts, definitions and purpose of the source, as well as on the collection and processing methods used, is also important. This helps to create a better understanding of potential quality issues, and should form the basis for data editing rules in the processing stage. During data processing, it is important to record what has been done to which records and values. This not only provides essential information for assessments of processing quality, but also furnishes a mechanism to investigate potential problems in the process and undo any errors.

Statistical outputs should be accompanied by sufficient metadata to allow users to retrieve them, interpret them correctly, and form an opinion on their quality.

² For a comprehensive collection of papers on different data editing issues, see the working papers of the Statistical Data Editing Work Sessions organized by the ECE. Available at: <http://www1.unece.org/stat/platform/display/kbase/UNECE+Work+Sessions+on+Statistical+Data+Editing>.

(j) Summary of Problems and Solutions

Vale (2007) concludes that there are many different barriers to the effective use of administrative data. Some are source or country-specific, whereas others are more general in nature. Access problems may include: legal restrictions; policy considerations; organizational arrangements; and technical standards. On the other hand, quality problems include coverage and accuracy. There could also be cases of internal resistance, for example if the NSO does not trust the data.

The main barriers commonly identified by statistical organizations are problems of access and quality. The list above also clearly shows that there are many problems to overcome when using administrative sources. It also indicates that there have been similar problems other sources, and that in most cases, it is possible to find full or partial solutions. The list cannot cover all potential problems that may be encountered, particularly those that are source-specific, but the intention is to give ideas that can be adapted to meet specific circumstances.

Overall, it appears that most problems encountered in the use of administrative data for statistical purposes, which is common with many other areas of statistics, can be overcome, or at least reduced, by effective planning and management, good knowledge of data sources, creative thinking and willingness to exchange experiences and learn from others. As shown above, administrative data often require different processing than statistical sources. Simply substituting administrative data for statistical data without changing the statistical production process rarely works in practice.

Finally, it must be recalled that despite all of the problems, the benefits of using administrative data still often greatly exceeds the costs.

2

Proposed Framework for Improved Systems of Administrative Agricultural Statistics

Section 2 discusses the proposed analytical framework, structure, conduct and performance for assessment of the Administrative Data System Design for Agricultural Statistics (ADSAS) and a quality framework for assessing data quality. Various frameworks exist that can be used to assess agricultural systems. Ballivian et. al. (undated) discussed the use of a standardized quality assessment tool of administrative registers applied in the Yucatan state in Mexico. The tool helps to identify areas that need improvement, as well opportunities, problems and their root causes, so that corrective actions can be implemented to eliminate those causes. The tool is further discussed in section 4.10 and the detailed indicators are in Appendix II: Table B. The design framework given in Table 2.1 will be used for the review of the systems while Table 2.2 will be used for the assessment of quality.

2.1. ANALYTICAL FRAMEWORK ADAPTED FROM THE AGRICULTURAL MARKET INFORMATION SYSTEMS IN SUB-SAHARAN AFRICA

A review of the analytical framework for assessment of Agricultural Market Information Systems developed by Kizito (2011) shows that it is comprehensive and can be used for the review of other systems' structure, conduct and performance. The framework was modified (see Table 2.1) for the purpose of assessing ADSAS in countries.

Structural design issues	Conduct design issues	Performance (quality of core data items covered)
<p>1. The ADSAS perceived mandate (aims, objectives, and clientele)</p> <ul style="list-style-type: none"> • Policy formulation and monitoring • Food security planning and monitoring • Attainment of efficient markets • Clientele (e.g. farmers, traders, consumers, government, donors) <p>2. Institutional home, organization, and coordination</p> <ul style="list-style-type: none"> • Public-, private-, farmer organization, or trader and NGO-based ADSAS • Provides complementary services that generate or increase value of information • Geographic coverage and range of commodities • Assuring coordination among stages <ul style="list-style-type: none"> o Integration of ADSAS Activities o Centralized or decentralized ADSAS activities o Specialization in ADSAS Products • Design of incentives for ADSAS staff • Profit orientation of the ADSAS <p>3. Nature of core data items covered (crop items, livestock items, poultry, aquaculture and fisheries products, agro-forestry production, agricultural inputs, land cover, apiary and horticulture). These are described in detail below.</p>	<p>1. Information provided</p> <ul style="list-style-type: none"> • Raw data • Analysis of raw data • Analytical reports <p>2. ICT used in the collection and dissemination</p> <ul style="list-style-type: none"> • Traditional ICT (e.g. radio, television, and fax) • Modern ICT (e.g. email, internet, SMS) • PDAs and GPSs <p>3. Funding strategies</p> <p>4. Data collection methods used</p> <ul style="list-style-type: none"> • Structured questionnaire and enumerators • Wiki approach (users SMS or update web) <p>5. Quality control methods used</p> <p>6. Feedback mechanism used</p>	<p>1. Coverage</p> <p>2. Comprehensiveness</p> <p>3. Timeliness</p> <p>4. Punctuality</p> <p>5. Completeness</p> <p>6. Relevance</p> <p>7. Accuracy</p> <p>8. Reliability</p> <p>9. Integrity credibility</p> <p>10. Accessibility to different clientele</p> <p>11. Clarity/interpretability</p> <p>12. Comparability</p> <p>13. Consistency/ coherence</p> <p>14. Sustainability of ADSAS</p> <ul style="list-style-type: none"> • Financial support • User support • Cost minimization

TABLE 2.1: Structure, Conduct and Performance Design Issues of an Administrative Data System for Agricultural Statistics
Adapted from Kizito (2011).

2.1.1. STRUCTURAL ISSUES

In this analytical framework, the structural design issues refer to the relatively stable features of the administrative sources related to agriculture. These include: (a) the perceived mandate (aims, objectives, and clientele) of the system, (b) the institutional home, organization, and coordination of the sources, and (c) the nature of the commodities to be covered. How the administrative sources fit in the overall integrated food and agricultural statistics system is also reviewed.

Under the perceived mandate, the framework is able to analyse how the administrative sources are used for policy formulation and monitoring, food security planning and monitoring, and attainment of efficient and fair markets.

Other uses of the sources that may be analysed include: support to the day-to-day implementation of agricultural policies; programmes and projects; providing information to farmers, traders, input dealers and agricultural processors; reporting on progress towards achieving international agreements and goals; and supporting investment decisions. This is important to demonstrate that there is effective demand for the administrative sources, a key indicator for sustainability. The administrative system's clientele includes farmers, traders, consumers, government and donors. More issues about the perceived mandate of the administrative systems will be continuously updated as more literature review is conducted.

In Table 2.1, under institutional home, organization and coordination, the framework can examine whether the existing administrative sources are public-, private-, farmer organization, or trader and NGO-based; whether they provide complementary services that generate or increase the value of information; their geographic coverage and range of commodities covered; and how they are coordinated, such as whether integrated, centralized, decentralized or specialized in one or more items. An inventory of all administrative sources will be made prior to applying the analysis framework. The framework also looks at the design of incentives for administrative staff and the profit orientation of the administrative system — factors that ensure sustainability. Under the nature of commodities covered, whether the existing administrative sources cover, among others, agriculture (crops), livestock, poultry and fisheries is determined.

2.1.2. CONDUCT ISSUES

In the context of this study, conduct refers to the behaviour of the administrative systems. In a way, the conduct issues are highly related to processes. The conduct or process includes the following: (a) the information provided by the administrative system, including data sharing agreements and processes; (b) the information and communications technology (ICT) used in transmission and diffusion of the administrative data (data management process); (c) the funding strategies; (d) the data collection methods; (e) the quality control methods used; and (f) the feedback mechanism used by the administrative systems.

Under the information provided, the framework examines whether administrative sources provide raw data and some analysis of the raw data and whether some analytical reports are written. The framework also covers the ICTs used by existing sources in both the collection and dissemination of administrative data. Another dimension of ICT usage is to determine whether

the administrative registers are digitized, and any other issues affecting the data entry and transmission. ICT can be disaggregated into traditional and modern ICT, which can affect the cost implications and sustainability of the administrative systems. Traditional ICT includes the use of technologies, such as Fax, radio transmission, broadcast national and community radios, and even the post office box to transmit and diffuse information. On the other hand, modern ICT includes, among others, technologies, such as handheld devices for data/information collection, short message service (SMS) and voice using cell phones, and other cell phone-based methods of transmitting and diffusing information, such as use of personal data assistants (PDAs) and data entry through scanning technology. For example, customs entry documents are processed directly online. The framework also examines the funding strategies of the administrative systems, such as from donors, central or national governments, local governments, non-governmental organizations (NGOs) and individual private organizations. The framework also looks at the data collection methods used by the administrative systems.

2.1.3. PERFORMANCE

In the structure, conduct and performance paradigm in Table 2.1, performance refers to the extent to which markets result in outcomes deemed good or preferred by society. This concept is adjusted from a market perspective to an administrative sources perspective to refer to outcomes that are deemed to be good or preferred by administrative sources' stakeholders. The performance issues mostly relate to data quality concepts, some of which are quite subjective and interrelated. The performance concepts relating to data quality draw on those used in various statistical systems, such as Statistics Canada, the National Agricultural Statistical System of the United States Department of Agriculture and the United Nations (Dion, Chartrand and Murray 2010; Jorner and Vogel 2010). In the context of ADSAS, this analytical framework looks at performance in terms of: (a) coverage; (b) comprehensiveness; (c) timeliness; (d) punctuality; (e) completeness; (f) relevance; (g) accuracy; (h) reliability; (i) integrity/credibility; (j) accessibility to different clientele; (k) clarity/interpretability; (l) comparability; (m) consistency/coherence; and (n) sustainability of ADSAS. Sustainability is examined in three aspects: (a) financial support; (b) user support; and (c) cost minimization.

2.2 THE HEALTH MATRIX NETWORK FRAMEWORK.

Another framework that could inform the process of review of relevant frameworks for agricultural administrative data systems is the Health Matrix Network, which was formed to assist low- and middle-income countries in

efforts to strengthen their health information systems. Its goal is “to increase the availability and use of timely and accurate health information in countries and globally through shared agreement on goals and coordinated investments in core health information systems”. HMN that supports countries to assess and plan their health information systems endeavours to document improved health outcomes that can be credibly attributed to increased use of information for decision- making in at least 80 developing countries by 2011 (AfDB, PARIS21 and Intersect 2007). In 2005, the Network developed a tool to assist countries to assess and plan their health information systems, namely a framework and standards for country health information system development. The tool looks at six components of a health information system, namely resources, indicators, data sources, data management, information products, and data dissemination and use. These components are assessed across sectors, agriculture included, as can be seen in the Table A (See Appendix I) (AfDB, PARIS21 and Intersect 2007).

Basically, the assessment should help to answer the following questions and lead to the development of strategies that could cut across the national statistical system (AfDB, PARIS21 and Intersect 2007):

Public awareness

- *Are the advocacy strategies employed sufficient to raise public awareness of the importance of the data produced by this sector?*

Use of resources

- *Are the arrangements for coordination between users and producers within the sector and between the sector and users and producers in other sectors adequate for the sector to function efficiently within the national statistical system?*
- *Do the existing legal and institutional frameworks adequately reflect the sector’s mandate in relation to the national statistics office and other agencies and sectors in the national statistical system? For example, do they clearly identify lead responsibilities for data production and dissemination?*
- *Are the organizational functions, relationships, operating procedures and management strategies appropriate to meet the sector’s mandate?*

- *Is the budget allocated to the statistical system adequate to meet current and projected needs?*
- *Are the sector's human resource strategies for recruiting, training, supervising and promoting staff adequate to allow the system to function properly and maintain standards?*
- *Is the physical infrastructure – information and communication technology, office equipment, office buildings, vehicles, among others, – adequate for the system to run efficiently?*

Productivity of data management

- *Could the data management system be improved to facilitate efficient data production, such as reduce duplication of effort and fill gaps in the system?*

Data availability

- *Are there clear definitions of all data produced? Are they archived so that they can be accessed by users and producers in the sector and in other agencies and sectors of the national statistical system?*
- *Does the system produce the appropriate set of indicators to assess sector performance?*
- *Does the system produce the appropriate set of indicators to monitor development goals, for example the Poverty Reduction Papers or the Millennium Development Goals?*
- *Does the data production calendar meet user needs and make efficient use of the system's resources?*
- *Are there agreed and appropriate standards for the quality of data produced by the system and are they implemented?*
- *Are the indicators clearly defined, accurately measured, representative, reliable and made available in a timely fashion?*
- *Are the existing dissemination policies and plans for statistical production adequate?*

- *Does the system provide adequate training to assist users in the sector and in other agencies and sectors in the national statistical system to make best use of the data produced?*

The information should enable undertaking a strengths, weaknesses, opportunities and threats (SWOT) analysis to be used in the design of the sector statistics plans.

The guide on mainstreaming sectors into the National Strategy for the Development of Statistics (NSDS) recommends that an assessment on the state of statistics in a sector should cover the above components. In addition, it proposes that information be collected on the mandate of the sector, key sector development plans/programmes, main stakeholders in the sector information system, major gaps in such areas as infrastructure, personnel, skills and funding and major challenges pertinent to statistical development in the sector. In order to ensure uniformity in information collection among sectors, a standard format should be developed and used in the assessment.

Efforts to develop and strengthen national statistical systems are ongoing. This process is being carried out based on the recommendation made in the Marrakech Action Plan for Statistics (MAPS) in 2004³ to develop and implement the national statistical systems. The main action point of MAPS is the design and implementation of NSDS. AfDB, PARIS21 and Intersect (2007) presents best practices in the design of NSDS that involve the design of sector statistics plans for each sector, including the agricultural sector, and using these plans as building blocks for NSDS – the sectoral approach to the design of NSDS. Currently, countries in Africa are designing sector statistics plans for agriculture as part of the NSDS design process or as stand-alone activities in cases in which NSDS have not been designed or were designed without using the sectoral approach.

The design of an agricultural statistical system involving the use of administrative data should factor in the process of its development, the Health Matrix Network initiatives and the NSDS frameworks, which are already being implemented in a number of developing countries with some success.

³MAPS was endorsed at the 2nd Roundtable on Managing for Results (2004) “*to improve national and international statistics to meet the measurement challenges without which baselines cannot be established and progress cannot be measured*”.

2.3. COMPLEMENTARY ARRANGEMENT FOR AN ADMINISTRATIVE DATA SYSTEM, SURVEYS AND CENSUSES

The rationale behind the proposed complementary approach is that based on experiences and the practices of various countries, using administrative data sources is cheaper than using censuses and sample surveys (Kiregyera 2002; United Nations 2007). In the course of the review and development of a system of agricultural statistics in which systems are found to collect similar data, the administrative system with appropriate improvements is preferred to the survey and census systems. When data are required for the smallest units, it is often best to use administrative systems to collect such data based on technical and cost considerations. The costs of getting such data from surveys and censuses are generally prohibitively high. Data for commercial crops, perennial crops and early warning can also be collected administratively. According to a brief on use of administrative data for statistical purposes in Pakistan, “since at all times one must strike a balance between the quality desired and the practical and economic realities, the administrative data will always be preferred if there is reasonable quality” (United Nations 2011). Administrative records can be a substitute for survey data and by so doing, the response burden on the citizens is reduced and NSOs are relieved of carrying out the function of direct data collection. As to whether they can be a substitute or complementary is an issue that has had to be examined along the way (United Nations 2007).

2.4. REVIEW OF THE QUALITY OF ADMINISTRATIVE DATA

The review will be done to determine the quality of each type of data being collected, including crops, livestock, fisheries, poultry and bee keeping, crop forecasting (early warning), production estimation statistics, market and price data, employment and labour, water and land use, agro-forestry, horticulture, inputs producers and distributors (table 2.2).

Data quality criteria	Timeliness	Punctuality	Relevance	Accuracy	Accessibility	Clarity/interpretability	Coherence/consistency
Crop items							
Livestock items							
Poultry							
Aquaculture and fisheries products							
AgroForestry production							
Agricultural inputs							
Land cover							
Apiary							
Horticulture							

TABLE 2.2: Data Quality Evaluation Framework

Note: Many systems produce data on more than one item.

Before data are used, a major issue is its quality. Statistics Canada defines data quality in terms of “fitness for use”. Six dimensions of quality have been identified within the concept of “fitness for use”, namely relevance, accuracy, timeliness, accessibility interpretability and coherence (Dion 2007). ADSAS will consider all these dimensions when assessing the quality of administrative data. In addition, sustainability and other quality aspects of the system will also be analysed. Among them are the following:

Timeliness – this refers to the continuous and consistent diffusion of information to stakeholders. It may refer to providing information to stakeholders when it is needed (Kizito 2011). Timeliness may also reflect the length of time between data being made available (date of publication) and the event or phenomenon being described.

Punctuality – the time lag between the date that data were actually released (date of publication) and the target release date (often pre-announced).

Relevance – the degree to which statistics meet the needs of current and potential users. It also covers methodological soundness and the extent to which concepts used reflect user needs.

Accuracy – the closeness of statistical estimates to true values. The challenge is often “What does accuracy mean for administrative data?” Distance from the true value may be attributed more often to a bias (coverage error) more than to a sampling error. Accuracy may also be described in terms of the major sources of error that potentially cause inaccuracy (e.g. coverage, sampling, response, non-response) (Dion 2007). In some countries attempts are made to compare

administrative data to survey or even census data whenever similar data become available (Trant 2011).

Accessibility – accessibility in the quality framework means accessibility to the final data users. In cases in which data are not published or at least not classified as “official” statistics by the NSO, its use will be greatly curtailed. Accessibility may also include the demand or the effective demand of the data, especially when generated from the private sector, in cases in which user or convenience fees or costs exist. Additionally, accessibility also covers such issues as data users, purpose of use and frequency of use, the physical conditions in which users can obtain data (where to go, how to order, delivery time, clear pricing policy of convenient marketing conditions, such as copyright), availability of micro or macro data and various formats, such as paper, files, CD-ROM, Internet, etc.

Clarity/interpretability – whether data are accompanied by sufficient and appropriate metadata, whether illustrations, such as graphs and maps, add value to the presentation of the data, and whether information on data quality is available. The illustrations make any data, administrative or otherwise, more user-friendly. It also examines the extent to which differences between statistics are attributed to differences between the true values of the statistical characteristic, or to methodological differences.

Coherence/consistency – data from different sources, in particular from statistical surveys of a different nature and/or frequency, may not be completely coherent in that they may be based on different approaches, classifications and methodologies. They may not, therefore, convey a completely coherent message to users, as many of them may be confused if two different measures of the same variable are published with different values.

This list of criteria can be used in two ways relating to administrative data. First, it can be used to assess the quality of the resulting statistics, and to compare data based on administrative sources with those based on surveys or censuses. Second, it can be used to help evaluate the quality of different administrative sources.

As the United Nations Economic Commission for Europe concludes, “moving from survey to administrative sources will clearly have an impact on output quality. Typically this impact may be positive for some quality criteria, and negative for others. In all cases, it is necessary to get an overall view of the impact, giving greater weight to those criteria the users consider to be the most

important. For example, users may feel that an improvement in timeliness more than compensates for a reduction in accuracy, particularly for short-period economic data. Another consideration should be the impact on time-series data, and whether it is possible to construct a consistent series of sufficient length following the change. It can be particularly important to give at least as much weight to the views of users as to the perceptions of statisticians, which may, in some cases be too heavily focussed on traditional notions of accuracy. Overall, it is vital that any judgement of the impact on statistical outputs is based on objective evidence rather than on supposition, as this is the only way to counter the potential for resistance to change” (United Nation 2011, p. 41).

Administrative data are produced on the basis of some administrative processes, and units and variables are defined out of administrative rules and demands. The definitions may differ from the needs of the official statistics, but the data are usually sufficient for administrative purposes. In simple terms: Using traditional data collection methods, the NSO is asking: "We asked our question, but do we get the right answer?" Using register-based (administrative) data the NSO is asking: "Surely we got the correct answer, but how does it correspond to our question?" The NSO is working towards the same goal, but from a different starting point. However, it is important to note that even though administrative data records can never fully replace direct data collection by sample surveys, these two methods can complement each other.

2.5. CORE DATA ITEMS TO BE REVIEWED

UNBS (2012) provides a checklist that is useful for assessment of data quality, while Chapter 3 of the Global Strategy (World Bank, FAO and United Nations, 2010) provides details of the variables to be reviewed, including core data items under the first pillar, “identifying a minimum set of core data and determining national priorities”. Only a subset of the minimum set of core data should be considered, the one from administrative sources.

Under the global strategy (World Bank, FAO and United Nations 2010), the core items include:⁴

(a) Crops: Wheat, maize, barley, sorghum, rice, sugarcane, soybeans, and cotton. The data to be reviewed for these core crops will include: (i) area planted and harvested, yield, and production; (ii) amounts in storage at the beginning of harvest; (iii) area of cropland that is irrigated; (iv) producer and

⁴ The items will vary from country to country.

consumer prices; (v) amounts used for own consumption, food, feed, seed, fibre, oil for food, bio-energy and net trade or imports and exports; (vi) (crop forecasting) – early warning indications, such as precipitation, windshield surveys of crop conditions and vegetative indices provided by satellite observations; and (vii) employment and labour.

(b) Livestock: Cattle, sheep, pigs, goats and poultry as they are major sources of food supply and agricultural income. Data required for these items include: (i) inventory and annual births; (ii) production of products, such as meat, milk, eggs, and wool; (iii) net trade or imports and exports; (iv) producer and consumer prices; and (v) employment and labour.

(c) Aquaculture and fisheries products: Aquaculture production entails the use of land and water resources. Fisheries provide livelihoods for small-scale and inland holdings. Data required include: (i) area cultured; production; prices; and net trade or imports and exports for aquaculture; (ii) quantity landed and discarded number of days fished; amounts processed for food and non-food uses; prices; employment and labour; imports; and exports.

(d) Forestry products: Forestry is a major category of land use. It provides income and plays a significant role in understanding the forces affecting climate change. Data required include: (i) area in woodlands and forests; quantities removed; and their prices for land associated with agricultural holdings; (ii) area in woodlands and forests; quantities removed; and their prices for products from non-agricultural holdings and respective utilizations, employment and labour.

(e) Agricultural inputs: Core inputs to agricultural production include labour, chemicals, water, energy, and capital stocks. These inputs are considered to be core because in combination with data about outputs, they provide measures of agricultural productivity important to monitoring and evaluating steps to reduce poverty and hunger. Data required include: (i) quantities of fertilizer and pesticides uses; (ii) water and energy consumed; (iii) capital stocks, such as machinery by purpose (i.e. tillage or harvesting); (iv) number of people of working age by sex; (v) number of workers hired by agricultural holders; (vi) employment of household members on the agricultural holding; and (vii) markets and prices.

(f) Land cover: Mapping products or digitized data from remote sensing should provide complete coverage for the entire land mass of a country with the following classifications: (i) cropland; (ii) forest land; (iii) grassland; (iv)

wetlands; (v) settlements; (g) other land; (h) water; and (i) employment and labour.

Some of these data items will be evaluated using the data quality framework given in Table 2.2 with appropriate scores, for example, very poor, poor, fair, good (adequate) and very good, depending on information availability.

2.6. REVIEW OF DATA USE

There is a wide range of uses to which agricultural statistics data collected by administrative sources can be put. For the purpose of this project, the review is limited to use for official purposes, such by the national statistical offices, government ministries, parastatals and agencies. In the course of the review, other checklists and data quality frameworks found to be appropriate will be reviewed and adapted for the assessment.

As indicated in Table 2.3, criteria are whether there is ease of access, whether the data is used and frequency of use. The review will also endeavour to identify systems with potential for providing data. Such systems can then be enhanced and improved for use.

Data use criteria	Who uses it	For what – current & potential	Accessibility	Frequency of use
Crop items				
Livestock items				
Poultry				
Aquaculture and fisheries products				
Agroforestry production				
Agricultural inputs				
Land cover				
Apiary				
Horticulture				

TABLE 2.3: Review of Data Use

In cases in which there are significant differences by regions or language groupings (French-Portuguese-Spanish versus English), in policies and administrative structures (centralized versus federal /decentralized systems) or in any other identified areas will be highlighted.

3

Specific Cases of Using Administrative Data

Agricultural statistics play an important role in policy-making and in studying the relationship between agriculture and other issues, such as food security, poverty, environmental health and economic development. Demand of agricultural information at lower geographical levels is increasing. Sample surveys and censuses are the most reliable way to produce agricultural statistics. However, in many developing countries, owing to resource limitations, it is impractical to carry out sample surveys at the spatial resolution and/or temporal frequency necessary to provide timely information to meet the demand. At the same time, a large amount of administrative data are available as a by-product of other administrative processes. The administrative data typically are collected at a higher frequency than survey data and provide better spatial coverage. In addition, they become more accessible with the improvement of computer technology. This brings forward the challenging problem of how to effectively use administrative data to improve agricultural statistics.

3.1. DEVELOPED COUNTRIES

The statistics agencies in the developed countries in North America and Europe have conducted extensive research in this area. The Center for Survey Statistics and Methodology at Iowa State University has worked with the National Agricultural Statistics Service of the United States Department of Agriculture over the past four years on the use administrative records and other remote sensing data to improve the estimation of crop acreage and to develop an automated sampling design for a real-based agricultural survey.

3.1.1 ADMINISTRATIVE DATA IN AN INTEGRATED NATIONAL STATISTICAL SYSTEM

Administrative data serve multiple purposes in an integrated national statistical system. As administrative data are often gathered naturally through governmental and economic programmes, the process of acquiring administrative data requires relatively little additional cost or respondent burden. Many sources of administrative data can be used for agricultural statistics, in particular. These include tax data, land ownership records, vehicle registration data, information on government subsidies, insurance information, records on agro-tourism and meteorological data.

Many federal statistical systems in developed countries have extensive experience in using administrative data in the production of official statistics. Brackstone (1987) reviews opportunities and challenges associated with the use of administrative records from a perspective informed by experiences at Statistics Canada. Wallgren and Wallgren (2010) and Wallgren and Wallgren (2011) discuss register-based statistics, with an emphasis on the use of register systems in Nordic countries, including methodological developments required for expansion and improvement of register-based statistics. Carfagna and Carfagna (2010) analyse the strengths and weaknesses of register-based statistics. They also discuss several ways to integrate administrative data in multiple stages of the survey process, including design, data collection and estimation. All in all, uses of administrative data in statistical offices in developed countries have the potential to reduce associated survey costs because they require smaller sample sizes, reduce respondent burden, improve small area estimates, enhance estimates of change over time, reduce variances of estimators through incorporation of auxiliary information and expand coverage of sampling frames. In achieving these ends, incorporating multiple sources of administrative data is recommended to improve coverage and consistency. Methodology for record linkage and evaluation measurement error is needed to maintain high quality databases that integrate multiple administrative files. These uses and associated methodological issues are discussed in more detail in Sections 3.1.2 and 3.1.3.

Interest in administrative data for official statistics has grown in recent years, owing to increases in computational resources coupled with government budget cuts. In a discussion in a special issue of *Statistica Neerlandica* on the topic of register-based statistics, Karr (2012) explains, “In today’s era of curtailed government expenditures, we are experiencing dramatic increase in the use of various kinds of administrative data as supplements to, or even replacements

for, traditional forms of official statistics data such as census and surveys. There is every indication that the trend will continue.” In response to this demand, several national statistical agencies have organized joint projects and workshops to address methods for effective incorporation of administrative data in official statistics.

- Representatives from numerous national statistical institutes from the European Social Survey (ESS) countries participated in “ESSnet Statistical Methodology – area ISAD: Integration of Surveys and Administrative Data”, a project dedicated to the dissemination and discussion of statistical methodologies and software tools for integration of survey and administrative data (ESSnet-ISAD 2008a; ESSnet-ISADb; ESSnet-ISADc). Two of the case studies described in the project report highlight the value of administrative data for reducing costs and correcting errors. To obtain estimates related to structural business statistics in Austria, administrative data from tax and social security authorities are used as a substitute for survey data for firms below a specified size threshold. Evaluations indicated that the procedure had potential to lead to good quality estimates of turnover and number of employees. The European Union Statistics on Income and Living Conditions Italian team explored combining tax data with survey data to build imputation models and check for under-reporting in estimating self-employed income. To protect against under-coverage, the maximum of the income reported on the tax record and the survey was used. Due to suspected under-reporting, especially on tax data, combining the two sources increased the estimate of average self-employment income by approximately 11.9 percent.
- Results of the project include three working papers dedicated to statistical and computational methodologies for data integration and case studies describing experiences of national statistical institutes in ESS countries.
- Statistics Canada hosted a symposium in 2010 entitled “Social statistics: the Interplay among censuses, surveys and administrative data,” and Eurostat/INSEE hosted a seminar on the use of administrative sources in statistics in 2004. Examples of applications discussed included combining census and administrative data to create efficient frames for health surveys, improvements to sub-national estimates and validating financial survey data using tax data.
- The Statistical Division of the United Nations Economic Commission for Europe (UNECE) published *Using Administrative and Secondary*

Sources for Official Statistics: A Handbook of Principles and Practices” (United Nations 2011). The handbook is the outcome of an international training course that has been presented to government statisticians from Europe, Asia and North Africa. Although the specific case studies in the handbook are primarily based on experiences in developed countries, the main ideas relating to strengths and weaknesses of administrative data, data linkage techniques and uses of administrative data in conjunction with surveys are informative for developing countries. Section 1 contains content from the handbook, and concepts from the handbook are reviewed in sections 3.1.2-3.1.3.

- The Federal Committee on Statistical Methodology sponsored an interagency committee to examine statistical uses of administrative data. Prell et al. (2009) define four elements of success for data-sharing endeavours: (a) vision and support by agency leadership; (b) narrow but flexible goals; (c) infrastructure; and (d) mutual interest. These were based on a comprehensive review of seven multiagency collaborations involving multiple sources of administrative or survey data. Iwig et al. (2013) propose a framework, structured around several quality dimensions, including relevance, accessibility, coherence, interpretability, accuracy, and institutional environment, for facilitating communication between the agency providing the administrative data and the statistical office.

3.1.2. USES OF ADMINISTRATIVE DATA IN DEVELOPED COUNTRIES

Administrative data have great potential to improve many dimensions of the production of official statistics for agriculture. At one extreme, administrative registers of sufficiently high quality can be used for direct tabulation of agricultural statistics. Even if register systems are not of sufficient quality or completeness to support direct tabulation, administrative data can be used to reduce respondent burden, to provide auxiliary information for model-assisted or model-based estimation methods, and to improve list or area frames and efficiency of sampling designs.

3.1.2.1. Direct Tabulation of Statistical Registers

In some cases, registers or systems of registers are sufficiently reliable to support direct tabulation of official statistics. If the administrative process covers the population of interest and collects data on desired analysis variables, the official statistics can be generated from a register system alone. Statistics Canada has used administrative data directly for statistics related to trade,

unemployment insurance claimants, income distributions, and annual migration (Brackstone 1987). After a test census in 2006, Austria transitioned to a register-based census in 2011 (Berka et al. 2012). Wallgren and Wallgren (2010) discuss experiences of a joint effort by Statistics Sweden and Eurostat to produce statistics on production and livestock from the Integrated Administrative and Control System (IACS). One finding of the study is that reliable estimates can be obtained for crops with area-based subsidies, but that the estimates based on IACS are severely biased for other crops.

3.1.2.2. Reduction of Respondent Burden and Costs of Data Collection

Administrative data can reduce respondent burden and data collection costs by serving as a substitute for survey data. For example, Statistics Canada has used tax data instead of survey data to obtain information on small businesses. It has found this to be practical, owing to the one-to-one relationship between tax filers and businesses and because concepts of interest to the statistical agency are measured through the administrative process (Brackstone 1997). A similar approach to reducing respondent burden is used in the United States Department of Agriculture Conservation Effects Assessment Project. For this survey, data that can be obtained from administrative databases are not collected as part of the farmer interview survey. Because the project requires detailed information from farmers, incorporating administrative records helps ensure that the interview survey is a manageable length. Statistics Canada has used annual tax income data and immigration data during intercensal years (Brackstone 1987). Carfagna and Carfagna (2010) discuss administrative data as a means to obtain additional information on rare domains, which can be difficult or expensive to target with sample surveys.

Because of cost constraints in developing countries, the potential use of administrative data to supplement survey data is an attractive option. This approach is also consistent with the guideline of the Global Strategy (World Bank, FAO and United Nations 2010) to collect information on major crops (core data items) on an annual basis and conduct surveys for specialty items at longer time intervals. Administrative data may provide information on relatively rare items or small domains on a timelier basis.

3.1.2.3. Improving Efficiency of Survey Based Estimators

Administrative data can provide population control totals that can be used to construct more efficient survey estimators through calibration (i.e. Deville et al. 1993). Carfagna and Carfagna (2010) provide an example in which estimated efficiency gains from incorporating administrative controls are larger than

efficiency gains from estimators that use remote sensing data as auxiliary information. In the National Resources Inventory, administrative data on federal and large water areas provide control totals (Nusser and Goebel 1997). The gain in efficiency depends on the correlation between the survey response and the auxiliary variable. If the survey variables are highly correlated with the administrative data, then using administrative data in constructing model-assisted estimators can lead to substantial efficiency gains.

Administrative records can also be used as covariates in constructing model-based small area estimates and forecasts. The United States Census Bureau investigates use of Internal Revenue Service-estimated poverty rates, food-stamp participation rates, and tax exemption rates to estimate child-poverty rates in small areas (Slud 2004). Kantanatha et al. (2010) and Nandram, Berg, and Barboza (2013) use auxiliary data on weather and prices to forecast the crop yield and crop commodity basis using data from the National Agricultural Statistics Service. In a cooperative agreement between Iowa State University and the National Agricultural Statistics Service, administrative data from the Farm Services Agency are used in a similar fashion to obtain estimates of acres planted to different crops (Kim et al. 2013). The National Agricultural Statistics Service also uses administrative data on imports, exports, and stocks to inform estimates and forecasts of production and harvested area.

3.1.2.4. Improving Frame Construction and Sampling Design

Administrative data can be used to maintain or improve an area frame in various ways (Carfagna and Carfagna 2010). One approach is to combine multiple registers to produce a list that completely covers the population of interest. Samples can then be drawn, treating the combined register as a single sampling frame. A second approach is to draw separate samples from multiple administrative lists and use methodology for multiple frame surveys to produce unbiased estimators and standard errors. A third approach, which has been particularly beneficial for agricultural surveys, is to combine administrative lists with area frames. Augmenting an area frame with a list frame can lead to efficiency gains if, for example, the list contains information about large operations or specialty items. One specific example of the multiple frame approach is the United States Census of Agriculture, conducted by the National Agricultural Statistics Service in which a list frame is augmented with an area frame to improve representation of small farms.⁵ The administrative data can be

⁵See www.agcensus.usda.gov/Publications/2012/.

used, in conjunction with remote sensing information, to create more efficient strata than the one obtained from the area frame alone, leading to a better sampling design. Kim, Zhu and Yang (2013) describe successful application of this approach in the context of redesigning the June Area Survey for the National Agricultural Statistics Service.

Many statistical offices use administrative data to maintain list frames. Wallgren and Wallgren (2010) and Wallgren and Wallgren (2011) discuss the value of linking multiple registers to create comprehensive list frames for sampling in Nordic countries. Statistics Canada has used tax records and payroll deduction systems to obtain timely information for maintaining list frames (Brackstone 1987). In the context of agriculture, Carfagna and Carfagna (2010) describe a project to investigate the use of Integrated Administrative and Control System (IACS) data to improve stratification of a list frame based on land cover and land use characteristics. Wallgren and Wallgren (2010) discuss the benefits of using IACS and the business register to inform the farm register. These experiences of developed countries demonstrate the importance of testing a stratification based on administrative data prior to implementation.

3.1.3. METHODOLOGICAL ISSUES WITH ADMINISTRATIVE DATA

Effective use of administrative data requires understanding multiple dimensions of data quality and “fitness for use” in the production of official statistics. Brackstone (1997), Wallgren and Wallgren (2007), Holt (2007), and Iwig et al. (2013) emphasize the need for procedures that check and improve the quality of the administrative data. As administrative data are often collected to meet specific aims of government programmes, rather than to produce estimates of characteristics of a target population, they are not necessarily directly applicable to the objectives of the statistical office. Iwig et al. (2013) provides guidelines for coordination with administrative offices to understand how quantities obtained through administrative processes relate to characteristics of interest to statistical agencies. Methods for record linkage are often required to form rich register systems from multiple administrative databases. Quality control procedures for evaluating measurement error and coverage problems are needed to protect against bias. These last two issues – linkage and measurement – and coverage error are discussed below. Other practical considerations are timeliness of data dissemination and a preserving confidentiality.

3.1.3.1. Quality Control: Measurement and Coverage Error in Developed Countries

Measurement errors in administrative data arise from multiple sources. Conceptual differences often exist between quantities collected through

administrative processes and quantities of analytical interest to a statistical agency. For example, administrative processes entail collecting information on unemployment insurance beneficiaries, a concept that is related to unemployment but differs from definitions of unemployment used by statistical agencies. False reporting can also stem from varying motives in administrative processes. For example, farmers may under-report acres in subsidy programmes to guard against consequences of inadvertent over-reporting (Carfagna and Carfagna 2010). Measurement errors in identifying variables can occur if establishments change but the identifier, such as a street address, does not. Changes in the nature of administrative processes can also lead to changes in collected data over time that makes consistent longitudinal analysis difficult. Zhang and Fosen (2011) examine the sensitivity of small area estimates based on the Norwegian employment register to varying degrees of misclassification error.

Coverage errors occur when the population that participates in the administrative process differs from the population of interest. This can result in both over-coverage and under-coverage. Carfagna and Carfagna (2010) describe studies to examine coverage problems in the IACS data. They conclude that if quality control procedures indicate substantial coverage problems, administrative data should be used only to support sample survey data selected from a frame, such as an area frame, that covers the entire population. Wallgren and Wallgren (2010) compare coverage problems in business and farm registers; they find that differences in coverage properties are related to different forms of coverage error.

The 2012 special issue of *Statistica Neerlandica* on register-based statistics addresses the problem of quality control for administrative data. Zhang (2012) develops a quality paradigm for administrative data that parallels the representation of errors in survey data of (Grove) 2004. Bakker (2012) develops model-based approaches to quantify the bias resulting from measurement error in administrative data. Berka et al. (2012) examine the effectiveness of the Dempster-Shafer theory to quantify uncertainty in each datum in each of several registers used for the register-based Austrian census.

Applications of administrative data to official statistics involve careful procedures to check for and correct measurement errors. In maintaining IACS, an administrative database for agriculture, a sample of declarations is selected on an annual basis and checked for irregularities, such as errors of commission and omission (Carfagna and Carfagna 2010). ESSnet-ISAD (2008b) present a

case study that illustrates the use of decision trees to harmonize definitions of variables related to pensions across multiple sources. Wallgren and Wallgren (2010) recommend combining multiple administrative data sources to improve coverage and check for errors. Iwig et al. (2013) takes a more proactive approach, providing guidelines intended to help statistical offices interact with administrative offices to reconcile definitions, unify objectives and improve the timeliness of data exchanges.

3.1.3.2. Record Linkage from Multiple Administrative Sources

One mechanism for improving coverage and reducing measurement error is to integrate multiple administrative sources to form register systems. This integration process requires linking units across files. Because administrative databases can contain different kinds of units, linkages across databases are not necessarily one-to-one, and procedures are needed for many-to-one or one-to-many matches (Wallgren and Wallgren 2010).

Statistics Sweden makes extensive use of record linkage to form rich register systems based on numerous administrative databases. One mechanism that facilitates use of register systems in Nordic countries is the existence of unique identification numbers for individuals and establishments. This identification system is used for deterministic linking. Statistics Sweden invests resources in register maintenance processes that preserve the integrity and confidentiality of the identification number (Wallgren and Wallgren 2011).

When data sets are linked through probabilistic linkage models or from incomplete linkage, the estimation procedures need to account for linkage error. Methods, such as those proposed by Kim and Chambers (2012), can be deployed. Such methods extend regression analysis techniques of Chambers (2009) to applications with more than two linked data sets.

Certain strategies for combining administrative and survey data also require linking operations. In a multiyear experiment in Italy, administrative records are linked with interview surveys to improve data quality on income. Owing to rules to preserve confidentiality, the Italian tax agency performs record linkage to provide the Italian Statistical Institute (ISTAT) with desired tax information for each sampled individual.

3.1.3.3. Software for Probabilistic Record Linkage in Developed Countries

Administrative data files are often stored in different formats. As a result, they may have different identifying variables or internal errors or inconsistencies,

and a one-to-one match is generally not possible. Probabilistic matching is one technique used to unify disparate data sources. Several software tools for probabilistic record linkage have been developed to meet this end (ESSnet-ISAD 2008c). They include the following:

- a) *AutoMatch*, developed at the United States Bureau of Census, now under the purview of IBM
- b) *Febri - Freely Extensible Biomedical Record Linkage*, developed at the Australian National University (*FEBRL*);
- c) *Generalized Record Linkage System (GRLS)*, developed at Statistics Canada (*Herzog et al. 2007, chap.19*);
- d) *LinkageWiz*, commercial software (*LINKAGEWIZ*);
- e) *RELAIS*, developed at ISTAT (*RELAIS*);
- f) *DataFlux*, commercialized by SAS [*DATAFLUX*];
- g) *The Link King*, commercial software (*LINKKING*);
- h) *Trillium*, commercial software (*TRILLIUM*);
- i) *Link Plus*, developed at the U.S. Centre for Disease Control and Prevention (*CDC*), Cancer Division (*LINKPLUS*).

The EENET (2008c) working paper evaluated the above-mentioned software packages on the basis of several criteria. The comparison criteria included (a) the cost of the software, (b) the flexibility of the software to handle different types of applications and (c) the extent to which the software had been adopted. The software packages were also evaluated in terms of the following capabilities:

- *Preprocessing/standardization* – the ability to convert raw input data to a standard format, reconciling inconsistencies and removing missing data to the degree possible;
- *Profiling* – automated determination of seemingly good matching criteria, recognizing that in many cases, subject-matter expertise determines acceptable matching criteria;

- *Comparison functions* – distance measures (equality, distance), i.e. Jaro, Hamming, Smith-Waterman, term –frequency-inverse document (TF-IDF);
- *Search space reduction* – when comparison of matching variables is impractical, these algorithms reduce the set of pairs to be compared;
- *Decision rule* – the mechanism (i.e. Fellegi and Sunter) for determining the likelihood of a match.

The software tools Automatch and Trillium are commercial. GRLS is commercial and specific to government applications. LinkageWiz and DataFlux are commercial but cost less than US\$ 5 000. RELAIS, FEBRL, The Link King, and Link Plus are free, and a SAS license is required for The Link King.

Automatch, GRLS, and LinkageWiz have tools for English words. FEBRL, RELIAS, and DataFlux are not written for a specific domain. The Link King, Trillium, and Link Plus have features that are relatively generalizable to multiple domains.

RELIAS is judged by the ESSnet-ISAD (2008c) report to have the lowest level of adoption. FEBRL, GRLS, LinkageWiz, and Trillium are judged to have moderate levels of adoption. Automatch and Link Plus have the highest levels of adoption.

3.1.4. DISCUSSION

In transferring this knowledge to developing countries, one can anticipate encountering new challenges, as well as challenges that parallel the experiences of statistical agencies in developed countries. Quality concerns related to coverage and measurement may need to be addressed. In addition, practical issues associated with preserving confidentiality and ensuring timeliness may be important. The guidelines of Iwig et al. (2013) on effective communication with government offices providing administrative data may be of assistance, to some degree, in coordinating with governmental offices in developing countries.

3.2. DEVELOPING COUNTRIES

After reviewing the use of administrative data in developed countries, we turn to past experiences in developing countries before discussing the lessons to be learned.

3.2.1. BACKGROUND

In many developing countries, crop forecasts plus data on crop production and livestock numbers are derived from returns of extension staff or chiefs; these returns are often monthly, quarterly or annual (UBOS 2007). They provide data, especially for the lower administrative levels and mostly for crops or livestock. Agricultural census and survey data are often only available for the major crops and livestock. Furthermore, the data are often only available at national and regional levels. Unfortunately, the data in the periodic reports of extension workers and/or administrative officers are often collected through subjective means, such as eye-estimates or general observations.

In the study of Naik et al. (2012) in the Indian state of Karnataka, a comparison of crop area between methods shows that the conventional method, in general, underestimated crop area and is not appropriate for capturing the changing cropping pattern. This is an enormous concern for a developing country that has an agriculture sector in transition towards commercialization and is adopting high value crops. The paper focused on how agricultural statistics could be strengthened in developing countries using new geospatial tools using the case of rural Karnataka as an example. These new technologies offer the potential to improve measurement by rigorously evaluating the data. Naik et al (2012) identified the major reason behind the poor quality of crop area and land use statistics as being failure of the village accountants to devote adequate time and attention under the conventional method. Okello et al. (2013) identified the same problem with livestock statistics.

The FAO (2002) report on the general status of the system of food and agriculture statistics in the Lao People's Democratic Republic indicated that an FAO-supported pilot sample survey on rice showed an overestimation against existing statistics generated through the hierarchical reporting system. This could partly be the result of having a reporting system that lacked unified definitions or standards. The report found that omissions in the reported data existed, but because of poor communication infrastructure, they were difficult to rectify.

Many major export/cash crops are often covered by one body – a parastatal agency. These agencies often have administratively collected data on marketed or exported amounts. In situations in which the production of a particular crop is managed by parastatals or private sector firms, countries tend to rely on those structures to collect crop production and other related data and feed them into the overall crop estimates made by the central statistics offices or the ministries

of agriculture. In most cases, these estimates are considered reliable. However, the central statistical offices/ministries of agriculture have little say in methods used to collect the data and the increasing use of commodity production or purchase contracts between government and those institutions can provide perverse incentives for the institutions to inflate or deflate their yield or production statistics (e.g. *Office du Niger* in Mali and Food Reserve Agency in Zambia) (Kelly and Donovan 2008).

As far as trade data are concerned, customs records in most developing countries are compiled by the respective customs office and are officially recorded statistics from the border points and ports. It is from those customs returns that exports and imports data, including agricultural exports and imports, are derived. Unfortunately, the informal trade is often not accounted for.

3.2.2. MAIN USES OF ADMINISTRATIVE DATA FOR STATISTICAL PURPOSES IN DEVELOPING COUNTRIES

The agricultural data collected through administrative systems are used by most developing countries in the compilation of gross domestic product (GDP). Consequently, they also serve as the official data on production, area and yield.

3.2.2.1 Administrative Data for Sampling Frame Construction and Sample Design

India and many other developing countries extensively use administrative records and other forms of administrative data for developing the sample frames for a wide range of activities. Among them are records of the following: small, medium plus institutional and large-scale/commercial farms; livestock data, such as slaughterhouse records and vaccinations; agricultural inputs dealers/manufacturers; and exporters and importers. The earliest and perhaps the most important form of administrative record use in Indian statistics is **the land-use data** that are generated on a regular basis by the land revenue administration of the state. These data are compiled from village land records maintained by the village *Patwari* (accountant). The land-use records are central to the whole process of agricultural production estimates for India. They are used as **sample frames in order to determine where crop cutting experiments would take place. These records are also used** as a basic statistical input into the estimation of production, which is derived as a product of the yield given by the crop-cutting experiments and the area under a particular crop as measured by the land-use records (Pronab, undated).

In Malawi, with the introduction of a farm input subsidy, administrative records are being used **to provide a sampling frame to decide on the allocation, beneficiaries and distributions of agricultural inputs.** It is also important to note that the quantity of agricultural inputs that farmers have, is to a large extent already available in administrative records. **Such records could be used to remind farmers about what they had. This reduces errors that results from recall agricultural surveys and the execution of other related survey data collections.** Apart from household identification information, agricultural offices, through officers based at the village, collect information on strategies, challenges and weather parameters of the area, which is crucial information for designing national initiatives (Pangapanga et al. 2013).

In Ethiopia, lists of *kebeles* from administrative offices were used as a sampling frame (Abaye 2010).

3.2.2.2. Crop Forecasting

The administrative data collected on the different aspects of weather are used in a number of countries to forecast crop yield and/or production for food security purposes.

In India, the National Crop Forecasting Centre uses such information as weather conditions, supply of agricultural inputs, pests and diseases in formulating scientific and objective forecasting methods (Srivastava, undated). In Mali, the *Systeme d'Alerte Précoce* (SAP)/Early Warning System/Unit collects and analyses information on such things as crop forecasts, satellite imagery, price trends and potential threats due to climate or pests in order to provide early warnings of impending food crises and make recommendations for actions to ameliorate the situation. SAP conducts on-the-ground assessments of potential food security hot spots, with a focus on high-risk areas in northern Mali. While the Malian SAP exhibits some characteristics of an effective early warning systems as noted in the FAO study on food security early warning systems in sub-Saharan Africa (Tefft et al. 2006), it is not immune from some of the weaknesses noted in the same FAO report, namely political pressure when recommendations do not conform to politicians' views and heavy reliance on data and information collected by donors or donor-funded activities.

The Zambian National Early Warning Unit oversees the crop and livestock monitoring done by the extension system and synthesizes the results to establish a preliminary forecast. This forecast appears in the food balance sheets until the Central Statistics Office survey-based numbers from a crop forecasting survey and later a post-harvest survey the Crop Forecasting Survey are available (Kelly

and Donovan 2008). A food balance sheet is the most widely circulated use of agricultural statistics in the country. It is used extensively by policy-makers when considering subsidies, import/export regulations and key food security issues.

The Directorate of Marketing and Inspection in the Ministry for Agriculture in India established the AGMARKNET portal for collecting and disseminating information on prices and arrival of all important commodities from different regulated agricultural markets across the country. The network links the country's agricultural produce markets with the state agricultural marketing boards to facilitate the dissemination of key information to farmers and other users. In addition, a **regular system is in place to record prices of agricultural commodity at the market yards and arrivals of agricultural produce**. This information is central to the existing system of recording agriculture output and consumption in value terms for the country (Srivastava, undated).

3.2.2.3 Integration of Administrative Data in Agricultural Surveys

Administrative records can be used to cover data gaps. Since the last agricultural survey conducted in Malawi was in fiscal year 2006/07, the ministry and government at large has been developing programmes based on administrative data (Malawi 2012b). Moreover, there are several ad hoc programmes that originate and require information from the Ministry of Agriculture. Administrative processes are used to obtain this information, as they can supply it in a timely manner as compared with agricultural surveys. It should, however, be noted that administrative data can be integrated in agricultural surveys only when: data instruments are designed through stakeholder consultations; an information needs assessment has been conducted; and the scope, quality and data items to be collected are properly defined and agreed (Pangapanga et al. 2013).

3.2.2.4 Administrative Data for Planning

Resources to collect administrative data are predictable and usually met by the administrative process itself. In brief, administrative data help government machinery for planning purposes and are available annually as required (Pangapanga et al. 2013). Through the national statistical system, the National Statistical Office Malawi is now promoting the use of administrative data for statistical purposes among national statistical system partners (Malawi 2012b).

3.2.2.5. Small Area Estimates

Owing to sample size constraints, the results from agricultural annual sample surveys are not likely to be available at lower levels, such as at the districts level. Ancillary information from the regular agricultural returns can be used for scaling down the highest level estimates from sample surveys to district level estimates, using small area estimation techniques. These techniques are being used in several countries. In Africa, one such application has been made in Ethiopia. The agricultural annual surveys conducted by the Ethiopian Central Statistical Agency had been providing crop-wise area estimates at regional and zone levels, but district-level estimates were not available. Meanwhile, the Ministry of Agriculture and Rural Development was generating area estimates through an approach that was very similar to the regular agricultural returns. The small area estimation approach was used to gather district-level estimates for crop areas from annual surveys, using data from the Ministry as an auxiliary variable, (National Bureau of Statistics, Tanzania 2012).

3.2.2.6. Non-probabilistic Procedures of Integrating Use of Administrative Data Sources

Sample surveys and censuses that use rigorous statistical methods remain the most reliable source of agricultural data. However, in many developing countries, a large portion of current data being produced and disseminated through national, regional or global databases, or publications are from a variety of sources, owing to the absence of regular statistical surveys or censuses conducted by countries (Keita and Chin 2013). Country assessment and in-country presentations at regular meetings often cite budgetary constraints as a factor behind not using frequent agricultural surveys. To fill the data gaps resulting from this challenge, there is an increase in the use of reporting systems in which extension staff in the field or other collaborators collect and compile agricultural statistics on, for example, crop area and yields, livestock and prices, and channel compiled data through the administrative hierarchy. Various non-probabilistic methods are used for data collection, including use of administrative sources, eye estimates by local informants, expert opinion or assessment, windscreen surveys, rapid appraisal methods and community/village. Among the limitations of these methods are lack of representativeness, high levels of subjectivity and inability to calculate a sampling error or confidence interval (Keita and Chin 2013; Galmés 2013).

Galmés (2013) also noted that although sample surveys are considered as the main source of agricultural and rural data given that they are “objective”, “scientific” and allow for the computation of sampling errors, complementary

use of other sources needs to be integrated in the sampling estimates, similar to what is being done in the global strategy. This is because, despite the availability of resources for implementing sampling surveys, they will neither be able to give accurate estimates for small administrative areas nor the required periodicity of less than one semester or so. In addition, non-sampling errors can affect the estimates and their assessment is impossible without external information.

As noted by many experts, even though non-scientific methods of data collection have limitations, they still play an important role in agricultural statistics systems, especially in developing countries because of the following reasons (Keita and Chin 2013):

- a) Limited statistical capacity and high cost of sample surveys and censuses: Most developing countries lack the statistical capacity to handle rigorous scientific statistical methods of surveys and censuses, or the financial capability to meet the high costs of the field components of surveys or censuses. They are dependent on donor expertise and funding.
- b) Need for regular, timely data: Although countries are increasingly demanding more frequent data on a yearly or even sub-annual basis, census and surveys in most of those countries are at best collected every few years, resulting in data gaps in the non-survey/census years. Administrative data systems play an important role in filling this data gap.
- c) Need for small area data: Surveys do not provide data at the lowest administrative areas and yet with decentralization, most countries are increasingly requiring data from local communities for development of local government plans and monitoring and evaluation systems
- d) Need for timely data for early warning data: In situations requiring rapid response, surveys and censuses may not provide timely data. Alternative non-probabilistic methods, though not accurate, may provide timely information for a quick response to an emergency.
- e) In many countries, statistical surveys cannot be used for obtaining early warning and pre-harvest crop forecasting data to inform governments, marketing boards and other users on the prospective food and crop situation and harvest because they are not produced

quickly enough to be available for use to provide early warnings of impending problems. In those situations, rapid assessment methods are the solution.

- f) Need for in-depth data: Surveys provide quantitative facts on situations. However, the reason behind the facts are qualitative in nature and are best collected using qualitative methods, such as focus group discussions, key informants interviews and institutional reviews, which can complement the survey data.

According to Keita and Chin (2013), statisticians need to analyse critically the alternative methods and their limitations in order to develop ways to improve (not promote) and make best use of the information generated to fill statistical gaps or provide timely information for decision making. Delorme (1982) proposes six principles for improving assessments:

1. Confine assessments to the smallest possible geographic area. The zone should be small enough so that the informant has direct knowledge of what is being estimated and reported on and can observe it.
2. Train assessors in the basic principles of enumeration, common definitions and concepts, visualization of units of measurement and standard frames of reference for the measures, and use of a systematic method of recording information
3. Assess variations and not absolute values as these are easier to perceive. Base year values are calculated from periodic sample surveys. This principle is particularly relevant to and borne out by experience in, integrated agricultural statistics systems in which benchmarks are provided by statistical surveys and censuses. Yearly/sub-yearly variations are often derived from an assessment. The principles primarily address the issue of technical errors, which can be reduced with appropriate training and development. Systematic errors remain more difficult to address, and although triangulation and cross checking of variables are useful, in practice, the data sources may be limited and simply not available. Discussions of results among experts, however, remain valid and are a step forward.
4. Search systematically for data that strengthen or support the assessments to triangulate the results on the same variable.

5. Look for connections between the quantities estimated. These relationships between variables provide a useful check.
6. Harmonize and clearly define concepts and nomenclatures. This may be more difficult in practice as systems serve different purposes and are run by different agencies. Thus, harmonization involves some degree of compromise in the concepts and definitions.

4

Issues in the Use of Administrative Data in Developing Countries

Apart from being a very rich source of agricultural data, as stated earlier, administrative data sources are generally cheaper and easier to sustain, but they tend to have a number of issues associated with them, especially in developing countries. The major ones are discussed in this section.

4.1. BACKGROUND

Several authors have pointed out that most data are collected and compiled primarily for internal use by personnel who have no training in statistical methods and usually without using standard statistical methods or procedures. It is believed that much of those data are guess estimates or a reflection of set targets rather than actual levels of indicators. Some of them are believed to be of a questionable quality (UBOS 2007; Kiregyera et al. 2007). There are several challenges in obtaining the data, including the methodology and equipment and staff used. For example, Uganda is still facing challenges of incomplete administrative data. The issues reported include the following: (a) the extension staff are overloaded with work, apart from data collection, and tend not to be properly trained in agricultural statistics collection; (b) administrators sometimes want to be paid before collecting and making reports; and (c) data collection has been greatly affected by decentralization, which has resulted in undermining allegiance by district staff to their respective ministry of agriculture and results in activities being carried out on an ad hoc basis, without concern of punitive action from the ministry as it is not their employer (UBOS 2007).

4.2 METHODOLOGY FOR OBTAINING THE DATA

The methodology used does not yield reliable estimates. In Uganda, for example, estimates made by the Department of Agriculture were guess estimates extracted from annual and monthly reports compiled by district agricultural officers. The information collected was on such things, as area, yield, production, prices and marketing of the main food and cash crops, namely cotton, coffee, tobacco, tea, sugar, cocoa, citrus, plantains, sweet potatoes, Irish potatoes, cassava, finger millet, sorghum, maize, wheat, rice, field peas, pigeon peas, cow peace, beans, groundnuts, simsim (sesame), castor and vegetables. Two methods were used in the estimation of annual crop areas (UBOS 2007):

- a) *Buganda Method* – randomly selected villages were supposedly completely enumerated (*Mitala Survey*) in respect of areas of all important crops. The area under each crop obtained by pacing and/or eye-estimation or pure guess work was aggregated and then divided by a "refined" number of taxpayers belonging to the sample villages. The average derived was then multiplied by the total number of taxpayers in the entire district to get the estimated area under the crop in the district.
- b) "*Outside Buganda Method*" – returns of plots counts were carried out by chiefs and compiled for the two major seasons in the year. Those plot counts were aggregated and multiplied by a general plot mean size, supposedly derived from pacing by the extension staff to obtain district crop area totals.

A variation of this, which was much more widely applied than the *Buganda Method* was to have both plot counts and mean plot sizes obtained by the extension staff rather than chiefs. Production was then estimated as the product of area and yield. The yield estimates were always arrived at subjectively by the respective district agriculture officer with the help of staff members led by an assistant agricultural officer. These systems unfortunately broke down in the late 1970s. Consequently, most of the data are made up at the district headquarters by the extension staff without any consultation with the chiefs (UBOS 2007).

In Mozambique, some provinces maintain a frame of livestock producers (*arrolamento*) with a cattle headcount that is updated periodically as an administrative data source at the district level (INE and MINAG 2011). It is reported that the data from the *arrolamento* are used by the Directorate of

Veterinary Services in some districts to collect livestock data, which the district uses for programming and operations, such as disease control. However, there is a concern that the *arrolamento* where livestock counts are conducted at the district level to provide annual headcount data for cattle are unable to give accurate livestock numbers because cattle dips were privatized and are not used by all cattle keepers.

4.3 DECENTRALIZATION AND LEVELS OF INFORMATION MANAGEMENT

Following decentralization, in several countries, the primary responsibility for collection of statistics has been passed to local governments, creating a number of levels at which data are managed, namely the central ministries, provinces/regions, district/municipality, sub-county/town council and village levels. At each of those levels, there is delayed information flow, limited capacity and data are only collected on an intermittent basis. This subsequently constrains programme monitoring by the local governments because of inadequate understanding of their own data (UBOS 2007). Unfortunately, with the devolution of planning functions to the lower levels, there is need for disaggregated data at these levels and the lack of it means that there is a critical need that is not being met (Kiregyera et al. 2007). To make matters worse, there are several data items. It is, however, not feasible to collect those data through censuses or surveys.

In Uganda, there is currently no follow-up on those extension staff who do not file reports on data. In fact, reports filed are often not reviewed by the district agricultural officers nor by any officer at the district headquarters. This is clearly the result of having no specific officer at the district level to review and collate the information in the reports. In the past, an assistant agricultural officer was in charge of statistics at the district. These officers generally had some basic training in statistics. Indeed, most of them held a certificate or even a diploma from the Eastern Africa Statistics Training Center in Dar es Salaam. Their responsibilities included, among other duties, the collection, analysis and reporting of agricultural statistics (UBOS 2007).

4.4 OPERATIONAL CONSTRAINTS

In the case of reporting on crops in Mozambique, operational constraints make it difficult for extension staff to go around their areas of jurisdiction to collect data. They have several other responsibilities and are not legally under the statistics authorities. In addition, much of the data from this source remain in raw form and are not turned into usable information for management (INE and

MINAG 2011). In Uganda, the extension staff are also reported to be overloaded with work, apart from data collection, and the chiefs sometimes want to be paid before collecting and making reports (UBOS 2007). In India, considering the *patwari* or village accountant's work load, his/their potential to collect the crop information can be stretched at most to half of the total allocated area. Moreover, the one-month time period that is allocated to complete the data collection process each season is insufficient (Naik et al. 2012).

4.5 DIFFERENCE BETWEEN ADMINISTRATIVE PURPOSES OF THE DATA AND STATISTICAL REQUIREMENTS

According to Pronab (source undated), problematic issues stemming from more extensive use of administrative records in the Indian context include:

- Divergence between the nature of the data required for administrative purposes, especially when the objective is to monitor programmes, and the nature of the data that would be required for statistical purposes.
- The data collection machinery is generally under the administrative control of the programme authorities rather than the statistical authority, which often results in the absence of appropriate questions and indicators due to the need to keep the record-keeping process manageable. In recognition of this issue, the Government of India has raised the status of the statistical officers in the line ministries significantly, with the aim to give them a more prominent voice in the process to design the administrative data collection system. A similar effort has also been placed on state governments to give more emphasis on statistical components of administrative records.
- The accuracy of the data comes under scrutiny. It has been found that in situations in which the data are collected for monitoring programmes, the quality of the data becomes highly questionable. This is a particular problem in the social sectors and for data collected by tax authorities. On the other hand, when the purpose is mainly for regulatory oversight, the quality of data tends to be high.

Most developing countries are still facing challenges of incomplete administrative data as a result of a number of factors, as stated earlier.

4.6. CROP FORECASTS

Regarding crop forecasts, statistics in the countries reviewed show that in each country there has been an overlapping of efforts, with the ministries of agriculture tending to use non-sampling methods to collect information for the early crop forecasting, while the central statistics offices mostly using statistical sampling methods to develop definitive estimates of the total production for the national accounts. Information on areas planted and anticipated harvests – both are needed for making crop forecasts – is collected using a variety of methods. Furthermore, it has been noted that in general, the ministries of agriculture and the central statistics office have no legal authority to impose any particular methodology on the offices reporting those production statistics.

As with Mozambique, there are problems with the linkages between the MACO preliminary crop forecast from monitoring and the Central Statistics Office Crop Forecasting Survey (INE and MINAG 2011). The diverse sources of data, as well as the use of different assumptions in modelling weather data often contributes to conflicting crop forecast results (e.g. Zambia in 1998 and 2005 and Mali in 2005). Resolving the problem of conflicting crop forecasts and food security assessments requires greater cooperation among the various actors, in order to be able to synthesize and reconcile differences in estimates obtained from different approaches. There is need to develop better methods of combining data from the various sources of forecasting data in terms of timing of the forecasts, use of statistical methods versus non-statistical sampling and use of different enumeration areas (Kelly and Donovan 2012).

Crop forecasts continue to be in the spotlight of controversy in many developing countries. This may be attributed to the combination of undocumented methods, reliance on key informants for early forecasts, and possible interference of politicians. There is no clear answer on how to combine statistical and non-statistical data. Advanced remote sensing analysis may assist in getting early crop forecasts that are more reliable. Clearer messages on the reliability of qualitative estimates may help in enabling the early forecasts to be used more appropriately as indicators rather than firm production numbers. There is a clear challenge in having good early estimates for food security analysis, yet retaining flexibility to respond to updated information (Kelly and Donovan 2008).

4.7. LIVESTOCK

Despite governments' and other regional and international institutions' wide-ranging use of routine livestock data, administrative records are often incomplete, out-of-date and unreliable. Insufficient resources and limited skills in data-handling and processing are the two most-cited reasons for the inadequate insufficient administrative records. In addition, as in the cases of Uganda and Tanzania, there is no unique format used throughout the country as the data are primarily collected to meet the data needs of the district authorities. However the Improved Agricultural Routine Data System has been developed to come up with a district-level integrated data collection format and a data management software LGMD-2 (Local Government Management Database). This is expected to transmit/generate a harmonized database for agriculture (National Bureau of Statistics, Tanzania 2012).

A paper by Okello et al. (2013) presents a methodology for rapid assessment of routine livestock data systems and identifies options for improvement. It has been developed by the Uganda Ministry of Agriculture, Animal Industry and Fisheries and the Uganda Bureau of Statistics (UBOS), in collaboration with the FAO-World Bank-ILRI-IBAR Livestock Data Innovation in Africa Project. The methodology to assess the routine livestock data system was applied to Uganda. It was concluded that the current system of routine livestock data collection in the country is inadequate, owing to missing information and the poor quality of the data. The paper proposes to streamline the current livestock-data reporting form by suggesting that the Uganda Ministry of Agriculture, Animal Industry and Fisheries requests district authorities to report on different items on a monthly, quarterly and annual basis. It then sketches four possible pilots to identify the first best institutional reform for an improved system of routine livestock data collection. The adoption of improved monthly, quarterly and annual livestock statistical reports – which is a no-cost action – is on its own expected to enhance the quality of routine livestock data.

4.8. IMPORT AND EXPORT

Apart from customs points, the informal trade data are covered under informal cross-border trade surveys in a number of countries. These surveys are carried out jointly by the central statistics offices, revenue authorities and the central banks. The data are then combined to derive the final trade data. The integration of formal and informal trade data are examples of integrating administrative and survey data. In relation to customs data, countries that have not put in place a system of collecting informal cross border statistics, customs data are problematic as there is a general sense that they do not reflect:

- Actual amounts that cross official borders due to negotiations to reduce taxes paid;
- Goods crossing informally where there is no customs post or in small quantities that are not captured by customs records (Uganda 2011; Uganda 2012).

4.9. FISHERIES

In Mozambique, most of the fisheries statistics are collected using the administrative system. However, the administrative methods used to collect the information need to be reviewed and harmonized. The other weakness identified is the lack of an integrated database. The information collected by the institutions is sent to the National Directorate of Economic Policy and Planning on sheets in excel format or row data. As the Directorate does not have a database that links to others databases, the information received through those sheets are again punched manually to excel sheets, resulting in a loss of information, which makes it difficult to disseminate (Kelly and Donovan 2012).

Fisheries management within the territorial waters in India rests entirely on the state governments. Information on fish catch/production from these waters and those that are brought onshore and within the state, disposal of fish-catch, preserved and processed items and aquaculture comes from the state. The constraints with fisheries statistics include:

- Delays in compiling and submitting fishery statistics. The flow of data to compile the all-India statistics is not rapid enough;
- Incomplete and incorrect or unreliable data;
- Incompatibility of data mainly because of variations in concepts, definitions and compilation methodology among regions;
- Poor accessibility;
- Lack of skilled manpower;
- Lack of commitment;
- Lack of standardized/compatible data collection and compilation methodologies and tools (CSO India 2011).

4.10. IMPROVING USABILITY OF ADMINISTRATIVE DATA

The traditional scientific sources of information for planning by governments, namely surveys and censuses, contribute a small percentage of the data required

for planning. According to the Strategic Plan (2010/11-2014/15) for Statistics South Africa, “Statistics South Africa produces less than 10 percent of the statistics required to inform national development while, the rest of the data are administrative data produced by other organs of (the) State”. Dr Sean Philips, the Director General of the Department of Performance Monitoring and Evaluation in the Presidency of South Africa, stated that "government performance cannot improve until the administrative data are improved". This underscores the importance that other data, especially administrative data, play in planning. As discussed earlier, administrative data are limited in terms of quality and usability for statistical purposes. There is need to improve the data to the point that they are reliable enough to contribute credible and reliable information to the planning and decision- making processes.

In the context of India, given the size and diversity of the country and the limitations in the reach of the government, administrative records tend to be incomplete. The classic case of this is the coverage of the civil registration system, which is so low that it cannot be used for measuring demographic parameters between the census years. Many of these problems can be tackled through cross checks and corrections made through survey data. For instance, in the case of civil registration, India operates a sample registration system, which provides reasonable estimates of demographic indicators for the inter-census period. Similarly, surveys carried out by the National Statistical Survey Organisation also provide important cross checks on a variety of statistical indicators (Pronab, undated).

According for Pronab (source undated), the best system would involve the use of limited surveys based on strong statistical principles to provide validation and corrective factors for the data generated on a regular basis through administrative accounts. This is similar to implementing a sample audit system, in which the purpose would not be to find fault, but to provide information that can be used to correct the inherent biases that may occur in administrative record keeping.

In the final analysis, however, the main factors governing the usability of administrative data for statistical purposes are the legal framework underpinning the data collection activity and the political importance attached to the government interventions concerned and of course governance.

Replacing a statistical survey with data from administrative registers is a difficult and important decision. This is especially applicable to official statistics in which comparability over time is an important component of quality. Therefore, when Sweden became a member of the European Union in 1995, and administrative registers in the agricultural area were created, studies were carried out to investigate the consequences of using administrative registers for statistical purposes (Wallgren and Wallgren 1999).

The studies showed that integrating administrative registers with censuses and sample-surveys could be cost-effective ways to produce statistics with sufficient quality. The integration phase in which data from several sources were integrated into a new statistical register was seen as essential for achieving sufficient quality. As a result of the studies, Swedish official statistics in the agricultural area are now based on extensive use of administrative data. This includes for example statistics on how agricultural land is used, the number of cattle, and the income of agricultural households. In later years, the work of Wallgren and Wallgren (2007) and Bakker (2011) have been used to increase knowledge on how to build statistics based on administrative data.

In recent years, several studies, for example Laitila et al. (2011); Daas et al. (2008); Daas et al. (2010); and Berka et al. (2012), have discussed quality frameworks for using administrative registers for statistical purposes. This includes the quality of the register itself, the possibilities of integrating administrative registers into statistical registers and how to document the quality of the statistics produced. Holmberg (2012) concludes that there has been a theoretical improvement on how to assess the quality of administrative data.

The indicators are also grouped according to the quality declaration for official statistics in Sweden, namely relevance, accuracy, timeliness, comparability/coherence and availability.

Quality in a broader sense as, for example, discussed by Eurostat (2007) includes requirements of user needs, institutional preconditions at the organization producing the statistics and the production process itself. Furthermore, evaluation of options may differ if the main goal is to reduce response burden as opposed to it being to reduce costs. It is important to have a phase in the decision-making process that is open-minded and also explores other alternatives, for example, solely modelling or discussing with the users if they would be willing to change their use of the statistics.

The problems of quality in the implementation of integrating the different sources can be reduced to considerably by using the agricultural censuses and surveys as quality checks. For example, in Mozambique, Michael Trant proposed a system of adjusting the regularly collected data based on the census of agriculture data (Trant 2012). This, however, presupposes that the censuses and surveys continue to be undertaken regularly and that a systematic approach is in place for evaluating the administrative registers themselves.

Ballivian et al. (undated) emphasize the need to assess the quality of administrative registers before using their data for statistical purposes. They document use of a World Bank-supported standardized tool for quality assessment of administrative registers, which was used in the Mexican state Yucatan. The tool was developed with quantitative indicators focusing on four quality aspects, namely administrative data source, metadata, data and statistical product. The last element determines the quality of statistics, tables and microdata files produced from administrative registers, either as a frame of sample surveys, a secondary information source or as generating statistics directly by processing data of the statistical register. The results from using the tool are used to identify improvement areas and opportunities, problems and their root causes, and then to implement corrective actions to eliminate those causes. This is the basis of the improvement plan of administrative registers that is currently being implemented by the State Secretary of Planning and Budget of Yucatan in Mexico. The attributes for each of the four quality aspects along with their indicators are shown in Appendix II.

4.11. LINKAGES AS INTEGRATING FACTORS

An integrated framework for agricultural statistics appears to be the best strategy for supporting a sustainable and user-oriented agricultural system. Specifically, this framework should consist of a number of integrated components, which together would support the entire agricultural data production process. This integrated approach needs to be coordinated and prioritized in order to prevent duplication of efforts, which often leads to inconsistent data. The integrating factors should lead to the establishment of linkages, such as (a) incorporating a common set of core items, (b) consistent application of common and compatible concepts, definitions and classifications, and (c) possibly use of a common field organization. As this framework should provide continuity, it must facilitate the accumulation of experience, which can then be used to improve the quality of data in subsequent rounds of data collection.

4.12. POTENTIAL SOURCES OF ADMINISTRATIVE DATA

In the wake of structural adjustment and privatization programmes in several developing countries, functions that were originally carried out directly by government have now been passed on to the private sector or at least to parastatal bodies. Some examples are the regulation of major cash crops production, marketing and export and the production, import or distribution of agricultural inputs. This means that administrative data produced by those agencies are not directly under the government.

A number of those systems need to be reviewed to determine their potential for providing data. This includes the parastatals and various associations. In situations in which the production of a particular crop is managed by parastatals or private sector firms, countries tend to rely on those structures to collect crop production and other related data and feed it into the overall crop estimates made by the central statistics office or the ministry of agriculture. In most cases, those estimates are considered reliable, but the country statistics office/ministry of agriculture have minimal say in methods used to collect the data and the increasing use of commodity production or purchase contracts between government and those institutions can provide perverse incentives for the institutions to inflate or deflate their yield or production statistics (e.g. *Office du Niger* in Mali and Food Reserve Agency in Zambia). Notably, many countries are revising their statistics law to ensure that the central statistics office regulates the whole national statistics system. This will make it easier to bring many of those agencies under the national statistical office's statistical control.

5

Conclusions

Literature review has begun to identify best practices and lessons learned that are adaptable for use in developing countries. There has also been a review of literature on administrative sources of agricultural statistics from developing countries in order to identify problems and gaps and new potential sources of data, especially from the private sector, that are available. This is partly due to the privatization of formerly official functions and the growth of the private sector.

A major weakness given the recent push towards crop and income diversification in Africa is the availability of accurate data on production, consumption and trade of horticultural crops, fish and livestock. Weakness in those data has serious implications for estimates of food security.

The next task (Task 2) is to review and analyse relevant country experiences and ongoing research in developed countries (including in Europe, where important research is being carried out) on the use of administrative sources for producing agricultural data and lessons for developing countries. It is also necessary to analyse the results of country assessments and other relevant documentation on administrative sources currently being used by developing countries (Task 3), and evaluate their strengths, weaknesses and suitability for use in agricultural statistics within an integrated and cost-effective agricultural statistics system.

The literature review clearly shows that statistics agencies in developed countries have conducted much research in this area, especially on the several uses of administrative records. Administrative data serve multiple purposes in an integrated national statistical system. On the other hand, the literature review on developing countries indicates that despite questionable administrative data quality, many countries are already collecting and using administrative data in many situations. In fact, administrative data are the major source of data in

many countries, especially for agricultural statistics. The major problems are mostly data quality issues. This is where lessons and ongoing research in developed countries will be used in Task 2 and Task 3. There is, however, concern that the best practices in developed countries may be impractical to implement in developing countries because of infrastructural differences.

References

Abaye, A. 2010. The 2003 Pastoral Areas Livestock Census and the Data Needs in the National Strategy for the Development of Statistics (NSDS) in Ethiopia, A paper presented at the Fifth International Conference on Agricultural Statistics, Kampala, Uganda, 12-15 October 2010. Available at: www.fao.org/fileadmin/templates/ess/documents/meetings_and_workshops/ICAS5/PDF/ICASV_1.1_008_Paper_Aberash.pdf.

African Development Bank (AfDB), Partnership in Statistics for Development in the 21st Century (PARIS21) & Intersect 2007, *Mainstreaming Sectoral Statistical System: A Guide to Planning a Coordinated National Statistical System*, Tunis and Paris.

AFRISTAT. 2014. *Observatoire économique et statistique d'Afrique subsaharienne*. Available at : www.afristat.org.

Andersen E. & Utne H. 2007. *The First Register-Based Census in Norway in 2011: How to Comply with International Recommendations*. Seminar on Registers in Statistics – methodology and quality, 21-23 May 2007, Helsinki, Finland.

Asian Development Bank. 2010. *Administrative Data Sources for Compiling the Millennium Development Goals and Related Indicators*. A reference handbook on using data from education, health, and vital registrations systems featuring practices and experiences from selected countries., *MDGs and related indicators*, Manila.

Asian Development Bank & Partnership in Statistics for Development in the 21st Century. 2002. Report of the ADB/PARIS21 *Joint High Level Forum on Statistical Capacity Building for Asian Countries*, 7-9 November 2002, Manila. Available at <http://www.paris21.org/sites/default/files/341.pdf>.

Budak D. & Volk T. 2009. *Agriculture Statistics Database — Survey on Data and Methodology in Turkey*, Working paper 4; Statistics in the Western Balkan and Turkey.

Humberto Silva, A., Morais, M. & Dias, D. 2013. Forecasting Municipal Crops in Brazil – A Comparison of Alternatives Data Sources for Better Accuracy of Results Brazilian Institute of Geography and Statistics. Paper

presented at the Sixth International Conference on Agricultural Statistics, 23-25 October 2013.

Bakker, B. 2011. Micro-Integration: State of the art. In *Report on WP1: State-of-the-art on Statistical Methodologies for Data Integration*, ESSNET on Data Integration. Available at: www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2010/wp.10.e.pdf.

_____ 2012. Estimating the Validity of Administrative Variables. *Statistica Neerlandica*, 66 (1): 8-17.

Ballivian, A., John-Abraham, I., Medina, A., Ortegon, J., Romero, I., and Segui, F. Quality improvement of administrative registers statistically exploited to generate the indicator-based decision-making system in the State of Yucatan Mexico. Available at: <http://www.statistics.gov.hk/wsc/IPS071-P2-S.pdf>.

Berg, E & Fuller, W. 2014. Small Area Prediction for the Canadian Labour Force Survey. *Journal of Survey Statistics and Methodology*, 2(3): 227-256.

Berka, C., Humer, S. and Moser, M. 2011. Combination of Evidence from Multiple Administrative Data Sources: Quality Assessment of the Austrian Register-based Census 2011. *Statistica Neerlandica*, 66(1): 18-33.

Berka, C., Humer, S., Moser, M., Lenk, M., Rechta, H. & Schwerer, E. 2012. Combination of Evidence from Multiple Administrative Data Sources: Quality Assessment of the Austrian Register-Based Census 2011, *Statistica Neerlandica*, 66(1): 18–33.

Bethlehem, J. 1988. Reduction in Nonresponse Bias through Regression Estimation, *Journal of Official Statistics*, 4(3): 251-260.

Bhoomi. (2012). *Pahani (RTC)*. Retrieved from Bhoomi: [http://bhoomi.karnataka.gov.in/faq.htm#What is Pahani\(RTC\)?](http://bhoomi.karnataka.gov.in/faq.htm#What is Pahani(RTC)?)

Bisa Banda, B. 2011. Malawi: Panorama Report II. Prepared for *Country STAT FOR SUB-SAHARAN AFRICA*.

Brackfield, D. 2007. Administrative Data Framework. OECD Publication: Paris. Available at: <http://www.oecd.org/std/fin-stats/38486345.pdf>.

Brackstone, G. 1987. Statistical Issues of Administrative Data: Issues and Challenges. *Survey Methodology*, 13(1): 29-43.

Carfagna, E. & Carfagna, A. 2010. Alternative sampling frames and administrative data; which is the best data source for agricultural statistics? In Benedetti R., Bee B., Espa G. and Piersimoni, F. (eds) *Agricultural Survey Methods*. Chichester, United Kingdom, John Wiley & Sons.

Chambers, R. 2009. Regression Analysis of Probability-linked Data. *Official Statistics Research Series* (4).

CSA. Undated. www.csa-mali.org.

Daas, P., Arends-Tóth, J., Schouten, B. & Kuijvenhoven, L. 2008. Quality framework for the evaluation of administrative Data. Proceedings of Q2008 European Conference on Quality in Official Statistics, Statistics Italy and Eurostat, Rome.

Daas, P. Ossen, S. & Tennekes, M. 2010. *Determination of Administrative Data Quality: Recent results and new developments*. Proceedings of Q2010 European Conference on Quality in Official Statistics, Statistics Finland and Eurostat, Helsinki, Finland.

Delorme, P. 1982. *Estimating crop area and yield in agricultural statistics*, ESD Paper 22, FAO

Eurostat 2007. Handbook on Data Quality Assessment Methods and Tools Manfred Ehling and Thomas Körner (eds), Wiesbaden

Fellegi, I.P., and Sunter, A.B. (1969). A theory for record linkage. *Journal of the American Statistical Association*, 64, 1183-1210.

Holmberg, A. 2012. Discussion on Assessing Quality of Administrative Data. *Statistica Neerlandica*, 66(1): 34–40.

Deaton, A. & Kozel, V. 2005. Data and Dogma: The Great Indian Poverty Debate. *World Bank Research Observer*, 20(2):177-199.

Deville, J., Sarndal, C. & Sautory, O. 1993. Generalized Raking Procedures in Survey Sampling. *Journal of the American Statistical Association*, 88(423): 1013-1020.

Dion, M. 2007. *Metadata: an integral part of Statistics Canada's data quality framework*. Paper prepared for the *Fourth International Conference on Agricultural Statistics, (ICAS-IV)*, 22-24 October 2007, Beijing.

Dion, M., Chartrand D. & Murray, P. 2010. Statistics Canada's quality assurance framework applied to agricultural statistics. In Benedetti R., Bee B., Espa G. and Piersimoni, F., (eds) *Agricultural Survey Methods*. Chichester, United Kingdom, John Wiley & Sons.

Durrant, G. & Skinner, C. 2006. Using Missing Data Methods to Correct for Measurement Error in a Distribution Function. *Survey Methodology*, 32 (1), pp. 25-36.

Elliot, M. & Davis, W. 2005. Obtaining Cancer Risk Factor Prevalence Estimates in Small Areas: Combining Data from Two Surveys. *Applied Statistics* 54(3): 595-609.

ESSnet, Statistical Methodology Project on Integration of Survey and Administrative Data (ESSnet-ISAD eds). 2008a. *Report on WP1: State-of-the-art on Statistical Methodologies for Data Integration*. Available at: www.istat.it/it/files/2013/12/FinalReport_WP1.pdf.

_____ 2008b. *Report of WP2: Recommendations on the Use of Methodologies for the Integration of Surveys and Administrative Data*.

_____ 2008c. *Report of WP3. Software Tools for Integration Methodologies*.

Food and Agriculture Organization of the United Nations (FAO). 1997. *Improving Agriculture Extension, Reference Manual*. FAO: Rome.

_____ 2002. General Status of the System of Food and Agriculture Statistics in LAO PDR, Food And Agriculture Organization Of The United Nations: Bangkok.

Fay, R. & Herriot, R. 1979. Estimation of Income from Small Places: An Application of James-Stein Procedures to Census Data. *Journal of the American Statistical Association*, 74(366): 269-277.

FEWS NET & World Food Programme (WFP). 2007. Overall Trade Flows, *Informal Cross Border Food Trade in Southern Africa Report*, Issue 30. Available at: <http://home.wfp.org/stellent/groups/public/documents/ena/wfp201343.pdf>.

Folsom, R. 1991. *Exponential and logistic weight adjustments for sampling and non-response error reduction*. Proceedings of the Social Statistics Section, American Statistical Association.

- Fuller, W.** 1987. *Measurement Error Models*. Online, John Wiley & Sons.
- _____ 2009. *Sampling Statistics*. Chichester, U.K., John Wiley & Sons.
- Galmés M.** 2013. *Integrating expert opinion in agricultural statistics*. Paper presented at *the Sixth International Conference on Agricultural Statistics*, Rio de Janeiro, Brazil, 23-25 October 2013.
- Groen, A.** 2012. Sources of Error in Survey and Administrative Data: The Importance of Reporting Procedures. *Journal of Official Statistics*, 28(2): 173-198.
- Gronvall, A. & Karlsson, A.** 2013. The usefulness of quality frameworks when deciding on replacing Surveys with administrative registers – the Case of the Swedish sheep register. Paper presented at the Sixth International Conference on Agricultural Statistics, Rio de Janeiro, Brazil, 23-25 October 2013.
- GEO.** 2011. *Participating programs: Examples of Global to Regional Agricultural Monitoring Systems*. Available at: https://www.earthobservations.org/cop_ag_gams_pp.shtml.
- Groves, R.** 2004. *Survey Errors and Survey Costs* (Vol. 536). John Wiley and Sons, online.
- Holmberg, A.** 2012. Discussion on Assessing Quality of Administrative Data. *Statistica Neerlandica* 66(1): 34-40.
- Holt, D.** 2007. The Official Statistics Olympic Challenge: Wider, Deeper, Quicker, Better, Cheaper. *The American Statistician*, 61(1): 1-8.
- Holt, D. & Smith, T.** 1979. Post Stratification. *Journal of the Royal Statistical Society: Series A*, 142(1): 33-46.
- Iannacchione, V., Milne, J. & Folsom, R.** 1991. Response probability weight adjustments using logistic regression. Proceedings of the Survey Research Methods Section, American Statistical Association.
- India, Ministry of Earth Sciences.** Undated *On Operational Agro-Meteorological Advisory Services, Vision 2030*.

India, Ministry of Statistics and Programme Implementation. 2011. *Manual on Fishery Statistics*. Available at: http://mospi.nic.in/mospi_new/upload/Manual_Fishery_Statistics_2dec11.pdf.

InWEnt. 2007. *Data Harmonization and Data Requirements for a Sector Monitoring and Statistical Strategy*. Report of two workshops in the frame of the GTZ Sector Project Statistical Capacity Building for Improved MDG-Monitoring, held at Siavonga, Zambia, May 2007. Lusaka: InWEnt in collaboration with GTZ.

Iwig, W., Berning, M., Marck, P. & Prell, M. 2013. Data Quality Assessment Tool for Administrative Data. Available at: www.bls.gov/osmr/datatool.pdf.

Jorner, U. & Vogel, F. 2010. Quality in agricultural statistics. In Benedetti R., Bee B., Espa G., and Piersimoni, F., (eds) *Agricultural Survey Methods*. Chichester, United Kingdom, John Wiley & Sons.

Kafuli, D., Cusack T., Keyser, J., Olesh, G. & Wright J. 1999. Final Report of the MAFF/USAID Crop Forecasting Study: Understanding the Impact of Crop Forecasting on Public and Private Decisions and Improving Crop Forecasting in Zambia. Washington, DC: Associates in Rural Development RAISE Consortium.

Kantanantha, N., Serban, N. & Griffin, P. 2010. Yield and Price Forecasting for Stochastic Crop Decision Planning. *Journal of Agricultural, Biological, and Environmental Statistics*, 15(3): 362-380.

Karr, A. 2012. Discussion on Statistical Use of Administrative Data: Old and New Challenges. *Statistica Neerlandica*, 66(1): 80-84.

Keita, N. and Chin, N. 2013. *The place of 'assessment' in current agricultural statistics for developing countries: making best use of available information for timely crop production estimates in the absence of a system of agricultural sample surveys*. Paper presented at the *Sixth International Conference on Agricultural Statistics*, Rio de Janeiro, Brazil, 23-25 October 2013.

Kelly, V. & Donovan C. 2008. Agricultural Statistics in Sub-Saharan Africa: Differences in Institutional Arrangements and their Impact on Agricultural Statistics Systems: A Synthesis of Four Country Studies, *Michigan State University International Development Working Paper No. 95*.

- Kim, G. & Chambers, R.** 2012. Regression Analysis under Probabilistic Multi-Linkage. *Statistica Neerlandica*, 66(1):64-70.
- Kim, J.** 2010 Calibration Estimation Using Exponential Tilting in Sample Surveys. *Survey Methodology*, 36 (2); 145-155.
- _____ 2011. Parametric Fractional Imputation for Missing Data Analysis. *Biometrika*, 98 (1): 119-132.
- Kim, J. & Hong, M.** 2012. An Imputation Approach to Statistical Inference with Coarse Data. *Canadian Journal of Statistics*, 40 (3): 604-618.
- Kim, J. and Rao, J.** 2012. Combining Data from Two Independent Surveys: A Model-assisted Approach. *Biometrika*. 99 (1): 85-100.
- Kim, J. & Riddles, M.** 2012. Some Theory for Propensity Scoring Adjustment Estimator. *Survey Methodology*, 38(2): 157-165.
- Kim, J., Zhu, Z., and Yang, S.** (2013). Improved Estimation for June Area Survey Incorporating Several Information. Proceedings of the 59th ISI World Statistics Congress, Hong Kong.
- Kiregyera, B.** 2002. *Assessment of the national agricultural Statistical systems and their capacity to meet the PRSP requirements*. Paper presented at the Seminar on a New partnership to Strengthen Agricultural and Rural Statistics in Africa for Poverty Reduction and Food Security, Paris, 16-17 September 2001.
- Kiregyera, B., Megill, D., Eding, D. & Bonifácio, J.** 2007. Mission report from a review of the National Agricultural Information System in Mozambique. TA for the Scandinavian Support Program to Strengthen the Institutional Capacity of the National Statistics, Mozambique.
- Kizito, A.** 2011. The structure, conduct, and performance of agricultural market information systems in Sub-Saharan Africa. Michigan State University, (Ph.D. Dissertation).
- Laitila, T., Wallgren, A. & Wallgren, B.** 2011. Quality Assessment of Administrative Data. *Research and Development – Methodology Reports from Statistics Sweden*, 2011:2 Statistics Sweden.
- Lee, S.** 2006. Propensity Score Adjustment as a Weighting Scheme for Volunteer Panel Web Surveys. *Journal of Official Statistics*, 22(2): 329-349.

- Little, R.** 1988. Missing-data Adjustments in large Surveys. *Journal of Business and Economic Statistics*, 6(3): 287-296.
- Lohr, S. & Prasad, N.** 2003. Small Area Estimation with Auxiliary Survey Data. *The Canadian Journal of Statistics*, 31(4): 383-396.
- Malawi, Ministry of Finance and Development Planning.** 2012a. *Malawi Growth and Development Strategy II 2011-2016*. Available at: www.mw.one.un.org/wp-content/uploads/2014/04/Malawi-Growth-and-Dedvelopment-Strategy-MGDS-II.pdf.
- Malawi, National Statistical Office.** 2012b. *National Statistical Office Strategic Plan 2012-2016*. Available at: www.paris21.org/sites/default/files/MALAWI_NSO_SP_2012-2016.pdf.
- Manzi, G., Spiegelhalter, D., Turner, R., Flowers, J. & Thompson, S.** 2011. Modelling Bias in Combining Small Area Prevalence Estimates from Multiple Surveys. *Journal of the Royal Statistical Society, Series A*, 174(1): 31-50.
- McNairn, H., Champagne, C., Shang, J., Holmstrom, D. & Reichert, G.** 2009. Integration of optical and Synthetic Aperture Radar (SAR) imagery for delivering operational annual crop inventories., *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(5): 434-449.
- Mendes J., Zavala, A., Feuser, C. & Ferreira F.** 2013. *The influence of rural credit in agricultural gross added value*. Paper presented at the *Sixth International Conference on Agricultural Statistics*, Rio de Janeiro, Brazil, 23-25 October 2013.
- Mehta, R.** 2000. Remote sensing applications in agricultural statistics system: rationale, scope, and aims. *Indian Journal of Agricultural Economics*, 55(2): 87-95.
- Merkouris, T.** 2010. *Combining information from multiple surveys by using regression for efficient small domain estimation*. *Journal of the Royal Statistical Society: Series B* 68: 509-521.
- Naik, G., Basavaraj, K., Hegde, V., Paidi, V. & Subramanian, A.** 2012. Reliability of agricultural statistics in developing countries: reflections from a comprehensive village Survey on crop area statistics in India. *Working Paper No. 381*, Indian Institute of Management, Bangalore.

Nandram, B., Berg, E. & Barboza, W. 2013. A Hierarchical Bayesian Model for Forecasting State-level Corn Yields. *Environmental and Ecological Statistics*, 21(3): 507-530.

National Bureau of Statistics, Tanzania, 2012. Assessment of the Agricultural Routine Data System in Tanzania; JICA, December, 2012.

Nordbotten S. 2008. The use of administrative data in official statistics – past, present, and future – with special reference to the Nordic countries. In Carlson, M., Nyquist, H., & Villani, M. (eds) *Official Statistics: Methodology and Applications in Honour of Daniel Thorburn*, (pp. 205–223) Brommatryck & Brolins, Sweden.

Nusser, S.M., & Goebel, J.J. (1997). The National Resources Inventory: a long-term multi-resource monitoring programme. *Environmental and Ecological Statistics*, 4(3), 181-204.

Okello, P., Serugga, J., Pica-Ciamarra, U. & Baker, D. 2013. *Administrative records for official statistics: food for thought on Uganda's livestock data.* Paper presented at the *Sixth International Conference on Agricultural Statistics*, Rio de Janeiro, Brazil, 23-25 October 2013.

Pangapanga, P., Kanyanda, S., Kussein, G. & Tembo, L. 2013. *Economic constraints in agricultural statistics: could administrative data complement agricultural surveys in Malawi National Statistical Systems?.* Paper presented at the *Sixth International Conference on Agricultural Statistics*, Rio de Janeiro, Brazil, 23-25 October 2013.

Parihar, J. & Oza, M. 2006. FASAL: an integrated approach for crop assessment and production forecasting. In *Asia-Pacific Remote Sensing Symposium* (pp. 641101-641101). International Society for Optics and Photonic.

Pfeffermann, D. 2002. Small Area Estimation – New Developments and Directions. *International Statistical Review*, 70(1): 125-144.

Prasad, N. & Rao, J. 1990. The Estimation of the Mean Squared Error of Small-area Estimators. *Journal of the American Statistical Association*, 85 (409): 163-171.

Prell, M., Bradsher-Fredrick, H., Comisarow, C., Cornman, S., Cox, C., Denbaly, M., Wilkie Martinez, R., Sabol, W. & Vile, M. (2009). Statistical Uses of Administrative Data. Available at <http://www.bls.gov/osmr/fcsm.pdf>.

Pronab, S. Undated. Challenges of Using Administrative Data for Statistical Purposes: India Country paper. Available at: http://www.google.co.ls/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CCUQFjAA&url=http%3A%2F%2Funstats.un.org%2Funsd%2Fdnss%2FdocViewer.aspx%3FdocID%3D2426&ei=stPUVKygHMir7Aa51ICwBg&usg=AFQjCNGruAxcq-JrOC3_CrZDra5_EV508A&bvm=bv.85464276,d.ZGU .

Raghunathan, T., Xie, D., Schenker, N., Parsons, V., Davis, W., Dodd, K. and Feuer, E. 2007. Combining Information from Two Surveys to Estimate County-level Prevalence Rates of Cancer Risk factors and Screening. *Journal of the American Statistical Association*, 102(478): 474-486.

Rao, J. 2003. *Small Area Estimation*. Wiley Series in Survey Methodology. Chichester, U.K., John Wiley and Sons.

Ray, S., Panigrahy, S. & Parihar, J. 2008. Precision farming in Indian Context - Role of Remote Sensing. In Anandaraja, N., Chandrakandan, K., & Ramasubramaniam, M. (eds) *Extension of Technologies from Labs to Farms* (pp. 491 ff.). New Delhi: New India Publishing Agency.

Recide, R. 2010. *Food and Agricultural statistics in the Philippines*. Paper presented at the twenty-third session of the Asia and Pacific Commission on Agricultural Statistics, Siem Reap, Cambodia, 26-30 April 2010.

Rizzo, L., Kalton, G. & Brick, J. 1996. A Comparison of Some Weighting Adjustment Methods for Panel Non-response. *Survey Methodology*, 22(1): 43-53.

Rosenbaum, P. 1987. Model-based Direct Adjustment. *Journal of the American Statistical Association*, 82(398): 387-394.

Rosenbaum, P. & Rubin, D. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika*, 70(1): 41-55.

Savran F., Demiryürek K., Özçatalbaş O., Akin A. & Boz I. 2011. The Agricultural Extension in Turkey. *Scientific Research and Essays*, 6(8): 1821-1838.

Shepherd, A. 1997. Market Information Services: Theory and Practice. *FAO Agricultural Services Bulletin 125*. Rome, FAO. Available at: <ftp://ftp.fao.org/docrep/fao/003/X6993E/x6993e01.pdf>. Accessed on 26 May 2015.

Slud, E. 2004. Small area estimation in SAIPE using GLM versus FH models. Proceedings of American Statistical Association Selection on Survey Research Methods.

Srivastava, A. Undated. Agricultural Statistics System in India. Available at: http://www.iasri.res.in/ebook/TEFCPI_sampling/AGRICULTURAL%20STATISTICS%20SYSTEM%20IN%20INDIA.pdf.

Srivastava, S. 2011. Case study: Harnessing new tools and techniques for making agricultural statistics more efficient and evidence based support food security policy decisions in India. Available at: [https://wpqr4.adb.org/LotusQuickr/agstat_ap/Main.nsf/h_Toc/B4354BBD9836357D48257DF70038873F/\\$FILE/Harnessing%20new%20tools%20and%20techniques%20for%20making%20agricultural%20statistics%20more%20efficient%20and%20evidence%20based%20to%20support%20food%20security%20policy%20decisions%20in%20India.pdf](https://wpqr4.adb.org/LotusQuickr/agstat_ap/Main.nsf/h_Toc/B4354BBD9836357D48257DF70038873F/$FILE/Harnessing%20new%20tools%20and%20techniques%20for%20making%20agricultural%20statistics%20more%20efficient%20and%20evidence%20based%20to%20support%20food%20security%20policy%20decisions%20in%20India.pdf).

Tefft, J., McGuire, M. & Maunder, N. 2006. *Planning for the future. An assessment of food security early warning systems in sub-Saharan Africa*. Rome, FAO.

Trant M. 2011. Mission Report, Master Plan Project for Agricultural Statistics, 17-29 April 2011; Ministry of Agriculture/INE, Maputo, Mozambique

Uganda, Bank of Uganda 2011. *The Informal Cross Border Trade Survey Report 2011*. Available at: www.bou.or.ug/bou/bou-downloads/publications/TradeStatistics/ICBT/All/ANNUAL_INFORMAL_CROSS_BORDER_TRADE_SURVEY_REPORT_2011.pdf.

_____ 2012. *The Informal Cross Border Trade Survey Report 2012*. Available at: www.bou.or.ug/bou/bou-downloads/publications/TradeStatistics/ICBT/All/ANNUAL-INFORMAL-CROSS-BORDER-TRADE-SURVEY-REPORT-2012.pdf.

Uganda National Bureau of Standards (UNBS) 2012. *Guidelines for Production of Quality Statistics – Uganda Standard*. Reference Number: US943:2012.

United Nations, Economic Commission for Europe 2007. *Pros and cons for using administrative records in statistical bureaus*. Paper presented Seminar on *Increasing the Efficiency and Productivity of Statistical Offices, Session II*, 11-13 June 2007, Geneva.

_____ 2011. *Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices*.

UNSD. 2014. Development of National Statistical Systems. Available at: <http://unstats.un.org/unsd/dnss/default.aspx>.

Uruguay. Undated. Trazabilidad: 100% del ganado vacuno uruguayo está registrado e identificado. *Gabinete Productivo*. Available at: <http://gp.gub.uy/es/node/419>.

Vale, S. 2007. *How real are the barriers to the use of administrative sources for statistical purposes*. Paper presented as the 56th session of the *International Statistical Institute*, Lisbon, 27-29 August 2007.

Wallgren, A. & Wallgren, B. 2007. *Register-based Statistics - Administrative Data for Statistical Purposes*. John Wiley & Sons, Chichester, UK.

_____ 2010. Using administrative registers for agricultural statistics. In Benedetti, B., Bee, M., Espa, G., & Piersimoni, F. (eds) *Agricultural Survey Methods*. John Wiley and Sons, Chichester, UK.

_____ 2011. To understand the possibilities of administrative data, you must change your statistical paradigm. Proceedings of the American Statistical Association Section on Survey Research Methods, 357-365.

Wei, G. & Tanner, M. 1990. A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithm. *Journal of the American Statistical Association*, 85(411): 699-704.

World Bank, Food and Agricultural Organization of the United Nations (FAO) & United Nations. 2010. *Global Strategy to Improve Agricultural and Rural Statistics*. The International Bank for Reconstruction and Development/World Bank, Washington, DC.

World Bank and Ministry of Agriculture, Animal Industries and Fisheries, Uganda. 2012. Proposed Plan to Operationalise the Non-ATAAS Component of the Development Strategy and Investment Plan. Available at

<http://www.agriculture.go.ug/userfiles/Final%20Synthesis%20Report%20-%20NON-ATAAS%20DSIP%2014%20November%20%202012-2.pdf>. (.

Ybarra, L. & Lohr, S. 2008. Small Area Estimation when Auxiliary Information is Measured with Error. *Biometrika Trust*, 95(4): 919-931.

Zhang, L. & Fosen, J. 2011. A Modeling Approach for Uncertainty Assessment of Register-Based Small Area Statistics. *Journal of the Indian Society of Agricultural Statistics*, 66: 91-104.

Zhang, L. 2012. Topics of Statistical Theory for Register-based Statistics and Data Integration. *Statistica Neerlandica*, 66(1): 41-63.

Zhou, M. and Kim, J. 2012. An Efficient Method of Estimation for Longitudinal Surveys with Monotone Missing Data. *Biometrika* 99(3): 631-648.

Websites visited:

1. CSA www.csa-mali.org
2. DNSI www.dnsi.gov.ml

Appendix I

TABLE A: Using the Health Matrix Network Framework

Using the Health Matrix Network Framework					
Components	Agriculture	Education	Labour	Health	Etc.
	Scope				
Resources	These include the legislative, regulatory and planning frameworks to ensure a fully functioning Sector Information System and the resources available for the system. Such resources involve manpower, logistics support, information and communications technology, and coordinating mechanisms within and between the six components.				
Indicators	A core set of indicators and related targets, covering the main domains of sector information is the basis for a sector information system plan and strategy. These include determinants, such as sector system inputs and outputs, sector service coverage and quality.				
Data sources	The two main data sources are administrative records, and censuses and surveys.				
Data management	This covers all aspects of data handling from collection, management and flow to its processing and analysis.				
Information products	Data must be transformed into information that serve as the basis for evidence and becomes knowledge to shape policies and actions in the sector.				
Dissemination and use	The value of statistical information can be enhanced by making it readily accessible to decision-makers, giving due attention to behavioural and organizational constraints and incentives to use the data.				

Appendix II

TABLE B: Elements, attributes and indicators of quality of administrative registers

Quality elements	Quality attributes	Quality indicators
Administrative data source	1. Relevance	1.1. Utility
		1.2. Intended use
		1.3. Demand for information
		1.4. Satisfaction of primary users
	2. Information security and limitations on the use of the information.	2.1. Legal framework
		2.2. Personal data protection
		2.3. Limitations due to confidentiality regulations
		2.4. Confidentiality agreements
		2.5. Secure data transfer
		2.6. Confidentiality, integrity and availability of information
		2.7. Data protection
		2.8. Data backup policies
	3. Data delivery commitment.	3.1. Costs associated with the delivery
		3.2. Delivery agreements
		3.3. Frequency of deliveries
3.4. Dates of last five deliveries		
3.5. Punctuality		
3.6. Risks because of lack of data		
3.7. Alternative method to replace the lack of information		
3.8. Means of data delivery		
3.9. File format		
3.11. Data selection		

Quality elements	Quality attributes	Quality indicators
	4. Control and Continuous improvement.	4.1. Data collection
		4.2. Consistency control
		4.3. Change control
		4.4. Continuous improvement
	5. Data treatment.	5.1. Control of objective units
		5.2. Control of variable content
		5.3. Control of Outliers
		5.4. Changes
		5.5. Reasons for not changing
		5.6. Changes according to procedure
		5.7. Use of Database Management System
		5.9. Database documentation
		5.10. Database integrity
		Metadata
2. Completeness and clarity	2.1. Definition of population units	
	2.2. Description of variables	
	2.3. Communication of changes in definitions/concepts	
3. Use of unique keys	3.1. Identification keys	
	3.2. Comparability of identification keys	
	3.3. Unique combinations of variables	
4. Comparability	4.1. Comparability of the objective unit definition	
	4.2. Comparability of variable definitions	
Data	1. Technical controls	
		1.2. Redefinition of concepts and metadata in case of more than one data source
		1.3. Correspondence between data and metadata
		1.4. Record linkage method

Quality elements	Quality attributes	Quality indicators
		1.5. Verification of effectiveness of the record linkage method
	2. Coverage	2.1. Overcoverage
		2.2. Classification errors
	3. Record linkage	3.1. Rate of record linkage
	4. Completeness	4.1. Rate of unit non response
		4.2. Rate of item non response
	5. Measurement	5.1. External control (audit)
	6. Identification keys	6.1. Rate of records with unique key
	7. Data processing	7.1. Data editing
		7.2. Imputation
	8. Data accuracy	8.1 Data accuracy
	9. Coding	9.1. Use of standard coding
		9.2. Verification of coding
		9.3. Rate of coding errors
9.4. Rate of records without code		
10. Data freshness	10.1. More than 90 percent of the objective units created during year t have been registered before the end of the year t+1	
11. Multiple records	11.1. Rate of multiple records of the same unit	
12. Other controls	12.1. Rate of units with valid values into identification keys	
	12.2. Tables of the statistical operation have been validated through automatic procedures	
Statistical product	1. Comparability	1.1. Length of comparable time series
		1.2. Comparability of microdata over time
	2. Relevance	2.1. Identification of users
		2.2. Information about users
		2.3. Rate of final user satisfaction
2.4. Utility (intended uses)		

Quality elements	Quality attributes	Quality indicators
	3. Coherence	3.1. Coherence of statistics with different periodicity
		3.2. Coherence of statistics with the same socioeconomic scope
	4. Availability and clarity	4.1. Accessibility by Internet
		4.2. Rate of completeness of metadata
	5. Accuracy	5.1. Coefficient of variation
		5.2. Rate of unit non-response
		5.3. Rate of item non-response
		5.4. Rate of imputation
		5.5. Rate of editing
		5.6. Rate of over coverage
		5.7. Rate of classification errors
	6. Timeliness and punctuality	6.1. Punctuality of the statistical product dissemination
		6.2. Length of time between its availability and the event or phenomenon it describes
6.3. Duration between reference time point of administrative data and date of availability to the statistical office		