Food and Agriculture Organization of the United Nations

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

Technical Report 8

Final Strategy and Methodological Report

Submitted to

Food and Agriculture Organization of the United Nations

Under

The Global Strategy to Improve Agriculture and Rural Statistics

By

School of Statistics and Planning (SSP)
College of Business and Management Sciences (CoBAMs)
Makerere University
UGANDA

Contact Person:
Dr. Leonard Atuhaire
Associate Professor at SSP – CoBAMs
Address: P.O. Box 7062, Kampala
Email: latuhaire@bams.mak.ac.ug

and

The Centre for Survey Statistics and Methodology (CSSM)
Iowa State University (ISU)
UNITED STATES OF AMERICA

2016

Contact Person:
Dr. Zhengyuan Zhu
Associate Professor at ISU and the Director of CSSM
E-Mail: zhuz@iastate.edu

Table of Contents

List of tables and figures

Acronyms and abbreviations

AAS	Annual Agricultural Survey
ADSAS	Administrative Data Systems for Agricultural Statistics
ADB	Asian Development Bank
AfDB	African Development Bank
ANADER	*Agence Nationale d'Appui au Développement Rural* (Côte d'Ivoire)
ARDS	Agricultural Routine Data System
ASDP	Agricultural Sector Development Plan
ASSD	African Symposium for Statistical Development
ASSP	Agricultural Statistics Strategic Plan
CAPE	Crop Acreage and Production Estimation

CAB	Cotton Advisory Board (India)
CAPI	Computer-Assisted Personal Interview
CAS	Centre for Agricultural Statistics (Lao People's Democratic Republic)
CDO	Cotton Development Organisation (Uganda)
COOIT	Central Organisation for Oil Industry and Trade
CPI	Consumer Price Index
CSO	Central Statistics Office
CSSM	Centre for Survey Statistics and Methodology (Iowa State University)
CV	Coefficient of Variation
CWC	Central Water Commission (India)
DESMOA	Directorate of Economics and Statistics, Ministry of Agriculture (India)
DNSI	*Direction Nationale de la Statistique et de l'Informatique* (Mali)
DRC	Department of Revenue and Customs (Bhutan)
FAO	Food and Agriculture Organization of the United Nations
FCBL	Food Corporation of Bhutan Limited
FSA	Farm Services Agency (USDA)
FSI	Forest Survey of India
FTP	Field Test Protocol
GCES	General Crop Estimation Surveys (India)
GDP	Gross Domestic Product
GPS	Global Positioning System
GSARS	Global Strategy to improve Agricultural and Rural Statistics
IACS	Integrated Administrative and Control System
ICS	Improvement of Crop Statistics (India)
INS	*Institut Nationale de la Statistique* (Côte d'Ivoire)
IMDB	Integrated Metadata Base
JICA	Japan International Cooperation Agency
LGA	Local Government Authority
LGMD	Local Government Management Database
LSB	Lao Statistics Bureau
MAAIF	Ministry of Agriculture, Animal Industry and Fisheries (Uganda)
MAF	Ministry of Agriculture and Forestry (Lao People's Democratic Republic)
MAFC	Ministry of Agriculture Food Security and Cooperatives (United Republic of Tanzania)
MALF	Ministry for Agriculture, Livestock, Forestry and Fisheries
MANR	Ministry of Agriculture and Natural Resource (United Republic of Tanzania)
MAWF	Ministry of Agriculture, Water and Forestry (Namibia)
MDA	Ministries, Departments and Agencies
MINADER	Ministry of Agriculture and Rural Development (Côte d'Ivoire)
MIT	Ministry of Industry and Trade (United Republic of Tanzania)
MLF	Ministry of Livestock and Fisheries (United Republic of Tanzania)
MLFD	Ministry of Livestock and Fisheries Development (United Republic of Tanzania)
MMSE	Minimum Mean Squared Error
MoAF	Ministry of Agriculture and Forests (Bhutan)
MoF	Ministry of Finance (Bhutan)
NAADS	National Agricultural Advisory Services (Uganda)
NAGRIC	National Animal Genetics Resources Centre (Uganda)
NARO	National Agricultural Research Organization (Uganda)
NASS	National Agricultural Statistics Service (United States of America)

NBS	National Bureau of Statistics
NCA	National Census of Agriculture
NHB	National Horticultural Board (India)
NRCS	National Resources Conservation Service (United States of America)
NRDCL	National Resources Development Corporation Limited (Bhutan)
NRSA	National Remote Sensing Agency (India)
NSA	Namibia Statistics Agency
NSI	National Statistical Institute
NSO	National Statistics Office
ONDR	*Office Nationale de Développement de la Riziculture* (National Office of Rice Development, Côte d'Ivoire)
OCPV	*Office d'aide à la commercialisation des Produits Vivriers* (Côte d'Ivoire)
PDA	Personal Digital Assistant
PHC	Population and Housing Census
PMO-RALG	Prime Minister's Office for Regional Administration and Local Government Authority (United Republic of Tanzania)
PPQ&S	Directorate of Plant Protection, Quarantine and Storage (India)
PPS	Probability Proportional to Size
RAAD	Routine Administrative Agricultural Data
RNR	Renewable Natural Resources
RS	Remote Sensing
SAP	*Système d'Alerte Précoce* (Early Warning System)
SASA	State Agricultural Statistics Authority (India)
SSP	School of Statistics and Planning, Makerere University (Uganda)
TAD	Traditional Administrative Data
TRS	Timely Reporting Scheme
UBOS	Uganda Bureau of Statistics
UCA	Uganda Census of Agriculture
UCDA	Uganda Coffee Development Authority
USDA	United States Department of Agriculture
VAEO	Village Agricultural Extension Officer (United Republic of Tanzania)
WAEO	Ward Agricultural Extension Officer (United Republic of Tanzania)

Preface

The Global Strategy to improve Agriculture and Rural Statistics (hereafter, Global Strategy or GSARS), adopted by the United Nations Statistical Commission in 2010, aims to improve the quality and sustainability of statistics on agriculture in developing countries. One of the key components of the Global Strategy's Global Action Plan is its Research Plan, which provides support for the research and development of cost-effective methods that will serve as the basis for 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 is the topic on *improving the methodology for using administrative data in agricultural statistics* (World Bank, 2010). Administrative data are non-statistical sources of information obtained through, for example, government programs or agricultural extension, and can benefit the final statistical product in ways ranging from reduced costs to improved small area estimates.

Under this topic, the aim is to research methods to improve the collection, management and use of administrative data for the production of agricultural statistics in developing countries. This publication constitutes a primary expected product of this research, a final technical report that includes a country-tested methodology. The development of this final report proceeded in two main stages. The first was the drafting of seven technical reports reviewing the current uses and challenges associated with the use of administrative data in both developed and developing countries[1]. The second stage consisted in the formulation, execution and documentation of in-country field tests. To appropriately consider the variability in administrative information systems across countries (e.g. differences in the approaches to collecting and using administrative data between, on one hand, French-, Portuguese- and Spanish-speaking countries and on the other, English-speaking countries), the country tests engaged one Francophone country and two Anglophone countries.

Technical Reports 1 through 5 provide literature reviews, identify challenges specific to developing countries, and propose procedures for in-country testing. On the basis of the findings of the first four technical reports, which consist of literature reviews and analyses of surveys conducted with the heads of statistical offices of developing countries, two types of pilot projects were considered suitable. The first type of pilot project, termed "desktop analyses", involves analyses of collected data. The second type, the "in-field tests", requires coordination with country representatives to review various dimensions of their administrative data systems for agricultural statistics.

[1] Technical Reports 1 through 5 were published in 2015 and are titled as follows: Technical Report 1 – *Reviewing the Relevant Literature and Studies on the Quality and Use of Administrative Sources for Agricultural Data* (GSARS, 2015a); Technical Report 2 – *The Role of Administrative Data in Developed Countries: Experiences and Ongoing Research* (GSARS, 2015b); Technical Report 3 – *Critical Analysis of Agricultural Administrative Sources Being Currently Used by Developing Countries* (GSARS, 2015c); Technical Report 4 – *Analysis of Agricultural Administrative Data Gaps and Ways of Improving the Quality and Use of Administrative Data Sources for Agricultural Statistics* (GSARS, 2015d); Technical Report 5 – *Strategy and Methodology for Improving the Use of Administrative Data – A Protocol for in-Country Testing* (GSARS, 2015e). Technical Reports 6 and 7 were made available in 2016: Technical Report 6 – *Findings from the Field in-Country Testing* (GSARS, 2016a); Technical Report 7 - *Strategy and Methodology for the Improvement of the Collection and Management of Administrative Data in an Agricultural Statistics System* (GSARS, 2016b).

Technical Reports 6 and 7 document the in-field tests and the desktop analysis. In-field testing (also known as field-based testing) focuses on organizational infrastructure, data collection methods, the capacity for enhancement of statistical analysis procedures, and the effectiveness of data dissemination techniques. The implementation of the field-based testing entailed a review of the administrative data systems and sources in selected countries, conducted through visits to statistical offices, with a view to suggesting areas for improvement. The countries that participated in the field-based tests were the United Republic of Tanzania and Côte d'Ivoire; thus, both English and French-speaking countries were represented. Technical Report 6 presents the results of the visits to test countries. The desktop analyses focused on the reconciliation of data from different sources and proceeded in two broad phases: the first phase required coordination with the relevant institutions to obtain, prepare and understand the different data sources; the second phase involved the use of systematic statistical procedures to synthesize multiple sources of information in an objective manner. The test countries selected for the desktop analyses were Namibia and the United Republic of Tanzania.

This final report, classified as Technical Report 8, synthesizes Technical Reports 1 through 7. The final Technical Report draw upon the experiences of developed countries noted in the literature reviews in Technical Reports 1 and 2, as well as on the in-field tests and the interviews conducted with the upper management of statistical offices in Africa. Selected results from the in-field tests and the desktop analysis provide examples throughout the text.

Acknowledgments

This Final Technical Report were produced by the Global Office of the Global Strategy to improve Agriculture and Rural Statistics under the overall coordination of Flavio Bolliger, Research Coordinator and Naman Keita, Senior Consultant in Agricultural Statistics. The consultants worked under the supervision of Dramane Bako, International Consultant – Statistician and Carola Fabi, Senior Statistician.

The research project was subcontracted to two institutions, which worked together in partnership: the School of Statistics and Planning/College of Business and Management Sciences of Makerere University (Uganda) and the Centre for Survey Statistics and Methodology of Iowa State University (USA). The research teams of the two institutions were composed as follows:

- **Makerere University**: Agnes M.N. Ssekiboobo (Team Leader), Dr Elijah S. K. Muwanga Zake, Dr Leonard Atuhaire (Contact Person), Dr Abraham Yeyo Owino, Felix Wamono, Johnson Kagugube, Dr Andrew Muganga Kizito, and Ankouvi Nayo (ENSEA, Abidjan).
- **Iowa State University**: Dr Emily Berg, Dr Zhengyuan Zhu (Contact Person), Dr. Jae-Kwang Kim, Dr. Jongho Im, Dr. Jie Li, and Colin Lewis Beck.

The document benefited from contributions by Anders and Britt Wallgren (international experts, formerly of Statistics Sweden) and Dalisay Samarita Maligalig (international expert, University of the Philippines at Los Banos).

We also wish to extend their gratitude to the participants in the FAO validation workshops[2] held in Rome, and in the review meetings held in Rome and New York, as well as the reviewers of the various draft versions, whose feedback enriched the technical reports drafted under this research project.

Several institutions and individuals contributed in various ways during the performance of this research. Among these are the Directors General of Statistics of the National Statistical Offices of the African countries that attended the African Symposium for Statistical Development (ASSD), which took place in Kampala from 12 to 14 January 2015.

The authors are grateful to Kali Ndamona (Namibia Statistics Agency) for her generous assistance and collaboration on the desktop analysis. In particular, we wish to thank her and Sikongo Haihambo for providing data from the 2013/2014 Namibia Census of Agriculture. We further thank several representatives of Namibia's Ministry of Agriculture, Water and Forestry, who provided valuable information on administrative data on

[2] The meeting of 5-6 September 2016 was attended by: Zhengyuan Zhu (Iowa State University, USA), Emily Berg (Iowa State University, USA), Jongho Im (Iowa State University, USA), Leonard Atuhaire (Makerere University, Uganda), Agnes M.N. Ssekiboobo (Makerere University, Uganda), Abraham Yeyo Owino (Makerere University, Uganda), Linda Young (NASS-USA), Cynthia St-Germain (StatCan-Canada), Dalisay Samarita Maligalig (University of the Philippines at Los Banos), Lassina Paré (Agricultural Statistics Department, Burkina Faso), Titus Mwisomba (National Bureau of Statistics, United Republic of Tanzania), Ndamona C. Kali (Namibian Statistics Agency), Anders Wallgren (International Expert, Sweden), Moussa Kaboré (International Expert, Burkina Faso), Flavio Bolliger (Global Strategy), Naman Keita (Global Strategy), Carola Fabi (Global Strategy), Valerie Bizier (Global Strategy) and Dramane Bako (Global Strategy)
The meeting of 13-14 April 2015 was attended by: Zhengyuan Zhu (Iowa State University, USA), Emily Berg (Iowa State University, USA), Jae-Kwang Kim (Iowa State University, USA), Agnes Ssekiboobo (Makerere University, Uganda), Elijah Muwanga-Zake (Makerere University, Uganda), Abraham Yeyo Owino (Makerere University, Uganda), Miguel Galmes (ex-Consultant), Dalisay Maligalig (Asian Development Bank), Titus Mwisomba (National Bureau of Statistics, United Republic of Tanzania), Hamidreza Jalali (Statistical Centre of Iran), Ndamona C. Kali (Namibian Statistics Agency), Jose Hernandez (INEGI), Aida Khalil (Global Strategy), Naman Keita (Global Strategy), Carola Fabi (Global Strategy), Michael Rahija (Global Strategy) and Christophe Duhamel (Global Strategy).

agriculture in the country: these are Matheus Ndjodhi, Alec Bishi and Frank Chitate. Likewise, we wish to thank Cristof Brock and Antoinette Venter (Namibia Agronomic Board).

The active participation of the Ministry of Agriculture and Rural Development (MINADER), the *Agence Nationale d'Appui au Développement Rural* (ANADER) and the *École Nationale Supérieure de Statistique et d'Economie Appliquée* (ENSEA) in the execution of the field tests in Côte d'Ivoire is deeply acknowledged. Our special thanks go to Dr Nouhoun Coulibaly, Director General of Planning, Project Control and Statistics (DGPPS), Soro Kouhonan, Koffi Gabriel, Konan Hilaire, KouameYoboua Malan Mathieu and Robert Touboui (MINADER); Camara Walley (ANADER) and Hugues Kouadio (Director), Ankouvi Nayo and Jean Stephane Nzi (ENSEA) for their contributions to the field exercise and their highly appreciated cooperation and expertise. ENSEA is further thanked for the valuable technical support provided by its staff throughout the project. Our gratitude is also extended to the Regional Officers of MINADER and ANADER in the districts of Abengourou and Adzopé, as well the field team that worked tirelessly to successfully implement the data collection phase of the field tests. The contribution of the individual farmers who responded to the different questionnaires that were administered is appreciated.

The authors also thank the United Republic of Tanzania's National Bureau of Statistics, the Ministry of Agriculture Food Security & Cooperatives (MAFSC), the Ministry of Livestock and Fisheries Development (MLFD) for their support in undertaking the field tests in the country. In particular, we thank Dr Albina Chuwa, Director General, National Bureau of Statistics (NBS), Mwisomba Titus (NBS), Jingushi Mana (JICA), Dr Sophus Mlete and Mlau Da Silva (MLFD); John Maige, Elias Abel Mhehe and Elias Masunga (MAFSC), for their valuable contribution and participation in the implementation of the field tests. Our appreciation also goes to the Regional and District Officers of MLFD and MAFSC for their supervisory role, the extension staff of both MAFSC and MLFD who collected the data, and the farmers of Bagamoyo and Morogoro districts who provided the data.

Executive summary

Researchers, government agencies, and policymakers have proposed several definitions of the term "administrative data". The succinct definition used for this report describes administrative data as the "information collected primarily for administrative 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"[3]. Administrative data are sometimes used for academic research, although its collection is typically geared towards practical and applied research questions. In both developing and developed countries, the information gained from administrative data can contribute towards the improvement of public policy in areas such as food security, poverty, environmental health and economic development.

Agricultural statistics benefit from administrative data, as numerous agencies and businesses linked to the agriculture industry regularly collect relevant data as a by-product of other administrative processes. For example, reports from agricultural field/extension staff, tax data, land ownership records, information on government subsidies, import/export data, administrative farm registers and other registration or licensing systems, records on agro-tourism, farmers' associations, and private businesses data all contain potentially useful information. In addition to cross-sectional data, there is also an increase in data that monitor or trace agricultural products across time. For instance, in Uruguay, individual livestock are now tracked within a herd (Boland et al., 2007). This generates large volumes of time-series data that are precise and regularly updated.

Supplementary administrative data is important because it may address populations for which no government data are available, or enhance existing data sets with additional information. Due to the resource limitations commonly encountered in developing countries, this is extremely valuable, because formal survey sampling is limited and delays may prevent data from reaching policymakers in a timely manner. Administrative data, on the other hand, are capable of providing a better spatial coverage, are often collected more frequently, and are cheaper. Such data are also more accessible, due to their faster processing times and the constant improvements in Internet and computer technology. From a research perspective, administrative data can improve a variety of statistical methodologies, such as sampling frame construction and sample design; the use of administrative records to cover data gaps from surveys and censuses; forecasting; planning; and provision of small area estimates and administrative uses. As a result, more precise estimates may be obtained, which may ultimately lead to more informed policymaking efforts.

However, there are drawbacks to the use of administrative data. Several issues stem from the fact that statistical offices use administrative data for a purpose different than that for which the data was originally collected. This creates a variety of complications, such as changes in administrative processes, which in turn result in inconsistencies in estimates across time and a reduced data availability; a lack of standardized definitions of variables, units and identifiers, which may pose challenges to the synthesis of multiple sources; undercoverage 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 the use of administrative data often involves processing and cleaning massive volumes of data; and the challenge of maintaining confidentiality if the administrative data involves confidential information. All of these issues may hinder comprehension of the data or gaining significant improvements in estimation, regardless of the statistical techniques employed. However, these limitations give rise to research and policy opportunities to develop new

[3] See www.adls.ac.uk/adls-resources/guidance/introduction/.

methods and best practices, ultimately to improve the quality and reliability of administrative data, in developing countries especially.

Both developing and developed countries use administrative data to improve agricultural statistics. As a whole, developing countries mostly employ administrative data in the determination of the Gross Domestic Product (GDP). Individual developing countries, however, use administrative data in a variety of ways. India and many other developing countries use administrative records and other forms of administrative data to develop sample frames for a wide range of activities (Sen, undated). In Malawi, agricultural offices, through officers based within villages, collect information on the strategies, challenges and weather parameters of the area, which is crucial when designing national initiatives (Pangapanga et al., 2013). Administrative data is also used when forecasting crop yield and/or production for food security purposes (India); covering gaps in national agricultural surveys (Malawi); and small-area estimating agricultural statistics at the district level (Ethiopia).

Despite the importance of administrative data for developing countries, much of the administrative data is collected and compiled without employing standard statistical procedures or researchers trained in statistical methods. Research has shown that a large proportion of administrative data consists of guessestimates and is believed to be of questionable quality (UBOS, 2007; Kiregyera et al., 2007). Other than import/export and agricultural price information, which are often circulated broadly, agricultural administrative data are not as widely disseminated. As mentioned earlier, there is also an issue of untimely and incomplete flow of data from the lower to the higher reporting levels. This may lead to delays in the ability of governments to make policy decisions or a general lack of understanding of their own country's data (UBOS, 2007).

In developed countries, there is a much stronger link between administrative data and the national statistics system. In many cases, administrative data are sufficiently reliable to support the direct tabulation of official statistics. For example, Statistics Canada has directly used administrative data for statistics related to trade, unemployment insurance claims, income distributions and annual migration (Brackstone, 1987). Agricultural statistics in particular may benefit from several sources of administrative data. Some examples include tax data, land ownership records, vehicle registration data, information on government subsidies, insurance information, records on agro-tourism and meteorological data. In addition, collaborations between universities and government agencies have been established to improve the government statistics that use administrative data. The Center for Survey Statistics and Methodology of Iowa State University has worked with the National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA) over the past four years on the use administrative records and other remote sensing (RS) data to improve the estimation of crop acreage. Other areas of research using administrative data include the use of integrating factors, the improvement of area frames, the use of administrative data as covariates, and measurement error models. Due to the significant budget cuts made within the National Statistics Offices (NSOs) of developed countries, the importance of administrative data and researching methodologies to maintaining a high quality of data will continue to grow.

1 Introduction

1.1 Background and Motivation

The Global Strategy to improve Agriculture and Rural Statistics (hereinafter, Global Strategy) was adopted by the United Nations Statistical Commission (UNSC) in 2010. The strategy aims to improve statistics on agriculture, livestock, aquaculture, small-scale fisheries and forestry production in developing countries and ensure the sustainability of their maintenance. Its main objective is the construction of statistical capacity in developing countries for key basic food and agricultural statistics (World Bank, 2010).

One of the key components of the Global Strategy's Global Action Plan is its Research Plan, which aims to convey cost-effective statistical methods in technical guidelines, handbooks and training material. These resources are intended for use by consultants, country statisticians and training centres. One of the key priorities of the Research Plan, which was to be implemented in 2014, was a research project titled "Improving the methodology for using administrative data in agricultural statistics" (World Bank, 2010).

Administrative data have potential value for improving agricultural statistics in developing countries for several reasons. One of the major factors that limit the ability of developing countries to conduct regular surveys and

censuses is a lack of financial resources. Because administrative data are often collected for non-statistical purposes, statistical offices do not need to incur the cost of data collection when using administrative data, as they would for surveys and censuses. Administrative data are often complete enumerations of the target population defined by the administrative program.

The value of administrative data in developing countries is further supported by a non-probability sample survey of the heads of the statistical offices of 13 African countries (GSARS, 2015c). Of the 13 countries represented in the survey, administrative sources were rated as the major source of over 50 per cent of data on agricultural inputs and agro-processing, in almost all instances. Administrative sources are the major source of data on external trade in over 85 per cent of the African countries that participated in the country assessment. Based on this assessment, the types of data for which over half of the African countries use administrative sources as their major data source include the value of wood from forest production (53.6 per cent) and the quantity of wood from forest production (55.9 per cent). Almost half of the countries that participated in the FAO assessment use administrative data to ascertain the value (47.2 per cent) and quantity (47.5 per cent) of fishery and aquaculture production. About one quarter of these countries report using administrative data sources as their major source of data for crop production value (23.7 per cent), livestock production quantity (27.3 per cent) and livestock production value (29.7 per cent). In Côte d'Ivoire, Kenya, Namibia, Sierra Leone, and the United Republic of Tanzania, the main source of crop production quantity data was reported to be of administrative nature. Over a fifth (21.6 per cent) of the African countries that participated in the FAO assessment derive livestock production value data from administrative estimates or forecasts.

Despite the widespread use and potential benefits of using administrative data for official agricultural statistics, several challenges remain. Administrative data may not always be of adequate quality for use in official statistics. It may be difficult to integrate administrative data with survey or census data. In addition, it is often necessary to coordinate with multiple institutions when using administrative data, which may also present logistical obstacles.

In light of the potential benefits and challenges of using administrative data, this research seeks to evaluate the feasibility of using administrative data for the production of official agricultural statistics in developing countries. A precise definition of administrative data is provided, and sources of administrative data for agricultural statistics are identified. The benefits and challenges of using these data for agricultural statistics are considered. In addition, ways to improve usability are explored, as well as the issues associated with costs and institutional arrangements.

1.2 Overview of project objectives and approaches

The research project aims to develop strategies and methodologies to improve (1) the collection and management of data from administrative sources and (2) the use of administrative data in an integrated agricultural statistics system in developing countries. This requires investigation of cost-effective approaches and methods for the production of reliable, regular and geographically disaggregated agricultural data, including the combination of surveys and administrative data. Thus, the expected primary products of this research comprise (i) eight technical reports, which include country-tested and validated methodology to improve and make available administrative data for producing agricultural statistics in developing countries and (ii) a proposed strategy on how to use administrative data in cost-effective agricultural statistical systems. This Handbook constitutes the eighth of the eight reports. The contents of the first seven reports are examined briefly

in Chapter 2. The technical reports and proposed strategies develop a sound methodology for improving and using administrative sources for agricultural statistics in developing countries, taking into consideration existing approaches relating to administrative information systems in the various countries.

The research activities commenced in September 2014 and have been divided into eight tasks, performed by the Center for Survey Statistics and Methodology (CSSM) of Iowa State University (ISU), USA, and the School of Statistics and Planning (SSP) of the College of Business and Management Sciences (CoBAMS) of Makerere University, Uganda. This collaboration allowed combining experiences from developed and developing countries.

1.3 Overview of the institutional arrangements in Namibia, United Republic of Tanzania and Côte d'Ivoire

The potential of using administrative data for statistical purposes depends to a large extent on the institutional arrangements in place within a particular country. Some countries have centralized systems, in which the activities for all sectors are conducted by a single statistical institution. Examples of countries with centralized systems include Canada and Namibia. Other countries have decentralized systems, in which different national or state departments administer the statistical activities relevant to the sectors that they control. Examples of countries with decentralized systems include the USA and India.

To obtain a diverse representation of countries in Africa, pilot projects were conducted in Namibia, the United Republic of Tanzania and Côte d'Ivoire. This report constantly makes reference to best practices and areas for improvement gleaned through the pilot projects. To provide readers with background information for understanding these examples, this section sets out a brief overview of the institutional arrangements in Namibia, United Republic of Tanzania and Côte d'Ivoire.

1.3.1 Namibia

Namibia has a centralized statistical system. The Namibia Statistics Agency (NSA) is responsible for the collection and dissemination of statistical information in Namibia. The NSA collects and publishes information related to demographics as well as to agriculture. Namibia has engaged in three major data collection efforts related to agriculture in the last 20-25 years. Following the 1994-1995 agricultural census, Namibia conducted a series of annual agricultural surveys spanning the years 1996/1997 through 2000/2003. In 2013/2014, the NSA conducted a second agricultural census, essentially a large-scale survey covering vast aspects of agriculture and aquaculture.

The NSA and the Namibia Ministry of Agriculture, Water, and Forestry (MAWF) have established memoranda of understanding that facilitate collaboration and data sharing. The MAWF collects several sources of administrative data that may potentially enhance agricultural statistics. A prominent example is Namibia's livestock tracing system, which is currently the only comprehensive livestock tracing system in Africa. This and other examples for Namibia are discussed as relevant to the general issues discussed in this final Technical Report.

1.3.2 United Republic of Tanzania

The United Republic of Tanzania has had a decentralized statistical system since 1996, with a hierarchy of only two levels: a central government layer with regional offices, and a layer of local government. All statistical data collections in the country are governed by the Statistics Act 2002, which was revised in 2015. The Statistics Act 2002 regulates the collection of data and compilation of statistics and provides that the responsibility for the production, coordination, supervision and dissemination of official statistics lies with the National Bureau of Statistics (NBS), the central statistical authority and custodian of official statistics.

The United Republic of Tanzania was one of the first countries to develop an Agriculture Statistics Strategic Plan (ASSP) in Africa. The ASSP supports evidence-based policymaking by defining the data needs of national development initiatives. It sets the key priorities for the production of sustainable agricultural statistics and provides a road map for developing the Agriculture Statistics System over a timeframe of five years (2014-2015 through 2018-2019). The NBS, which is responsible for statistical coordination in mainland Tanzania, implemented an outline of activities described in the ASSP. In Zanzibar, the Office of Chief Government Statistician (OCGS) is responsible for the implementation of the ASSP. This effort was carried out in collaboration with Agriculture Sector Lead Ministries (ASLMs) and Development Partners led by FAO. In particular, the ASLMs are the Ministry of Agriculture, Food Security and Co-operatives (MAFC), the Ministry of Industry and Trade (MIT), the Ministry of Livestock and Fisheries Development (MLFD) and the Prime Minister's Office for Regional Administration and Local Government Authority (PMO-RALG). In Zanzibar, the actors are the Ministry of Agriculture and Natural Resource (MANR) and the Ministry of Livestock and Fisheries (MLF). Other stakeholders are farmers, development partners, farmer organizations, crop and livestock boards and trade companies.

Among other data sources for agriculture, the most prominent is the National Sample Census of Agriculture (NSCA), which is conducted every five years under the aegis of the NBS. The last two censuses were conducted in 2007-2008 and 2016. The other sample survey which touches on agricultural aspects is the National Panel Survey, which has also been conducted by the NBS every other year since 2008-2009. However, this survey provides estimates at national level. The latest population and housing census, conducted in 2012, had some agriculture-related questions in the form of an agricultural module. The main purpose of collecting data on agriculture in the population census is to provide a sampling frame of agricultural households, which may be subsequently used in NSCA and agricultural annual sample surveys. The agricultural module also collects data on the number of livestock in the individual household. This information provides a spatial distribution of crops and livestock, which is useful in planning agricultural and livestock surveys. Annual agricultural sample surveys have not been conducted in the country since the late 1990s.

The Agricultural Routine Data System (ARDS) was developed to meet the data needs for the monitoring and evaluation of the Agricultural Sector Development Programme (ASDP). The ARDS plays an important role in delivering field-level agricultural information to districts, regions and ASLMs. However, it has been difficult to engage in effective monitoring, supervision, planning and policy formulation, partly due to the inadequate functioning of the ARDS. After decentralization in 1996, the information flow had been routed as follows: villages → wards → districts → regions → PMO-RALG. Central ministries received information through the PMO-RALG. One of the purposes of improving the ARDS was to revive the flow of routine reporting, which originates in the Crop and Livestock Development Report, from the Local Government Authority (LGA) to the

ASLMs via regions, using a village/ward formats. Currently, in the improved ARDS, the data flow route is the following: village → ward → districts (where it is consolidated and transferred to the LGMD-2 server) → PMO-RALG, through Regions. Information is collected on the particular characteristics of the enumeration areas, the holdings, the holders' household, crop production, rainfall, disasters and plant health, as well as on livestock and poultry production and health.

1.3.3 Côte d'Ivoire

In Côte d'Ivoire, agricultural statistics are produced and disseminated by the Department of Statistics and Documentation of the Ministry of Agriculture and Rural Development (MINADER). MINADER usually produces two publications per year, the *Agricultural Statistics Yearbook* and the *Food Balance Sheets*. The data for these two publications come from two major sources of collection: the census and administrative data collection. Extension work is carried out by ANADER, a private limited company with agents up to village level in most regions of the country. ANADER occasionally collects information on some food crops. The information is forwarded to the regional governments and MINADER.

Due to political instability in the country, there have been wide temporal gap between successive agricultural censuses, with the last two censuses having been conducted in 2001 and, next, in 2016. Therefore, administrative agricultural data is increasingly essential to the production of agricultural statistics, because the previous estimates were based on the sampling frame and input from the 2001 census. The Ministry of Agriculture (MINAGRI), today MINADER, is responsible for receiving and validating all the agricultural administrative data collected by the different institutions. However, this administrative data (especially that on food crops) is obtained mainly from along the value chain; this implies that the production utilized at the household level is excluded. MINADER is aware of these underestimations, as well as those due to informal trade, because informal cross-border trade surveys are not conducted. Therefore, there is no system for generating crop area and production estimates for food crops.

The administrative data collection efforts are carried out by 18 national institutions operating in the agricultural sector, the major ones of which include: the Cotton and Cashew Regulatory Authority (ARECA) and the Council of Cotton and Cashew, which provide data on the areas, production and purchase prices to producers and exporters of cotton and cashew nuts; the Council of Coffee and Cocoa (previously ARCC and CCGC), for data on the marketed production of coffee and cocoa; the Professional Association of Natural Rubber (APROMAC) for areas planted with rubber and rubber products in village plantations and research stations, as well as purchase prices from producers; sugar companies such as the Operating Company of Sugar Cane Plantations (SUCRIVOIRE) and Sucrerie Africaine – Côte d'Ivoire (SUCAF-CI) (African Sugar Producer – Côte d'Ivoire) for data on the areas planted with sugar cane, and on sugar production (industrial and village plantations) and the marketed price; *Organisation Centrale des Producteurs Exportateurs Ananas Et De Bananes* (Central Organization of Producers-Exporters of Pineapple and Banana) OCAB for data on the production of banana, pineapple and mango; the National Office for Rice Development (ONDR, from the French name *Office Nationale de Développement de la Riziculture*) for the areas, production and price of paddy rice; and the Customs Department for the volumes of imported or exported agricultural products.

In 2014, MINADER'S Technical Working Group (TWG) tasked with checking the validity of agricultural data under the CountryStat country program ceased operation. However, whenever data is collected, it is validated at the regional level between MINADER and ANADER. Meetings are regularly held to discuss the reliability of the data generated between the INS, MINADER, ANADER, the ONDR and other state institutions responsible

for generating statistics on the cash crops. The data is then sent to the INS for the compilation of national accounts.

1.4 Structure of the report

This final Technical Report have 10 main chapters. Following the Introduction, Chapter 2 provides an overview of the previous seven Technical Reports published by the Global Strategy under this research topic. Chapter 3 summarizes the definitions and data sources relevant to administrative agricultural data. Chapter 4 examines the benefits and challenges of working with administrative data. Chapter 5 reviews how developing countries currently collect and use administrative data, highlighting areas where improvements have been made. Chapter 6 discusses the various uses (statistical and non-statistical) of administrative data. Chapter 7 outlines the legal issues to consider when sharing data across different agencies, countries, and other non-governmental institutions. The costs for setting up an Administrative Data System for Agricultural Statistics are the focus of Chapter 8; sources for funding and strategies for maintaining a system are also presented. Chapter 9 provides start-up strategies to improve existing administrative data systems. The final Technical Report concludes with a long-term strategy for the construction of more modern administrative data systems on the basis of statistical registries.

2 Methodology (research approach) and overview of previous publications

This Chapter summarizes the overall methodological approach adopted by the research project and the results previously obtained. Each of the seven tasks has been covered in a Technical Report submitted to the Global Strategy by the SSP and the ISU.

Task 1 consisted in the review of relevant literature and studies on the quality and use of administrative sources for the production of agricultural data. The first report, Technical Report 1 on *Reviewing the Relevant Literature and Studies on the Quality and Use of Administrative Sources for Agricultural Data* (GSARS, 2015a), reviewed the available literature and studies on the quality and use of administrative sources for agricultural data in developed and developing countries. Task 1 was essentially planned to identify best practices and lessons learnt that may be adapted to developing countries, and then to review the literature on the administrative sources of agricultural statistics from developing countries to identify any gaps. A conceptual framework was proposed to assess the quality and use of administrative data in agricultural statistics, and was later used in Task 3 of the project.

Technical Report 2, titled *The Role of Administrative Data in Developed Countries: Experiences and Ongoing Research* (GSARS, 2015b), covered **Task 2** and focused on country experiences and ongoing research in developed countries, with the aim of formulating lessons for developing countries. The discussion conducted under Task 2 demonstrated that administrative data have multiple uses in developed countries and have the potential to improve the quality of the final statistical product in several respects. The use of administrative data to improve a statistical product that is also based on survey or census data differs from the primary use of administrative data in developing countries, which consists mostly in use as an alternative to survey or census data or as a basis for the construction of sampling frames.

Task 3, detailed in Technical Report 3 (GSARS, 2015c), which provided a *Critical Analysis of Agricultural Administrative Sources Being Currently Used by Developing Countries*, evaluated the strengths, weaknesses and suitability of administrative data for use in agricultural statistics within an integrated and cost-effective agricultural statistics system, based on country assessments, a survey of the leadership of statistical offices in Africa and other documentation. The challenge was to identify effective uses of administrative data in developed countries that may potentially be applied in developing countries. Relevant uses include those currently made in developing countries but which exhibit potential for improvement, as well as completely new uses – and sources. Technical Report 3 also addresses what adaptations, if any, would be required to transfer given methodologies from developed to developing countries. The Report also highlights good practices gleaned from selected developing and developed countries, in preparation for Task 4.

Technical Report 4, titled *Analysis of Agricultural Administrative Data Gaps and Ways of Improving the Quality and Use of Administrative Data Sources for Agricultural Statistics* (GSARS, 2015d), drafted to fulfill **Task 4**, presented a gap analysis, highlighting areas for methodological and institutional improvement and proposing possible solutions upon which the field test and desk analysis protocols could be based. Therefore, among other goals, the desk analyses were expected to provide developing countries with resources to shift the use of administrative data away from direct use only (as a substitute for surveys and censuses) towards using them to improve statistics based partly on survey or census programs. Technical Report 4 highlights the key issues, challenges and weaknesses of administrative agricultural data systems and posits recommendations to address these problems. Among the solutions suggested are the establishment of channels of communication where none exist, or strengthening those that are available; and reviving and maintaining the user-producer committees that provide the avenues for dialogue.

Under **Task 5 (**explored in Technical Report 5 detailing a *Strategy and Methodology for Improving the Use of Administrative Data – A Protocol for in-Country Testing* **–** GSARS, 2015e), a strategy and related methodology were formulated for creating a cost-effective agricultural statistics system that would combine sample surveys and administrative data sources. As anticipated above, the problems associated with the direct use of administrative data as the final statistical product were identified in the first four tasks and their respective reports.

Task 6 consisted in the selection of two developing countries with different typologies in which to conduct the in-country testing of the proposed methodology – the pilot, which consisted in a system for collecting, processing, analysing and managing agricultural administrative data. The ARDS of the United Republic of Tanzania was identified as one of the best agricultural administrative data collection systems in Africa, due to its consistency and coverage and the support received from the Japan International Cooperation Agency (JICA). Therefore, the ARDS was used as the basis of two of the pilot projects. The strategy adopted was to identify the improvements required, which were then piloted (see Technical Report 5). A similar system was set up in two villages in Cote d'Ivoire, which was selected to satisfy the requirement, stipulated in the Letter of Agreement (LoA), that the pilot was to cover one English-speaking country and one French/Portuguese-speaking country in Africa. The third part of the pilot project was conducted in Namibia. The first part of this project required meeting with statistical and administrative agencies to discuss existing data sources and identify potential uses. The second part consisted in an investigation of the feasibility of applying a measurement error model to obtain a single estimate of crop area from multiple survey and census sources.

Field protocol for in-country testing

The pilot in-country testing was guided by a Field Test Protocol (FTP) which set out details of the objectives, mandates (roles and responsibilities) and routines of the field testing. Details of how the pilot field testing was carried out are currently available in the Global Strategy's Working Paper No. 11 on *Improving the Methodology for Collecting and Using Administrative Data in an Agricultural Statistics System: Field Protocol for In-Country Testing*", last updated on 11 November 2015 (GSARS, 2016a). A Concept Note summarizing the requirements was also produced prior to commencement of the pilot.

Technical Report 7 (GSARS, 2016b) presented and discussed the findings from the field country testing. As explained in that Report, the system, strategy and methodology were revised following the implementation of **Task 7** and after discussions at the validation workshop held in Rome on 5 and 6 September 2016.

This report covers **Task 8** and synthesizes the findings discussed in the first seven reports as final Technical Report.

3 Definition and data sources

This chapter presents the definition of administrative data adopted, after briefly reviewing the many definitions available (see Section 1.2 of GSARS (2015a) and Section 1.1 of GSARS (2015b)). The typologies and common data sources of agricultural administrative data are outlined, and the findings from the field tests on identifying administrative data sources are discussed.

3.1 Definition of "administrative data"

Administrative data are derived from diverse sources, including government records, subjective reporting systems and private organizations. This implies the sheer complexity of the problem of defining administrative data.

Traditionally, several authors have 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 (Brackstone, 1987). The United Nations (UN, 2011) report on an internal Eurostat task force that in 1996-1997 examined ways to better coordinate work relating to the use of administrative sources across different domains of statistics. Sen (undated) provides an alternative definition: administrative data is distinct from statistical data when the specific identity of the respondent or data source is central to the use of the data.

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*[4].

This broader definition encompasses two large classes of administrative data. For the first class, data are measurements of well-defined farm entities arising naturally through participation in a program. Examples of this first type of administrative data include information collected through taxation and subsidy programs. The highly structured European systems make extensive use of this first type of administrative data in producing

[4]See www.adls.ac.uk/adls-resources/guidance/introduction/

official statistics. For the second type, an extension officer makes a determination based on his or her observations and expert judgment. This second class of administrative data is emphasized in this final Technical Report, because existing extension programs routinely collect information on agriculture in many developing countries.

Two terms are introduced to distinguish these two classes of administrative data. The term "traditional administrative data" (TAD) is used to describe data arising from measurements of well-defined farm entities that are required for participation in government programs or private organizations. In the context of agriculture, TAD includes information collected through taxation, regulatory processes (i.e. farm inspections), farm assistance programs (i.e. subsidies and insurance) and monitoring programs (i.e. livestock tracing systems). The term "Agricultural Routine Data System" (ARDS) describes the second class of data. The properties of these two types of data differ due to differences in the data collection methods. TAD is often considered more objective than ARDS data because TAD arises from quantitative measurements of well-defined entities.

In this section, two types of administrative data are distinguished, TAD and ARDS. In the rest of this report, the two types are treated as a single integrated category, and distinguished only when the characteristics of the two data types should be considered separately.

3.2 Sources of agricultural administrative data

In the context of agriculture, the administrative sources include: regular returns or reports by agricultural field or 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 from manufacturers and distributors; farm registers 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, (GSARS, 2015a) . This section provides an overview of administrative data sources, separated by collection types.

3.2.1 Soil information

Topographical maps and maps of soil characteristics are often maintained in administrative processes.

Box 3.2.1.1. The Natural Resources Conservation Service (NRCS) is the soil conservation service of the United States Department of Agriculture (USDA). The NRCS maintains the Soil Data Mart, a database of the soil characteristics of land in the United States of America. Information on soil characteristics from these soil maps can be used for stratification in surveys (see e.g. Goebel, 2009) and as auxiliary information in constructing estimates.

3.2.2 Crop insurance and subsidy programmes

Government assistance programmes generate administrative data. Subsidies and crop insurance can provide such information as the areas planted with particular crops (Carfagna & Carfagna, 2010). Access to the administrative databases of government subsidy and insurance programmes requires forming good working relations with the administrative agency in question (Prell et al., 2009). Expert reviewers have raised the issue that such sources are scarce in developing countries, although some reports (Roberts, 2005; Clay, 2013) note an increase in the insurance programmes being established in developing countries.

Box 3.2.2.1. The Integrated Administrative Control System (IACS) is a database generated for managing and controlling payments to farmers. The IACS contains information on crop areas on farms in subsidy programs.

Statistical offices in Denmark, Germany and Italy utilize the IACS database for various purposes (FAO, 2010).

Box 3.2.2.2. The USDA's Farm Services Agency (FSA) administers several agricultural programs, including subsidies and incentives to conserve land. Beckler (2013) describes how the National Agricultural Statistics Service (NASS) uses FSA data: "[t]he FSA is an agency within USDA and is tasked with administering a variety of agricultural assistance and conservation programs that provide price support, disaster assistance, loans, and other services to agricultural producers. The omnibus United States Farm Bill, generally renewed every five years, provides authorizing legislation to FSA for the programs it administers. FSA collects an abundance of information from agricultural producers on the various application forms required to participate in the programs. Some of these data and FSA's geographical information system data are used by NASS as administrative data (also called administrative records). NASS uses these administrative data in a variety of ways, including: (1) building and maintaining sampling frames, (2) as ground truth data for remotely sensed data, and (3) to supplement data collected on NASS's censuses and surveys."

3.2.3 Land registration and cadastral records

A cadastre consists of records defining the "extent, value and ownership of land" (Bins & Dale, 1995). Maintained by several European countries, cadastres are used for taxation purposes, and to provide precise descriptions and continuous records of land ownership. Land registration systems generally share several characteristics with European cadastres and contain a wealth of information on land use, including crop management.

Box 3.2.3.1. India has a decentralized statistical system in which tasks are distributed various ministries at national and state level. The land revenue administration system managed by state governments is a source of administrative data that can be useful in the compilation of agricultural statistics (Sen, undated; Goel, 2002). This resource consists of information on land use and crop management gathered by village-level accountants. Examples include crop areas, fruit orchards, irrigated areas and irrigation sources (Goel, 2002). In the land registration system, which covers 88 per cent of the crop area (Goel, 2002), data are tabulated directly from land records and registration information is used as a sample frame for surveys of crop yields and production (Goel, 2002; Sen, undated). The aim of the Timely Reporting System, a process whereby village heads collect data for a 20-per-cent subsample instead of all crop areas, is to accelerate data collection (Republic of India, 2013).

3.2.4 Government regulation and monitoring programmes

Government regulation and monitoring programmes, whether voluntary or mandatory, produce substantial administrative data. Regulatory activities include the monitoring of production processes, financial institutions and insurance practices, and the resulting administrative data are used by statistical offices in a variety of ways.

Data from the regulation and monitoring of agricultural production and consumption play a significant role in agricultural statistics: in some countries, landowners are required to register their land, and information may be derived from farm food-safety and health inspections and vaccination records. Systems, such as livestock/cattle tracing systems that monitor the births, deaths and movement of registered livestock, are becoming increasingly important as sources of administrative data.

Box 3.2.4.1. An example of a database generated to regulate an industry yet is relevant to agricultural statistics is Belgium's SANITEL. SANITEL is a relational database that was created to regulate the cattle and pig industry and contains a permanent inventory of the animals in Belgium. It provides a complete inventory of the counts and movements of cattle and pigs, and contains information on health status and the detection of antibiotics, hormones or contaminants. The database is managed by the Central Association for Animal Health, not by a statistical office. The information in the database is obtained from regulatory activities: "[e]very keeper of pigs is required to complete a health certificate showing the capacity of his holding. Subsequently, every three or four months, approximately, he has a visit from an approved veterinarian so that he can declare the type and number of animals actually present" (European Communities, 2003). Since 2002, Belgium has reduced the number of pig surveys from four to two, with a view to replacing the survey data with information from SANITEL to compile its gross indigenous production forecasts (European Communities, 2003).

Box 3.2.4.2. Namibia's Ministry of Agriculture, Water and Forestry (MAWF) collects extensive data on livestock from government-sponsored vaccination and monitoring programs. An annual Livestock Census, conducted as part of the annual vaccination campaign, results in an enumeration of livestock in communal and commercial agricultural operations. In addition, Namibia has a livestock tracing system that enables the monitoring of births and deaths as well as of the movement of cattle. Interestingly, Namibia is currently the only African country with a comprehensive cattle tracing system.

Box 3.2.4.3. In the Lao People's Democratic Republic, the Agricultural Statistics Yearbook (ASY) aims to provide information relating to crop production, livestock, fishery, forestry, irrigation and other important information to be used as the basis to monitor the progress of the implementation of the country's Agriculture and Forestry Plan, as well as to develop policies and plan for agricultural and forestry development. The Ministry of Agriculture and Forestry (MAF) continues to use the data from its administrative reporting system to monitor the government's development plan and to establish resource allocation, because this system provides timely data without requiring additional budget allocations.

3.2.5 Private-sector sources of data

Private organizations involved in agriculture, such as licensing or regulatory bureaus, grain associations, commodity associations, cooperatives, factories, slaughterhouses, distributors of agricultural inputs and agricultural extension workers affiliated to universities regularly gather agricultural information that may be used in official statistics (USDA, 2011).

Box 3.2.5.1. The Meat Board of Namibia and the Namibia Agricultural Union are two administrative agencies within the MAWF. They provide the Namibia Statistics Agency (NSA) with the data required to produce a monthly livestock report, which contains information on the number of livestock marketed and indexes measuring the magnitude of changes over time with respect to a 2010 base year (NSA, 2015).

Box 3.2.5.2. Keita and Chin (2013) cite a study conducted in Cabo Verde in which information from private organizations was critical because there were no consistent survey or census data. They explain that Cabo Verde "is an island country with irrigated agriculture and cash crops concentrated in a limited number of well-known zones" and that the local agricultural production culture fosters a system of farmers' organizations and cooperatives for certain cash crops.

A multidisciplinary team investigated the feasibility of filling the data gaps existing in the time series for the production of primary horticulture crops, irrigated crops and cash crops. It was soon evident that, with the exception of non-irrigated agriculture, there was no centralized survey system for estimating agricultural production: the primary information came from decentralized surveys, administrative data and expert knowledge.

The investigation consisted in: (i) a review of the relevant documentation, surveys and studies; and (ii) collection of data from administrative sources, regional offices, research institutions and experts. Several methods were compared to determine the range of estimates that could possibly be obtained from the available information, and a

specific method was developed to estimate horticulture production. A primary input to the horticulture production estimate was the information on seeds sold by the two main seed marketing companies. Other sources included average yield, obtained from a crop-cutting survey conducted by the statistical service of the Ministry of Agriculture, and estimates of seed quantity per hectare, which were computed by research institutions. The final production estimates were cross-checked with data from consumption surveys.

3.2.6 Reporting systems and expert opinions

Individuals who regularly participate in agricultural production and research processes naturally gain considerable expert knowledge. Subjective assessments by expert reporters provide information for statistical offices in many countries (Keita & Chin, 2013; Galmes, 2013; Hamer, 2013). These reporters include individuals involved in agribusiness, university research and administrative agencies (Hamer, 2013), who often possess expert knowledge of a particular domain of interest.

Systems for expert reporting are of particular interest to this report due to their prevalence in developing countries. In Africa, the agricultural reporting systems set up by ministries of agriculture can provide weekly, monthly, semi-annual or annual reports of plantings, production, crop conditions and weather. The collection of Routine Administrative Agricultural Data (RAAD) is often administered through the Ministries for Agriculture, Livestock, Forestry and Fisheries (MALFs) on a regular (weekly, monthly or annual) basis. Often, RAAD even provides data on the smallest administrative units, such as districts or villages.

Box 3.2.6.1. The Agricultural Routine Data System (ARDS) is a primary source of information on agriculture in the United Republic of Tanzania. The ARDS was developed by the Agricultural Sector Development Programme (ASDP) in consultation with many regions and districts, with the objective to meet the data needs for monitoring and evaluation the ASDP itself. Standardized data are collected at the village level and aggregated to the level of wards, districts or regions, and finally the country. GSARS's Figure 5 (2015e) shows the flow of data collection, processing, and aggregation in the ARDS. The PMO-RALG issued a letter to instruct each Local Government Area to operate an ARDS. Available data from the monthly ARDS report provide information on crops (area, yield, production and prices), plant health services, livestock slaughtered, meat inspections, and livestock products and health.

Box 3.2.6.2. Namibia's MAWF administers a questionnaire called the crop assessment checklist (also known as the cereal production checklist). This checklist enables the gathering of qualitative and quantitative information on weather and crop conditions. Farmers are asked to comment on weather and rainfall conditions, as well as on the likely impacts of infestations. They are also asked to comment on the affordability of the necessary farm inputs (i.e. seeds and fertilizers) and on the utility of government subsidy programs. Farmers provide information on the percentage of area planted to cereal crops and explain how the current year's planted areas compare with those of the previous season, as well as with those during typical seasons. Farmers are then asked to "indicate" probable production, relative to the previous season, for maize, millet and sorghum. For non-cereal crops and livestock, farmers are asked to assess current and future conditions relative to previous seasons. The remaining questions pertain to household food security and the availability of advice and interventions from extension staff and government assistance programs.

Box 3.2.6.3. Bhutan has a decentralized statistical system. Line ministries and agencies collect, compile, analyse and disseminate data while the National Statistics Bureau (NSB) is mandated to lead the country's statistical development and statistics dissemination. The NSB does not have a separate unit established for the compilation of agriculture statistics. It does not collect agricultural data; however, it estimates agriculture's gross value added using the statistical data from the Ministry of Agriculture and Forests (MoAF). Harvested area and crop production data are estimated annually through a sample survey that is conducted by the Department of Agriculture under MoAF. Livestock data, including that on fisheries, are collected through the annual livestock reporting system. Forestry

information is compiled annually from the administrative reporting system (Asian Development Bank, 2016).

Box 3.2.6.4. The statistical system of the Lao People's Democratic Republic is decentralized and involves several institutions, each with its specific assignment. The ASY is the annual publication of the Department of Planning and Cooperation of MAF that compiles agricultural data from the administrative reports. Most of the crop production data and other agricultural data series come from administrative reports in which the government's agricultural personnel assesses crop production by observing harvests and interviewing key informants (i.e. farmers and village heads) in their localities. The information presented in the ASY is mainly collected through the MAF administrative reporting system, in which village heads are required to report on crop production and area, livestock, fisheries, and irrigation to the competent District Agricultural and Forestry Office (DAFO) every six months. DAFO officials derive summary district statistics from the village reports and forward these to the relevant Provincial Agricultural and Forestry Office (PAFO). A copy of the DAFO report is also forwarded to the Provincial Planning Office, which may eventually bring it to the attention of the Lao Statistics Bureau (LSB), which is under the Ministry of Planning and Investment (MPI). PAFOs derive the provincial summary statistics from the DAFO reports and forward a copy of said statistics to the Center for Agricultural Statistics (CAS) at the MAF. These reports are also forwarded to the Provincial Planning Offices, which may eventually bring them to the attention of the LSB (Asian Development Bank, 2016).

3.2.7 Taxation data

Tax data often provide information on farm expenses. Taxation data have long been used in statistical processes (Nordbotten, 2008), providing information on individual and household incomes, business types and sizes, and changes such as migrations, as well as information on the start or end of commercial operations in years when census are not carried out. The various roles of tax data in producing short-term business statistics are described by the OECD (2015). Statistical offices in Europe, Australia, Canada and the United States of America make extensive use of taxation data when producing agricultural statistics: Canada, for example, plans to use tax data to measure farm expenses for most farms in its 2016 Census of Agriculture (Smith et al., 2013).

4 Importance of and challenges posed by administrative data in the context of agricultural statistics

This section reviews the importance and potential benefits of administrative data for agricultural statistics, considering (i) the reporting system and (ii) other administrative data sources. The challenges and limitations of working with administrative data are also discussed.

4.1 Main benefits

The use of administrative data sources to generate statistics presents several benefits.

According to Principle 5 of the Fundamental Principles of Official Statistics, adopted by the United Nations 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"* (UN, 2011).

Most National Statistics Offices (NSOs) have experienced budget cuts without receiving directions on how statistical programs were to be reduced while still maintaining a high level of data quality. Paradoxically, these budget cuts have occurred alongside a growing interest in data-driven public policy and advanced statistical methodologies. For instance, in several countries, there has been a rising demand for statistics on small population groups or areas at more frequent intervals than are possible with ten-year population censuses, especially in light of the diffusion of decentralization policies.

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

Two of the principal motivations to explore the use of administrative data are reductions in cost and respondent burden (Carfagna & Carfagna, 2010). With smaller budgets, statistical agencies must make use of all available data. Concomitantly, increasing nonresponse rates in surveys lead to concerns about nonresponse bias and to heightened efforts to reduce respondent burden. In addition to cost savings and a reduced respondent burden, administrative data have the potential to improve surveys and censuses in terms of timeliness, small-area statistics and, in some cases, accuracy of responses. These reasons for the use of administrative data in surveys are discussed in further detail below.

4.1.1 Cost savings

Reductions in cost make the use of administrative data attractive to official statisticians. It is no coincidence that attention and energies are increasingly devoted to administrative data when budgets for government statistical agencies are cut. The cost savings resulting from the use of the information that is already available within administrative data is a straightforward initial step. Traditional methods of data collection applied to the production of official statistics – such as direct mailing, telephone interviews and personal interviews – are expensive. Administrative data, in contrast, have no associated data collection costs because the data are, by definition, already collected for existing purposes. However, administrative data are not free. Evaluating and assuring the quality of administrative data also entail expenses. Nevertheless, the costs associated with the acquisition and maintenance of administrative data are often much lower than those of conducting a traditional survey or census, because the saving is spread across data collection and data maintenance (UN, 2011).

4.1.2 Reduced respondent burden

Obtaining information from an administrative source that would ordinarily be obtained from surveys and censuses reduces respondent burden. This is an attractive characteristic of administrative data, especially given recent increases in nonresponse rates. Filling out forms and providing data according to regulatory requirements is routinely expected of businesses and individuals. If information has already been provided to a government agency, it would counterproductive if it were required again by other agencies. Likewise, if information such as

change of address, birth and death has already been filed with one agency, it should ideally be updated automatically in other government databases. Data sharing, and cooperative data set construction and maintenance – which support the sound use of administrative data – can also be said to underlie efficient, open and transparent governance.

4.1.3 Improvements in the efficiency of macro-level estimators

Even if information at the microdata level is sparse or difficult to obtain, administrative data may improve the efficiency of estimators at higher levels of aggregation. If released on a regular time schedule, administrative data can improve the consistency of published time series (Carfagna & Carfagna, 2010). One procedure to improve the efficiency of large domain estimates that has been subjected to comprehensive study is calibration. In theory, the total from the administrative source at the level of the domain of interest is assumed to be known. If the quantity observed from the administrative source is correlated with the quantity of interest within the survey, calibration can lead to improvements in the efficiency of estimators.

4.1.4 Improvements in the quality of microdata

The collection of administrative data is often mandated by government programs. For this reason, administrative data may be more complete and of higher quality than survey or census data. Therefore, the use of administrative data to substitute or augment survey- or census-based microdata can improve the accuracy and completeness of the resulting microdata. Evaluations of any discrepancies between survey data and administrative data can be used to identify outliers and correct errors in either data source.

4.1.5 Small-area statistics

Since the number of records in an administrative database is often large compared to surveys and censuses, administrative data have the potential to improve small-area estimates and expand the level of detail in published reports. Holt (2007) comments that several NSOs have expanded their use of administrative data in response to the increasing demand for small-area statistics. This finding is echoed by the Asian Development Bank: “compared with censuses and surveys, administrative data … can be presented at various levels of disaggregation, such as by geographical location or by age and gender” (Asian Development Bank, 2010).

Statistics New Zealand (undated) reinforces the value of administrative data for small-area estimation in the context of business surveys: “[t]he use of individual business tax and other administrative data within the official economic statistical program has major benefits in terms of the range and depth of statistical outputs that can be efficiently produced without imposing additional compliance costs on providers.”

Galmes (2013) advocates the use of information from expert informants, in conjunction with traditional surveys and censuses, as a mechanism for obtaining more reliable small-area estimates. The author notes that surveys have the advantage of providing a scientific measurement of sampling variation. However, cost constraints often prohibit the construction of survey-based estimates at granular levels of geographic detail. A combination of sampling surveys and administrative data from expert assessments can meet decision makers’ demands for

agricultural data in small administrative areas, without sacrificing the scientific integrity of traditional survey methods (Galmes, 2013).

Keita and Chin (2013) echo this view, citing an increasing demand for small-area statistics as one reason that alternatives to surveys and censuses will continue to play a role in agricultural statistical systems. The authors argue that the increasing decentralization of government, planning, and monitoring systems leads to a heightened need for information at low geographic levels. Similarly, poverty monitoring and policy implementation "require an understanding of local conditions for implementing poverty relief measures" (Keita & Chin, 2013). Because the sample sizes in surveys are often too small to provide reliable information for small areas, alternative methods of generating agricultural data are necessary. These include "administrative sources, eye estimates by local informants, expert opinion or assessment, village surveys, or other subjective assessment" (Keita & Chin, 2013).

One way to integrate survey data with information from administrative sources to produce small-area statistics is through a small-area model (Rao, 2003). In this approach, the administrative data are covariates in a model with the survey information as the response. If the administrative data are reliable at the small-area level and correlated with the response variable of interest, then the estimates for the small areas that incorporate the extra information from the administrative source are often more reliable than estimates from the survey information alone.

4.1.6 Timeliness

Administrative data sources are often updated more frequently than surveys or censuses. The Asian Development Bank comments on timeliness as an advantage of administrative data: "[f]requent and quick compilation of statistics is possible so that these can be released earlier than through data collected through censes and surveys." (Asian Development Bank, 2010). In some countries, including the United States of America, the time lag between two consecutive agricultural censuses may be of several years, whereas agricultural operations submit tax returns and apply for subsidy programs at relatively shorter time intervals.

Keita and Chin (2013) consider demands for timely data as a strong reason to pursue alternatives to surveys and censuses, such as administrative data sources. They explain that "[e]ven in countries where censuses and surveys are conducted, these activities are likely to be implemented with multi-annual frequency. Governments are increasingly demanding more frequent data on a yearly or even sub-annual basis. This is particularly the case for certain areas such as prices, where economic volatility has resulted periodically in rapid and wide ranging fluctuations from past trends."

Similarly, Galmes (2013) mentions expert opinion assessments as a pragmatic means of obtaining more timely estimates. The NASS crop progress and crop condition reports are an example of rapid dissemination of information from qualified informants. "The power of the weekly report lies in the flexibility of asking over 3,000 potential questions each week, the panel of experts who provide data, and the fast processing time" (Hamer, 2013).

Rao (2003, Chapter 3) discusses the role of administrative data in obtaining timely estimates of population changes in non-censal years. “In the absence of population registers maintained over time (as in some Scandinavian countries), it becomes necessary to develop suitable methods of population estimation in non-censal years, exploiting administrative files that contain valuable demographic information related to population change” (Rao, 2003, p. 27). Examples of administrative files used in population estimation include registrations of births and deaths, military migration information, school enrolments and income tax returns (Rao, 2003, p. 30). When applied to obtain small-area estimates, such methods often rely on implicit assumptions that the dynamics of the small area match those of the larger area (Rao, 2003, p. 30). To reduce the need for such assumptions, the methodology with which small-area estimation is conducted has evolved to make extensive use of explicit assumptions through mixed effects models (Rao, 2003). In this framework, the administrative data are often incorporated as a covariate in the model.

4.2 Challenges and limitations

An Administrative Data System for Agricultural Statistics (ADSAS) was used to review the structural, conduct and performance issues affecting existing administrative data systems in developing countries. Using ADSAS, a critical analysis of the agricultural administrative sources currently in use by developing countries was conducted. The results of this review are fully summarized by GSARS (2015c), which presents the results of country assessments and an analysis of other relevant documentation on administrative sources that are currently used by developing countries, evaluating their strengths, weaknesses and suitability for use in agricultural statistics within an integrated and cost-effective agricultural statistics system.

In another report (GSARS, 2015d), the gaps in the administrative data systems of developing countries were identified, and areas for methodological improvement noted. This involved an assessment of the types of data actually collected as compared to what should have been collected, the methods and frequency of data collection, the flow of data, the tools used for data collection, the capture and reporting formats chosen (e.g. whether any questionnaires or manuals were provided), and the use of the data and information generated to improve the system. The study also assessed the use of such data in informing policy, the frequency of system revision, accessibility, and the adequacy (in terms of skills and competences) of the human resources available to manage the system. This also involved evaluating any training and supervision given in relation to the system. In addition, an institutional framework review was conducted on aspects including system management, the personnel who actually collects the data, how the administrative data systems were synchronized with the agricultural survey and census systems, and other data management and maintenance issues.

GSARS (2015b) highlights the key issues, challenges and weaknesses relating to the administrative agricultural data system and advances possible recommendations to address the problems. It concludes that “administrative agricultural data systems are often set up in most countries with the major institution being the ministry responsible for agriculture and then other parastatal and private sector organizations. The challenge pronounced in most countries is lack of coordination of these institutions responsible for the collection, compilation and management of agricultural administrative data as well as lack of coordination with the NSOs. This is made worse by the fact that there is no legal framework that guides the required institutional arrangements”.

Further, agricultural statistics systems in developing countries have inadequate financial resources to effectively and efficiently run them. This is partly due to very low provision of funding for agriculture in the national

budgets. Developing countries generally need to lobby government for greater budgetary allocations in the national budgets and mainstreaming the costs of administrative agricultural data into the budgets at various levels of administration. Better coordination of the institutions collecting agricultural statistics could minimize costs of collection and compilation.

Many of these limitations arise from the nature of the processes underlying the collection of the administrative data source. Unlike traditional statistical surveys and censuses, administrative data are gathered for purposes that differ from the objectives underlying a statistical operation. The data may be entered by individuals in non-controlled settings, without enforcement of the strict protocols that dictate the data collection processes of carefully implemented surveys and censuses. Challenges arising from the nature of administrative data can adversely impact the quality and limit the utility of the administrative data source. Challenges associated with a lack of standardization, a lack of control over suppliers' methods, resistance from the public, incomplete coverage of both the population and the concepts of interest, and difficulties associated with computation and methodology are discussed in subsections 4.2.1 through 4.2.5 below.

4.2.1 Lack of standardization

Several applications of administrative data to the production of official statistics involve integration of the administrative data source with either other administrative data sources or with surveys or censuses. Challenges in this integration process arise when different data sources employ different definitions or coding systems. For example, simple differences in the labels used to identify units in micro data may hinder the linkage of disparate data sources. The subsections below discuss how differences between (i) definitions of units (ii) definitions of variables and (iii) coding systems may impede the use of administrative data for official statistics.

The lack of standardization implies that inconsistencies exist among data collection forms. When processes are decentralized and not standardized, different data collection methods may use different formats, which may make it difficult to integrate data sources. Chapter 5 contains a discussion of how using automated data collection can help to address this challenge.

a) Definitions of units

In this context, the unit is the smallest reporting entity in the micro data file. Definitions of units in administrative files are generally driven by the particular function of the administrative agency in question. As a consequence, the administrative agency's definition of a unit may differ from the definition endorsed by the NSO. For example, according to the FSA, a single farm producer may be associated with multiple farms, as defined by USDA/NASS. Studies have shown that this inconsistency between definitions of units results in increases in response burden when the FSA's database, instead of the NASS list frame, is used directly as a sampling frame (Barboza & Harris, 2009; Beckler, 2013). Numerous individual farms may comprise a single operation; therefore, the unit in the administrative database may differ from the statistical unit.

Benedetti et al. (2010) consider the challenge relating to units to be of such entity as to make it the foremost concern facing statisticians. The large choice of units (family, agricultural holding, household, parcel of land, point, etc.) and the dependency on the availability of a quality frame of units are particularly pronounced issues. The differences between the definitions of units may limit the utility of the administrative source, especially for purposes that involve linking micro data.

b) Definitions of variables

Differences in the definitions of related concepts can lead to significant inconsistencies between administrative and statistical sources. For example, the definition of income for tax purposes may differ from that required by policymakers who wish to analyse the data in the context of a survey of income. Such differences can cause systematic deviations between the quantities derived from the administrative source and the corresponding quantities (or estimates thereof) obtained from surveys (UN, 2011; Wallgren & Wallgren, 2010; Carfagna & Carfagna, 2010; Brackstone, 1987). The definitions and contents of administrative records are sometimes changed without prior notice to users and without provision of a grace period in which the new and old definitions are reported simultaneously. In the absence of an overlap period during which data are collected according to both definitions, it is impossible to disentangle any real change that may be occurring from the effect of the revised definitions. The impact of changes in definitions are more pronounced when files are updated continuously, as may be the case with a register (UN ESC, 2007).

c) Differences in coding systems

Challenges that arise due to the use of different coding systems are closely related to those associated with differences in the definitions of variables and units. For example, the NSO may require a more granular coding system than that needed by the administrative agency. When merging sources with different coding systems, inconclusive situations arise when one code in the administrative source maps to multiple codes in the system used by the NSO. Possible mechanisms for handling this challenge are discussed in Chapter 7 (UN, 2011).

4.2.2 Lack of control

Because administrative data, by definition, are not collected by the NSO, the latter generally has limited control over how such data are collected. The time schedule for collection and release of the administrative data may be subject to legal and policy-related processes that are difficult to manage or change. Changes in policy may alter the availability, characteristics or quality of the administrative data, which can produce inconsistencies across time. In several types of administrative data, the data are entered by lay individuals, and not by trained data collectors. This relatively uncontrolled data collection procedure may lead to errors and variability in how data are entered. The subsections below discuss how (i) changes in policy (ii) legal and political constraints, and (iii) uncontrolled data collection procedures may affect the quality and usability of administrative data.

As with the lack of standardization, a lack of control over the data collection process may also lead to inconsistent data formats. The NSO does not have any control over data collection, and different data collection efforts can use different formats. Further, administrative agencies may change data collection protocols over time or may use paper questionnaires; these may result in processing errors occurring when the information is digitized.

a) Policy changes

Changes in the policies and funding levels of administrative agencies may affect the quality of administrative data and change the way administrative data are collected. These changes are often beyond the control of the statistical agency, especially the NSO, and may have an impact on the availability and accessibility of the administrative source (UN, 2011). Brackstone (1987) explains that “discontinuities in time series based on administrative records can be caused by simple changes in the coverage of a program, the introduction of an

incentive to join or leave a program, or procedural changes that affect quality or completeness of records." NSOs are advised to work with the administrative agencies involved to anticipate such changes and mitigate their effects (Brackstone, 1987; Iwig et al., 2013). This recommendation for protecting against the impacts of changes in administrative procedures is discussed further in Chapter 6.

In some cases, a change in policy may actually improve the quality of the administrative data. Brackstone (1987) cites an example in which a change to the method of collecting tax data improved the coverage of the tax register. Specifically, the introduction of the Child Tax Credit in Canada required low-income earners to file a tax return to claim the credit. This legislation improved the coverage of the tax register in Canada, and thus potentially any estimates computed on its basis.

b) Legal and political constraints

Access to administrative data may be limited by legal and political constraints, which may be in place for good reason, such as to protect the confidentiality of the individuals in the population (Brackstone, 1987). Often, it is necessary to prepare a detailed Memorandum of Understanding (MOU) explaining the NSO's objectives and the data required to meet those objectives, before the flow of data from the administrative agency to the NSO may be established. Developing a high-quality MOU is often an expensive and time-consuming process (Prell et al., 2009). Guidelines for elaborating an adequate MOU are provided in Chapter 9 of this final Technical Report.

One dimension of the constraints related to legal and political issues is that of confidentiality. NSOs are under an inherent commitment to preserve the confidentiality of the statistical data. The complexity of these requirements increases, however, when administrative data, which collected and maintained by other agencies, enters the picture.

Box 4.2.2.1. Statistics New Zealand (undated) elucidates the complexities associated with preserving multidimensional legal and political requirements: "consideration of the legal, confidentiality and privacy issues associated with administrative data are also very important for Statistics New Zealand. Not only must Statistics New Zealand meet the requirements of the Statistics Act 1975, it must ensure that the security and other requirements of legislation such as the Tax Administration Act relating to access and use of individual administrative data are also met."

c) Data collection procedures

The method used to collect administrative data is largely beyond the control of the statistical agency. For example, tax forms are generally completed by individual filers. In many developing countries, agricultural returns are written by agricultural extension staff, or even chiefs thereof. The statistics offices in the customs office or ministry for agriculture, respectively, will often have no control. These forms of data collection lack standardization and may lead to reporting errors and inconsistencies (UN, 2011). In some cases, bias may arise from program-induced incentives (Brackstone, 1987; Carfagna & Carfagna, 2010). While casual data collection procedures may be acceptable for the purposes of the administrative office, the objectives of the users of statistical products usually demand a higher level of consistency and accuracy.

4.2.3 Coverage of population and concepts

The administrative database may not cover the full population of interest and may not include all necessary data items. For example, the administrative database may only consist of participants in a certain government program (Carfagna & Cafagna, 2010). The administrative process may only collect data on certain items, or may aggregate detailed categories which are individually of interest to the statistics office (UN, 2011; Wallgren & Wallgren, 2010). Information may be missing for particular subdomains of the population of statistical interest. For example, when evaluating the possibility of a register-based census for New Zealand, one consideration was that educational qualifications gained abroad would not be included for new immigrants, and emigrants may fail to de-register. This may be especially problematic in countries with substantial international mobility (Bycroft, 2010). Due to such limitations, multiple administrative databases must often be combined to achieve a register system with sufficient coverage.

A related solution involves integrating administrative and survey data. Due to the importance of pooling multiple data sources to overcome under-coverage problems, methods and techniques for linking records and data sets constitute an important component of the literature on using administrative data to produce official statistics. The use of record linkage to address quality issues arising in the use of administrative data is discussed in Chapter 6 of this final Technical Report. Chapters 7 and 8 of Technical Report 2 provide an overview of the statistical aspects of probabilistic record linkage.

4.2.4 Computational and methodological challenges

Administrative data often reach the statistics office in bulk and require editing, imputation and integration with other sources before they may be used in the statistical product. The need to manage massive quantities of data leads to significant computational challenges. Linkage is often not straightforward, due to the fact that there is no one-to-one correspondence between identifying variables on distinct administrative or survey data sets. Procedures for cleaning data and merging multiple sources often rely on either explicit or implicit model assumptions. Errors that arise in data cleaning operations may propagate to the final statistical product. Processes to manage massive data sets and edit error-prone or incomplete data create both computational and methodological challenges.

a) Computational challenges

Increases in the capacity and speed of computing environments increase the appeal of solutions utilizing massive quantities of administrative data. Although computational resources continue to expand, managing large volumes of data from disparate sources still poses significant IT and computing challenges.

b) Methodological challenges

Cibella and Tuoto (2008) report on a survey of the National Statistical Institutes (NSIs) of European countries that found that the vast majority of applications of administrative data to the production of official statistics involve harmonizing multiple data sources. A lack of standardization across data sources – for example, using different identifying variables – leads to challenges in merging disparate data sources. Desires to improve consistency across data sources through editing, or to improve the coverage of a register through imputation, lead to the need for such corrective practices. However, it is often difficult to develop adequate models to support the editing and imputation procedures.

When records are linked through probabilistic linkage models or through incomplete linkage, the estimation procedures should account for linkage error. However, quantifying the impact of this source of uncertainty on the mean square error of subsequent inferences poses a methodological challenge (Carfagna & Carfagna, 2010). A possible solution is the application of methods such as those proposed by Kim and Chambers (2012), which extend Chambers' regression analysis techniques (2009) to applications with more than two linked data sets.

4.2.5 Confidentiality and public perception

The issue of confidentiality is complex, especially with regard to administrative data. Consider two different scenarios, one in which some information in the administrative database is not protected by laws that ensure secrecy, and another in which the administrative information is confidential. In the first situation, the issue of confidentiality is not an obstacle to data sharing, from a policy or legal point of view. In the latter, an agreement between the statistics (often the NSO) and the administrative office is necessary to permit the access the administrative information. This permission may take the form of a MOU, a redefinition of the statistical system, or a government act or policy allowing the statistics office to access administrative data. These mechanisms for facilitating data transfer are discussed in Chapter 7 (GSARS, 2015b).

In the second scenario described above, individuals and enterprises may provide information to the administrative agency with the understanding that the reported information will remain confidential. As a consequence, the use of administrative data for statistical purposes may be met with scepticism from the public (Brackstone, 1987). To comply with this public concern for privacy, the statistics office is often advised to take ample measures to ensure the confidentiality of administrative data. For example, at Statistics Canada, administrative tax data are housed in a highly restricted and secure area. The need to accommodate public concerns related to privacy and confidentiality may increase the cost associated with administrative data and limit their accessibility.

4.3 Examples of challenges arising in developing and emerging nations

4.3.1 Uganda

In Uganda, there remain challenges with regard to the compilation of agricultural statistics from administrative records. First, farmers do not keep records on area planted, animals kept and production levels. Second, the quality and timeliness of the data is generally poor. Local-level financial and human resources to support administrative data generation are limited. For instance, the number of local governments compiling administrative data has been on a decline, although the Minister of Agriculture, Animal Industry and Fisheries (MAAIF) has been engaging in efforts to develop the capacity of the local government staff involved in generating agricultural statistics.

One of the key challenges facing the National Statistical System is the generation and utilization of administrative data. A large volume of administrative data is produced; however, it is of inadequate quality due to the following reasons:

- Poor data flow, due to unclear reporting mechanisms;
- Submission of incomplete returns or reports;

- Failure of some units to submit returns;
- Data may be collected but not used for planning purposes;
- Poor documentation of the data production processes;
- The reporting mechanisms of different sectors or institutions vary considerably, which delays the data collection process;
- The skills of the staff involved in data management are limited.
- High turnover of the professional staff; and
- Low level demand for agricultural statistics especially at lower levels of administration.

4.3.2 Côte d'Ivoire

Due to the political crisis afflicting the country, agricultural censuses were not carried between 2001 and 2015-2016. Therefore, the collection of administrative agricultural data is increasingly essential for the production of agricultural statistics, because the estimates based on the sampling frame and input from the 2001 census are obsolete. Administrative data are collected by 18 national institutions in the agricultural sector, including: the *Comité de Gestion de la Filière Café-Cacao/Conseil Café Cacao* (CGFCC/CCC), the *Association Interprofessionnelle de la Filière Coton* (INTERCOTON), the *Office d'aide à la commercialisation des Produits Vivriers* (OCPV) and the *Société d'Exploitation et de Développement Aéroportuaire et Météorologique* (SODEXAM) etc. Table 4.1 summarily presents the key stakeholders.

Several variables are covered, including the following: rainfall, maximum and minimum temperatures, land area cultivated and/or harvested, production, yields, agricultural products sold, prices, and quantities (produced, exported and processed).

Unfortunately, food crops (such as maize, millet and sorghum) are not covered in the administrative data collection conducted by the OCPV, with the exception of the monitoring of the prices of food crops (including rice). For rice, since 2008, the Department of Statistics and Documentation mandated the data collection to the ONDR.

Table 4.1. List of institutions producing agricultural statistics in Côte d'Ivoire.

No.	Acronyms	Names	Requested data
1	CGFCC/CCC	*Comité de Gestion de la Filière Café-Cacao/Conseil Café Cacao* (Management Committee of the Coffee-Cocoa Pathway/Coffee Cocoa Council)	Production, Export, Processing, Price, Distribution of exports by destination and by port of embarkation

No.	Acronyms	Names	Requested data
2	INTERCOTON	*Interprofession du Coton* (Inter-professional Association of the Cotton Sector)	Seed cotton production, Areas planted and yields, Cotton fibre production and yield, Producer prices
3	OCPV	*Office d'aide à la commercialisation des Produits Vivriers* (Office Assisting the Food Marketing of Products)	Prices of food in the markets
4	SODEXAM	*Société d'Exploitation et de Développement Aéroportuaire et Météorologique* (Company for Airport and Meteorological Development)	Temperatures and rainfall per station
5	OCAB	*Organisation Centrale des Producteurs Exportateurs Ananas Et De Bananes* (Central Organization of Producers-Exporters of Pineapple and Banana)	Production, Export, Areas, Yields
6	DCPE	*Direction de la Conjoncture et de la Prévision Économiques* (Directorate of Economic Conjuncture and Forecast)	Foreign trade of agricultural products for export and import
7	DGD	*Direction Générale des Douanes* (General Directorate of Customs)	As above
8	PALM-CI	Côte d'Ivoire Export Corporation of Palm Oil	Production of palm oil, Palmiste, Fine and cake
9	PALMAFRIQUE	African palm oil plantations company in Côte d'Ivoire	Production regimes (Industrial and village plantations), Areas (Industrial and village plantations)
10	MINEF/DISA/DPIF	*Ministère des Eaux et des Forêts* (Ministry of the Environment, Water and Forests)/Directorate of Information, Statistics and Archives/Directorate of Production and Forest Industries	Volume of wood export, volume of processed wood (lumber, veneer-slicing)
11	MIPARH/DPP	*Ministère des Ressources Animales et Halieutiques* (Ministry of Animal Husbandry and Fishery Resources)/Directorate of Planning and Programs	Animal products (national herd and production of meat and offal, milk and eggs by species-import of meat, offal and derivatives by species); Import milks, products and derivatives by species; Fishery products (import fishery products and derivatives)
12	ARECA	*Autorité de régulation du Coton et de l'Anacarde* (Cotton and Cashew Regulatory Authority)	Areas planted, Production of seed cotton, Yields, Production of cotton fibre, Purchase price to producers, Exportation
13	APROMAC	*Association des Professionnels de Caoutchouc Naturel de Côte d'Ivoire* (Association of Professionals of the Natural Rubber Sector in Côte d'Ivoire)	Areas planted (village plantations and research stations), Production rubber, Purchase price to producers
14	AIPH	*Association Interprofessionnelle de la filière Palmier à Huile* (Inter-professional Oil Palm Association)	Production regimes (industrial & village plantations), Total Production (industrial & village plantations), Production palm kernel (by framed structure)
15	SUCRIVOIRE	Operating Company of Sugar Cane Plantations	Area under Sugar production, Yields, Marketed production, Production of sugarcane (industrial & village plantations)
16	SUCAF-CI	*Sucrerie Africaine – Côte d'Ivoire* (African Sugar Producer – Côte d'Ivoire)	Area under Sugar production, Yields, Marketed production, Production of sugarcane (industrial & village plantations)
17	DOPA	Direction of Professional	Number of Cooperatives by region and speculation, List of

No.	Acronyms	Names	Requested data
		Agricultural Organizations	Cooperatives
18	ONDR	*Office Nationale de Développement de la Riziculture* (National Office of Rice Development)	Production, Yields, Areas, Price

In the vast majority of African countries, the collection of administrative data in general, and particularly that of relevance to the agricultural sector, is faced with major constraints. Several problems have also been reported regarding the quality of this data in Côte d'Ivoire.

In Côte d'Ivoire, the last agricultural census was conducted 13 years ago. Therefore, the forecasts and estimates made on the basis of agricultural census data are often of poor quality. In addition, the Ministry for Agriculture (MINAGRI) has not established a procedure to assess the quality of the administrative data. However, as part of the FAO CountryStat programme, since 2008 an annual workshop for the validation of statistics received by FAO has been held. This workshop is open to all stakeholders.

Following discussions with the Department of Statistics and Documentation, it has emerged that the main difficulties encountered in the production of agricultural statistics are due to:

- Inadequate or lacking material resources;
- Unskilled human resources in agricultural data processing;
- Lack of capacity-building programs for the staff;
- High turnover of agricultural statisticians or skilled staff, who seek better working and living conditions;
- Lack of the information required (especially in disaggregated form) and poor organization of data collection and archiving processes;
- Methodologies that are not always adequate (especially with regard to the food crops);
- Lack of a national strategy for the production of statistics generally, and for agricultural statistics in particular;
- An obsolete regulatory framework – decrees and orders have yet to be issued to implement the new legislation;
- Poor coordination of the many stakeholders involved; and
- Lack of national classification systems (sometimes, two different organisations transmit different values for the same variable for the same period).

4.3.3 India

a) Crop area statistics

- The main purpose of the Improvement of Crop Statistics (ICS) scheme is to monitor the performance of the primary reporting agency in the villages participating in the Timely Reporting Scheme (TRS) and in the Establishment of an Agency for Reporting Agricultural Statistics (EARAS). The ICS's findings

over a number of years reveal that there is a high degree of negligence in carrying out the *girdawari*, which thus casts doubt upon the reliability of crop area statistics.

- Another deficiency of crop area statistics arose with the development and modernization of agriculture, due to the fact that several new short-duration crops are now grown**.** Although the *patwari* (Village revenue agency) is required to undertake intermediate crop inspections between the two major seasons, this does not appear to take place regularly. Even if short-duration crops such as vegetables, flowers and mushrooms are covered in the crop inspection, they are not listed separately in the final crop abstract but rather grouped together under "Other crops".

- The major reason for the poor quality of area statistics is the failure of the *patwari* agency to devote adequate time and attention to the *girdawari* (complete enumeration of all fields). The *patwari* is overburdened with several functions and covers a large geographical area.

b) Crop forecasts

- The present system of crop forecasts is based mostly on subjective appraisals at various levels, and fails to reflect the ground situation correctly. This is especially the case with regard to preliminary forecasts, which must be fairly reliable if they are to act as the basis of policy decisions.

- The Directorate of Economics and Statistics, Ministry of Agriculture (DESMOA) suffers setbacks due to the failure to receive timely information from the states, and is often obliged to prepare these forecasts based on incomplete data.

- Frequent changes in the production figures – especially of food grains – between one forecast and another, with and the "final" and "fully revised" estimates cause confusion and doubt among users. When releasing these figures, DESMOA does not indicate the reasons for the changes.

c) Production of horticultural crops

- The pilot surveys conducted by the Directorate of Economics and Statistics, Ministry of Agriculture (DESMOA) are based on a sound technical methodology. However, the survey procedures are complex, time-consuming and rather difficult to implement in practice. Furthermore, the survey is limited to 11 states and its extension to the remaining states will be a lengthy process, due to the fact that many of these states do not possess the staff resources necessary to carry out the fieldwork.

- The estimates provided by the National Horticultural Board (NHB) relate to the entire country. However, they may not be completely reliable due to the fact that they are essentially based on subjective reports received from the ground-level staff. Indeed, there is considerable divergence between the estimates provided by the NHB and those computed by DESMOA for the states and the crops covered.

- Neither the NHB nor DESMOA provide estimates of the production of crops such as mushrooms, herbs and floriculture, which are however of emerging commercial importance (thus giving rise to coverage or completeness issues).

d) Land use

- The nine-fold classification of land use based on village records is inadequate because, for instance, it fails to provide information on characteristics such as social forestry, marshy and waterlogged land, and built-up land, which are nevertheless important for local development plans. It is established that the twenty-two-fold classification cannot be introduced in the village records.

- In most cases, the *patwari* cannot identify the characteristics of the various categories, nor can it manage the heavy burden that this work imposes.

e) Irrigation statistics

- The absence of a sound database for the minor irrigation sector has made it necessary to conduct a periodical Census of Minor Irrigation Works throughout the country, under the scheme of Rationalization of Minor Irrigation Statistics (RMIS). The primary fieldwork required under the census is entrusted to the *patwari* and the village-level worker (of the CD block), which operate under the supervision of block-level officials who also perform a five-per-cent sample check in randomly selected villages.

- The results of the sample check are used to apply a correction factor to the main census data. Data validation takes place at the district level and further compilation and tabulation work is done at the state level, with the help of software provided by the National Informatics Centre.

- The Central Water Commission (CWC), which is the nodal agency for water resource development in the country, is responsible for statistics on the water resources pertaining to major and medium irrigation projects. The River Management Wing of the CWC is engaged in hydrological data collection relating to all the important river systems in the country.

- Statistics compiled by the CWC on major and medium irrigation projects and those compiled by the Minor Irrigation Division – especially the irrigation potential created and actually being utilized – are alternative sources of the estimates of the total irrigated area.

- There is great variation between the statistics on the "area irrigated" published by DESMOA and the "irrigation potential utilized" published by the Ministry of Water Resources. Both data series are available with a considerable time lag.

- The existing system of generation and dissemination of data in respect of major and medium irrigation projects does not allow for the real-time monitoring of inflows of water and its utilization through canals and the distributary system. Reluctance on the part of the states to provide the data, in view of their vested interest in the sharing of water, poses a further obstacle.

- A large volume of useful data is available from the CWC on various aspects of irrigation, without any statistical analysis. These data should be used by the statistical system for the ultimate aim of improving the management of water resources.

f) Agricultural prices

- Wholesale price data are received by DESMOA mostly through post, which entails a delay.
- The state governments generally use part-time reporters, who are not fully conversant with the connotations of the different terms used in price data collection and do not pay adequate attention to the reporting work.
- The main deficiency in the collection of price data is due to the large scale of nonresponse.
- There is no coordination among the state agencies concerned, nor an adequate supervisory check over price collection.

g) Agricultural market intelligence

- Although the data to be supplied by the market intelligence units are of great use, these units have ceased to be effective in discharging their functions, mainly due to a lack of proper direction and control of their activities.
- The staff strength of the units has been considerably reduced, thus exacerbating the performance problems.

h) Fisheries statistics

- The flow of data from states is problematic. Consequently, much delay arises in the compilation of national statistics. As far as the deep-sea sector is concerned, although only a small number of licensed vessels are in operation, the data on fish catch do not flow in a regular manner. It is necessary to put in place a proper reporting mechanism for this purpose.
- The data on fish production from the inland sector are collected by the state governments. The resources required for regular data collection are, notably, significant and the cost incurred is not commensurate with the actual volume of fish production. Inland fisheries pose several problems due to the vast and diverse nature of water sources. It is therefore necessary to develop a cost-effective methodology.
- The data on fish production from aquaculture, supplied by the states, similarly suffer from poor quality and become available only with a considerable time lag. The types of culturing methods are not reflected in the data.
- The data on fisherman population, fishing craft and gear are available from both the state governments and the livestock census, while data on workers engaged in fishing are also available from the population census. However, the data from these sources are not comparable, due to differences in the concepts and definitions adopted and their application across states.

- There is an apparent inconsistency between the value of the output and the export earnings, with the latter being much higher. An exploratory study is required to reconcile the discrepancy.

i) Forestry statistics

- The main drawback in the compilation of forestry statistics (as is the case of several other sectors) is the inordinate delay in the availability of data. Except for the area under forest cover currently assessed by the biennial RS satellite survey, all other published data are available only after long time lags. The Forest Survey of India (FSI) faces the problem of delayed transmission of data from the states, which tend to accord low priority to reporting work. Almost half of the states fail to provide the statistics in time, which delays compilation at national level.

- The forest sector's present contribution to the GDP is considered an underestimate and therefore inaccurate, because it fails to take into account several important items such as head loads of fire wood, wood used for power generation and ecotourism.

- There is a significant discrepancy between the area under forest cover as published by the FSI and that computed by DESMOA, mainly due to the differences in the concepts and definitions followed by the two agencies.

j) Agricultural inputs statistics

- With structural adjustment occurring in several countries, the production, marketing, and export/import of agricultural inputs lie principally in the private sector. This greatly hampers data collection efforts. Although a limited amount of data on fertilizers are available from the input survey and from publications of the Fertiliser Association of India, these are incomplete and are not available in an acceptable timeframe.
- The collection and compilation of data with reference to agricultural implements and machinery is limited to tractors and power tillers; in addition, this too depends only on the data supplied by the manufacturers. Very often, the information is incomplete and there is no scientific mechanism for collecting statistics in the area. Data on farm practices and farm management are not available, although they are very much required for an adequate understanding of farm practices.
- Although a large proportion of the statistics on plant protection, quarantine and storage flow into the headquarters of the Directorate of Plant Protection, Quarantine and Storage (PPQ&S), they are not being fully compiled. The data have also not been organized for effective long-term use. The PPQ&S does not enjoy sufficient statistical support.

4.3.4. Lao People's Democratic Republic

In Lao People's Democratic Republic, large discrepancies have been found between the data reported in the Agricultural Statistics Yearbook from administrative reports and the most recent Census of Agriculture (2010/2011), Risk and Vulnerability Survey (2012/2013), and Lao Expenditure and Consumption Survey (2007/2008). The possible causes of these discrepancies are differences in coverage, in the concepts and definitions applied, and in the timing of data collection. This suggests that there is a need to control for possible

sources of discrepancies by developing a system-wide set of standards, concepts and methods to harmonize data collection and processing.

The summary statistics are not validated as one ascends the administrative hierarchy; therefore, the size of the measurement errors cannot be estimated. It is not feasible to conduct an analysis at the household level, because the central office holds only aggregated statistics. For example, the profiling of households that are food-insecure and the identification of vulnerable areas cannot be accomplished with an adequate degree of accuracy (Asian Development Bank, 2016).

5 Improving administrative data in agricultural statistics

An Administrative Data System for Agricultural Statistics (ADSAS) was used to review the structural, conduct and performance issues affecting the existing administrative data systems in developing countries (see GSARS, 2015a, Chapter 2; GSARS, 2015c, Chapter 3). Data collected by separate agencies using different methods, concepts, or formats, and by staff with varying skills, are usually divergent, inconsistent and unreliable. In a number of countries, there is a shortage of qualified staff, and poor incentives and motivation to work and poor or no supervision. This results in poor-quality data being collected and compiled.

The solutions suggested involve the establishment of new channels of communication or strengthening those which already exist; and the revival and maintenance of the user/producer committees that provide avenues for dialogue. The institutions should work towards using harmonized tools, definitions and concepts and should provide appropriate training to their staff in charge of collecting, compiling and managing administrative agricultural data. A hierarchical training protocol to be used at all levels should be designed: at national level, officers should be trained to then provide training to regional officers; these, in turn, should train district extension workers. New technologies such as Global Positioning Systems (GPS), Personal Digital Assistants (PDAs), scanners and Remote Sensing should be introduced, and efforts should be made to train and equip staff with the skills necessary to use these technologies for data capturing, processing and analysis, compilation, storage and dissemination.

These and other related issues are discussed in this chapter. GSARS (2015b) has also advanced proposals seeking to fill the identified gaps in the use of administrative data in an integrated agricultural statistics system.

Based on these findings, proposals were formulated for conducting desktop analyses and in-field testing. The proposals for desktop analyses focus on reconciling data from different sources and proceed in two broad phases. The first phase requires coordination with relevant institutions to obtain, prepare and understand the various data sources. The second phase envisages the use of systematic statistical procedures to synthesize multiple sources of information in an objective fashion. The test countries selected for the desktop analyses were Namibia and the United Republic of Tanzania. The in-field testing (also known as field-based testing) focuses on organizational infrastructure, data collection methods, the capacity for enhancing statistical analysis procedures, and the effectiveness of data dissemination techniques. The implementation of the field-based testing involves a review of administrative data systems and sources in selected countries through visits to statistical offices, and tests the improved procedures proposed for data collection and transmission, supervision and coordination, with a view toward suggesting areas for improvement. The countries that participated in the field-based tests were the United

Republic of Tanzania and Côte d'Ivoire, thus ensuring the representation of both English and French-speaking countries. GSARS (2016a) sets out detailed procedures for the desktop analyses and in-field testing.

5.1 The results of the country assessment reports: an analysis

The 2014 Africa country assessment report concluded that overall, the continent is relatively weak in terms of the resources allocated to agricultural statistical activities and of the statistical methods and practices dimensions in place. On the other hand, its record is strong in terms of the institutional capacity and availability of statistical information dimensions. However, these are general analyses and do not concern agricultural administrative data specifically. It was therefore decided to request for the original data for Africa from the African Development Bank (AfDB) to carry out a more detailed review. Further, it was decided to carry out another, more specific study on agricultural administrative data during the African Symposium for Statistical Development (ASSD), which took place in Kampala, Uganda between the 12th to 14th January 2015.

For Asia and the Pacific, the Asia-Pacific Commission on Agricultural Statistics rated Australia, Japan, Mongolia and New Zealand as "Excellent" in terms of the Institutional Infrastructure Dimension (Prerequisites); the Resources Dimension (Input); the Statistical Methods and Practices Dimension (Throughput); and the Availability of Statistical Information Dimension (Output) (APCAS, 2012).

The In-Depth Country Assessment for Bhutan (Thinley, 2014) concluded that some of the major issues and challenges in the field of agricultural statistics could be viewed from different perspectives, such as from those of producers or of users. Producers mainly face challenges of poor coordination, lack of professional support and funding, while users face difficulties due to inadequacy, poor quality and irregular release of data.

a) *Poor coordination*
Multiple agencies within and outside the Ministry of Agriculture and Forests (MoAF) are involved in generating Renewable Natural Resources (RNR) statistics. For instance, the major agencies outside the authority of MoAF are the Department of Revenue and Customs (DRC), which falls under the Ministry of Finance (MoF) and is involved in the recording of trade statistics; the Natural Resources Development Corporation Limited (NRDCL), active in the recording of forestry-related statistics; the Food Corporation of Bhutan Limited (FCBL), which is involved in recording food-related statistics, especially with regard to imports, exports and food reserves. There are also numerous agencies within MoAF responsible for producing RNR statistics.

b) *Inadequate and poor-quality data*
The general experience is that the available data are of poor quality. This is attributed to the lack of professional full-time statisticians and adequate funds. The staff, currently involved in the generation and handling of the RNR statistics, are from non-statistical backgrounds and are also required to fulfil multiple mandates at their places of work. This means that on occasion, the majority of their time was spent on non-statistical activities. Further, in the absence of adequate government funding, it is difficult to train staff, and certain statistical activities cannot be carried out as required.

c) *Irregular release of data*
Owing to the lack of adequate funding, the timely release of data is greatly hindered. In the absence of regular funding support, the service depends on contributions from donors and development partners and may be suspended, should such support not be readily available. At times, the collected field data may become obsolete by the time that adequate funds and experts become available.

d) *Lack of professional manpower*
The activities concerning RNR statistics are coordinated by the RNR statistical coordination section (RNR-SCS), which is housed in the Policy and Planning Division of MoAF. The RNR-SCS itself does not have qualified statisticians, except for some professionals who have taken short training courses and acquired experiences on the job. Most field data collection efforts are performed by field extension officials, who are supervised by district RNR sector heads under the overall coordination of the relevant subject matter departments (agriculture, livestock or forestry). For almost all RNR data collection activities, the extension officials who serve as enumerators do not possess statistical backgrounds or skills.

e) *Lack of adequate funding*
In the absence of a strong legal framework on statistics and adequate funds, it is difficult for the Government of Bhutan to allocate enough funds for statistical activities.

5.2 Structural issues affecting ADSAS

In this analytical framework, issues of structural design refer to the relatively stable features of the administrative sources related to agriculture. These include: (a) the system's perceived mandate (aims, objectives and clientele); (b) the institutional home, organization and coordination of the sources; and (c) the nature of the commodities to be covered. A review was also conducted of how the administrative sources fit into the overall integrated food and agricultural statistical system (GSARS, 2015a). A number of weaknesses were identified in the ADSAS structure, including the organizations collecting the administrative data, their structure, the core data items collected and the staffing levels and qualifications. These weaknesses are outlined below.

5.2.1 Organizations that collect and manage administrative agricultural data

It was found that in most developing countries, basic agricultural administrative data (i.e. crops, livestock, fisheries and forestry) is collected and managed under the ministries of agriculture, livestock, fisheries or forestry. However, in many countries, parastatal organizations collect administrative data, especially on commercial or cash crops. Private-sector agencies or organizations also often collect and manage various forms of administrative data, especially following the restructuring policies adopted in many of these countries. Say on prices, marketing, inputs, etc. These agencies may collect and manage the data without any direct participation of the relevant NSO. Often, different concepts and definitions are used, which may lead to the data on the same item being different.

In this respect, India provides a good example of how to coordinate federal and state institutions (Sen, undated).

5.2.2 Structure of organizations collecting administrative agricultural data

A number of ministries, departments and agencies (MDAs) charged with the collection and management of data have staff at headquarters and in the field (extension staff and chiefs, or even enumerators). However, well-qualified staff often cannot be retained. In many developing countries, there is a lack of staff and low staff retention rates, mainly due to poor working conditions and incentives. In several countries, otherwise sound data collection systems have not been sustained. Examples are Uganda's "*Buganda*" and "*Outside Buganda*" methods and FOODNET market information systems. Often, such systems are donor-funded and cease operation as soon as donor funding ends. Also, approximately 15 years ago, all districts reported on a regular basis; currently, very few districts report as per the desired schedule (MAAIF Verbal Communication).

A second weakness is that the field staff are often poorly supervised.

Box 5.2.2.1. The Tanzanian ARDS is a good example of supervision, or backstopping. The Backstopping Team consists of two Monitoring & Evaluation (M&E) TWG members and one Regional Officer. Furthermore, all local government officers gather at regional town meetings and report on the progress made and challenges encountered.

The most significant issue in the collection and management of agricultural administrative data in many developing countries consists in the multiple frequent changes to the administrative structure itself. For example, in Uganda, there have been several changes to the number and boundaries of districts. Furthermore, in the 1980s, a shift occurred from reliance on the purely administrative chiefs to the semi-political local council leaders. The latter were not accustomed to collecting data. Similarly, the decentralization policy pursued, meant that the extension staff were no longer answerable to the central governments.

5.2.3 The core items covered and the geographical coverage

There is generally a lack of data on food crops, and coverage is often at the national and regional levels. Therefore, the lower administrative levels tend to be excluded, although they are important in light of the decentralization policies adopted in most developing countries. This section **also** examines the specific weaknesses relating to certain core items.

As far as the core data items are concerned, the key issues arising for crops are the gaps occurring in the methods of making estimates under mixed and continuous cropping, and the estimation of production for food and minor crops. Routine data on livestock and other core data items – such as forestry, fisheries, agricultural inputs, exports and imports – are often incomplete, obsolete, inconsistent and unreliable due to limited skills in data-handling and processing, insufficient resources and other reasons. A common set of variables to report on for each core item should be identified. A harmonized format should be used throughout the local government level, to enable aggregating the data from the lowest level of data collection to district, regional and national levels.

The gap analysis conducted, and the recommendations made to deal with these gaps, have been useful in developing a draft strategy and methodology for improving administrative data and approaches for using this data as an element

of a cost-effective agricultural statistics system and designing a protocol for in-country testing of the proposed methodology. The draft strategy and methodology are discussed in Chapter 9 of this final Technical Report.

The crop forecasts system is often based on subjective appraisals, and the forecasts are prepared on the basis of incomplete data. Changes to the production figures are made between one forecast and another, thus casting doubt on the credibility of the figures reported. Resolving the problem of conflicting crop forecasts and food security assessments requires a greater collaboration among the various actors, so that the differences in the estimates obtained from the different approaches can be synthesized and reconciled. It is necessary to develop better methods of combining data from the various sources of forecasting data in terms of the timing of forecasts, the use of statistical methods versus non-statistical sampling, and the use of different enumeration areas (Kelly & Donovan, 2008). Objective or scientific techniques of crop forecasting should be developed. For forecast purposes, Remote Sensing (RS) technology can also be used to develop reliable estimates of crop area and condition at various stages of growth.

5.2.4 Data management

As discussed in Section 3.3.3 above, administrative agricultural databases are often owned by the agencies which generate the data. Data is often disseminated in hard copies (publications) or CD-ROMs at workshops; summaries are posted on the agencies' websites. The use of administrative data for statistical purposes requires sharing the administrative agency's file with the relevant NSO. When sharing or disseminating data, sufficient metadata should be provided for the user to fully understand the variable definitions and data collection procedures. Metadata can also help to ensure that values are correctly allocated to the relevant variables. Metadata are at the heart of the management of the interpretability of indicators. A current limitation identified in the use of administrative data for statistical purposes is that metadata are often missing or, at best, incomplete. The use of consistent archiving protocols with clear and thorough metadata can enable the more efficient use of administrative data. The issue of data archiving is explored further in Section 6.3.4.

5.3 Conduct issues in the ADSAS

In the context of this study, the term "conduct" refers to the behaviour of the administrative systems. In a way, the conduct issues are closely related to the processes in place. The conduct or process includes the following: (a) the information provided by the administrative system (including data sharing agreements and processes); (b) the ICT used in the 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 (GSARS, 2015a).

As discussed in Section 3.3, there are weaknesses in the data collection methods used; the technologies used in data collection, analysis, management and dissemination; and funding, especially the issue of sustainability.

5.3.1 Methodology

Different methods are used in developing countries, including equipment-in-area measurement (eye estimates, measuring equipment, GPS equipment, etc.) and yield or production estimation (e.g. farmers' estimates and crop-cutting). In this respect, several weaknesses have been observed.

Box 5.3.1.1. As discussed above, Uganda operates a decentralized statistical system. Therefore, the agricultural statistics system is also decentralized and comprises several players, under the leadership of the Uganda Bureau of Statistics (UBOS) and the MAAIF. Other than the MAIIF Directorates and Departments, seven semi-autonomous bodies report to the MAAIF for policy guidance and collecting some data (mostly administrative), mainly for their own operations. These bodies are the following: the National Agricultural Research Organisation (NARO); the Cotton Development Organisation (CDO); the Uganda Coffee Development Authority (UCDA); the Dairy Development Authority (DDA); the National Animal Genetic Resource Centre Data Bank (NAGRIC & DB); the National Agricultural Advisory Services (NAADS); and the Uganda Trypanosomiasis Control Council (UTCC).

Between 1963 and 1965, the first National Census of Agriculture and Livestock (NCAL) was held to provide benchmark data on the structure of the agricultural sector. Since then, a number of agricultural data collection exercises have been carried out by UBOS and MAAIF and the autonomous bodies that work with the Ministry to generate data. These include follow-up surveys conducted in 1967/68 & 1968; the NCAL 1990/91; two follow-up annual sample surveys in 1991/92 and 1992/93; the National Livestock Censuses (NLCs) 2008; the Uganda Census of Agriculture 2008/09; and the Population and Housing Censuses (PHCs) of 2002 and 2014, as well as the Uganda Panel surveys that have included an agricultural module. However, currently, very few agricultural statistics are collected on an annual basis at national and subnational levels.

There is no reliable and documented method of data collection at district level. There is also no clear data collection infrastructure. The district officials are not obliged to work for MAIIF, nor to provide it with reports.

The Fisheries sections of these districts established Beach Management Units (BMUs) as well as Fish catch forms (Form I BMU, Form II Parish, Form III Sub-county); however, implementation is still inadequate. UBOS designed enumeration areas throughout the country; however, these are not being used in the data collection exercise (UBOS, 2007).

Data collection was usually assigned to the extension officers, parish chiefs and local council officials without any facilitation. This renders the collected data completely inaccurate. Whenever data was collected, it would be stored as hard copies; only occasionally would it also be stored in digital format.

This data was analysed infrequently, and, even then, the analysis would be basic and conducted in simple Excel formats. Nevertheless, this is the data reported to have been used for planning and reporting purposes.

For the future, it is recommended that UBOS and other countries support the development of administrative data as a reliable source, through standardization of the data collection instruments and ensuring continuous coordination with the respective ministerial agencies. Under the Integrated Framework for the Development of Agricultural Statistics, there is a proposal for the development of a Village Registration System and an Agricultural Reporting Service.

5.4 Performance issues

In the context of the ADSAS, this analytical framework examines 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 terms of three aspects: (i) financial support; (ii) user support; and (iii) cost minimization (GSARS, 2015a).

5.4.1 Quality assessment

In developing countries, quality assessments for agricultural administrative data systems are rarely conducted. Most ADSAS in developing countries appear to not emphasize the documentation of agricultural data quality

parameters. The general impression, however, is that the quality of most administrative agricultural data requires improvement.

Table 5.1 illustrates the quality dimensions that were considered relevant to administrative data and how each of them can be measured and assessed. Where possible, efforts were made to make these measures quantitative.

Table 5.1. Measures for assessing quality.

Dimension	Description	Evaluation method
Relevance	The degree to which the available statistics meet the needs of current and potential users. This dimension also covers methodological soundness and the extent to which the concepts used reflect user needs.	Ascertain the interpretability of the data. Do the users readily understand the data? Hold focus group discussions with stakeholders. Which administrative data items are understandable or useful? Are there clear definitions of concepts, target populations, variables and terminology, as well as information describing the limitations of the data? Each administrative data set is accompanied by metadata on its contents, so that users can assess the data set's suitability for their purposes.
Accuracy / Representativeness	The closeness of the statistical estimates to the true values. The data should correctly estimate or describe the quantities or characteristics being measured. Accuracy may also be described in terms of the major sources of error that potentially cause inaccuracy (e.g. coverage, sampling, response, nonresponse). The data should adequately represent the entire population and relevant subpopulations.	The data should be produced in accordance with appropriate standards, classifications and practices. If sampling is carried out, it is necessary to ascertain whether it adheres to a standard sampling scheme. The percentage of eligible respondents that have not been included in the records should be determined. The data collected using different collection modes should be compared (the experiments proposed will also inform the level of accuracy). The coefficients of variation (CVs) must be computed. The administrative data should be compared to survey or even census data, whenever such data becomes available[5].
Accessibility Confidentiality and privacy protection	Accessibility should be considered in terms of accessibility to the final data users. The term may also encompass the demand or the effective demand of the data.	The data can be readily located and accessed in multiple dissemination formats (paper, files, CD-ROM, Internet, etc.). Metadata is available that explains the

[5] Trant, 2010.

Dimension	Description	Evaluation method
		variables and units of measure. Summary reports and microdata are available and can be accessed for research purposes. The number of data users and their frequency of use is known. Clear information is available on where to obtain the information, how to order it and its delivery time. The pricing policy is clear and convenient marketing conditions (copyright, etc.) are in place.
Coherence / Consistency	Data from different sources – and in particular from statistical surveys of a different nature or frequency – may not be completely coherent, in that they may be based on different approaches, classifications or methodologies. Such data may not, therefore, convey a completely coherent message to users (for example, users may be confused if two different measures of the same variable are published with different values).	Data comparisons and the linkages between administrative data and survey data may be considered as a possible criterion for evaluating administrative data consistency in the agricultural statistical system, as can the analysis of administrative data series. ✓ Compare the data from administrative sources with censuses and survey data. ✓ Compare the approaches, classifications and methodologies used in administrative data collection and analysis with those used in censuses and surveys.
Timeliness / Punctuality	This dimension refers to the continuous and consistent diffusion of information to stakeholders when it is needed.	Measure the length of time between the data being made available (date of publication) and the event or phenomenon that they describe. Ascertain the time lag between the date on which the data were actually released (date of publication) and the target release date (often preannounced). How frequently are the data updated – how often and at what time points? Ensure that the reference period is clearly specified so that appropriate adjustments can be made if the administrative data is to be integrated with surveys. The data should be available to the NSO when needed.
Disaggregation	This dimension examines the breaking down of information into smaller subpopulations and the examination of how they behave (seeking to identify patterns and trends).	The data can be stratified by sex, age and major geographical or administrative regions.
Comparability	This dimension focuses on the validity of comparisons between administrative sources; and on the validity of comparisons over time and space within a single source.	The same characteristics of the data should be compared between different administrative sources.

Dimension	Description	Evaluation method
		This should also occur within the same source, over time and space.

The above table can further be refined to devise data quality indicators. Examples of proposed data quality indicators are shown in Table 5.2 below.

Table 5.2. Examples of data quality indicators.

	Data quality indicators	Definition
	Completeness and timeliness of district reporting	Percentage of expected monthly district reports received within the specified period
	Completeness and timeliness of village/ward reporting	Percentage of expected monthly village/ward reports received within the specified period
	Completeness of concepts	Percentage of core data items on which data is not collected, e.g. in the ARDS
	Comprehensiveness of village reporting (ensuring absence of zero or missing values)	Percentage of monthly village reports that are not zero or that do not have missing values (for an average of, for example, five indicators such as crop area or crop production, or yield). Core data items should be defined and should be reported upon by all villages/wards. These items should then be reported to districts and the reports compared against this list.

The use of automated data collection procedures can also improve the overall performance of the ADSAS. Such methods can improve consistency over time and space and reduce processing errors. In addition, the use of consistent identifying variables and a systematic method for archiving data is key to improving the usability of administrative data. Digitizing and automating data collection is important for both TAD and for routine agricultural data systems.

5.4.2 Data quality control

The basic approach to achieving quality administrative agricultural data is to ensure that quality controls are embedded in every step of the statistical value chain. In addition, data quality checks should be instituted at every level of supervision and observed. The data quality control systems and support supervision should be adequate, and reports submitted on a timely and regular basis.

The field extension staff should be monitored and supervised to ensure that reliable data is actually collected from the farming households and transmitted regularly through the system. An example is India, in which a scheme for the ICS has been established. The scheme employs full-time staff for field supervision tasks, which reduces the occurrence of non-sampling errors and improves the system's efficiency.

Similarly, data quality should also be inspected at the national level, which should deliver guidance to the regional and district levels. If necessary, the national level should adapt the documents such as questionnaires/forms and training guides, so that data quality is uniformly improved across the country.

It is very important to institutionalize quality controls, and NSOs must prioritize the certification of administrative data obtained from other agencies for official use. In addition, NSOs should ideally establish an "audit" section (linked to the standards or methodology division) that can conduct a post-data collection assessment to identify areas for improvement or ascertain instances of non-adherence to relevant standards. This provides valuable feedback for the further improvement of both the data series and the standards. Quality frameworks should also be put in place and should define the concept of quality, providing criteria against which the quality of the outputs can be evaluated and certified. Peer-review mechanisms could also be an acceptable means to accredit the data produced (African Statistical Community, 2016).

Prior to publication, it is important that a highly structured approach to the final approval of the reports or publications be established. In particular, there must be a clear trail of the members of staff who have approved the publication. In this respect, it is useful to develop a clearance procedure for each publication, within which background information – on the sample, response rates, comparison with other data sets, etc. – should be provided before the report or publication is approved.

Monitoring processes should be in place to provide information to managers and supervisors on the quality achieved for the ongoing administrative data collection activities (e.g. response rates, editing rates and timeliness evaluations). Periodic reviews should also be undertaken to identify the steps necessary to maintain quality requirements.

The introduction of Computer-Assisted Personal Interviewers (CAPI) for data collection has enabled the significant reduction of errors and of processing times, and is overall more efficient. The entire team involved in administrative data collection efforts should be sensitized to and trained on all dimensions of data quality and on how a sound data quality may be achieved.

Technology can help to improve the field monitoring system. In particular, such technology can be used to monitor enumerators in the field from headquarters, in addition to the routine spot checks physically conducted by monitors.

5.4.3 Observations from the study

Box 5.4.3.1. In 2012, a study was carried out to rigorously assessing the United Republic of Tanzania's improved ARDS. The study focused on the relevance, effectiveness, efficiency and sustainability of the data collection system, and its strengths and weaknesses. The assessment also aimed to provide insights on what data are collected through the ARDS versus other data collection instruments, such as the multiannual agricultural sample censuses and the proposed annual agricultural survey. The assessment was conducted by three consultants with the oversight of the ASDP's M&E TWG, which is composed of officials of the ASLMs and Development Partners (DPs).

a) Relevance – appropriateness of the design

The ARDS emerged as a requirement for monitoring the ASDP. Before the ARDS, the data needs for project monitoring were met through the traditional data collection approach. The main contribution of the improved ARDS has been to systematize the data generation process. For monitoring the ASDP's activities, several

performance indicators (particularly output indicators) are obtained from the Village or Ward Agricultural Extension Officer (VAEO/WAEO) reports generated at LGA level. The ARDS design for monitoring the performance of the agricultural sector is appropriate. However, the implementation process presents some inherent difficulties. Regarding the questions of whether the ARDS is appropriate for delivering the data required by FAO and as inputs into the system of national accounts, it may be noted that the design of the ARDS is comprehensive in terms of coverage; however, the ARDS is not yet fully operative to provide the data at national level.

b) Effectiveness

One of the important aspects of the effectiveness of data collection system is data quality. The data collection methods in the ARDS present some inherent limitations compared to alternative methods adopted in sample surveys. For instance, the supervision of the data collection is inadequate. In addition, respondents tend to provide subjective responses to qualitative questions. For example, when a question is asked on a variable such as the area under a given crop in a village, for which there are no authentic records, the response may be subjective.

One of the recommendations is to adopt more intensive supervision. An example is India's ICS scheme, which, as mentioned above, employs full-time staff for field supervision tasks. Another recommendation is to improve record keeping on part of farmer groups and farm input providers.

c) Efficiency (cost-effectiveness)

Compared to other methods of data collection, the ARDS is less costly; however, in terms of data quality data, sample surveys enable better control. One of the main limitations is that it is not possible to associate objective measures of reliability with the results. This contrasts with sample surveys, in which sampling errors provide an objective measure of reliability. In the ARDS, data is collected from all the villages, and the errors associated with the results are non-sampling errors – which cannot be measured objectively. One of the general perceptions of the improved ARDS is that that the system is cost-effective and has facilitated data collection, uniform reporting and improved data accessibility. The recommendation is that, to reduce non-sampling errors and improve efficiency, there should be closer supervision during the data collection process.

d) Sustainability

The sustainability of the ARDS is linked with its relevance, effectiveness and efficiency. The improved ARDS has placed the routine data collection efforts on a systematic track. However, it remains at the initial stage. It is well-designed to the purpose for which it was initiated. In terms of resources, the minimum resources necessary for data collection are to be maintained. In the long run, the system must be sustainable with the internal resources available. It is recommended to expand the scope of the ARDS to cover all administrative units in the country. Specific country examples are given in Boxes 5.4.3.2. – 5.4.3.5.

Box 5.4.3.2. United Republic of Tanzania – ARDS

Based on the observations and discussions explored in the text above, key recommendations regarding the improved

ARDS are made on the following aspects (Statistics, 2012):

i) Data Collection Methods (Data collection for VAEO format); ii) Data Flow (From VAEO format to LGMD2) and Accessibility; iii) Data Quality and Reliability; iv) Data Management; v) Resources (Funding/Budget); vi) Human Resource and Capacity Building; vii) ARDS Data Use; viii) Sustainability; and ix) ARDS-Future Perspective.

- In view of an integrated approach to censuses and agricultural annual sample surveys, the role of the ARDS will have to be refocused. The number of items in the VAEO/WAEO formats may have to be reduced.
- The ARDS is to be maintained for routine administrative monitoring purposes. In view of the agricultural statistical system's refocused format, the ARDS may be used to provide early warnings of crop conditions, food shortages, surpluses and other anecdotal information that could develop or improve the master sampling frame.

Once the annual agricultural sample surveys are fully operational on a regular basis and provide reliable estimates for core crops and livestock products at national and regional levels, the ARDS data should help to generate district-level estimates using small-area estimation techniques.

Box 5.4.3.3. Uganda

For Uganda, the following actions are recommended:

- Establish a system for linking administrative units with NSO data collection units (enumeration areas);
- Recruit and train more data collectors, in addition to improving support supervision performed by higher-level officials;
- Standardize data collection instruments, and ensure continuous coordination with the respective ministerial agencies. Under the Integrated Framework for the Development of Agricultural Statistics, there is a proposal for the development of a Village Registration System and an Agricultural Reporting Service;
- Raise the visibility of the statistics unit (define minimum statistics staff structure across the NSS, identify high-profile and senior staff to advocate for the statistics function); and
- Improve and address capacity gaps at all levels of administration through a comprehensive statistical capacity development programme.

Production estimates continue to pose a significant challenge. Their computation continues to depend heavily on farmers' estimates, despite full awareness of the fact that the latter are consistently underestimated. The improvement of extension services could be accompanied by the introduction of a crop card system, or at least a recording of monthly production data on some main crops and livestock numbers.

Box 5.4.3.4. Côte d'Ivoire

To improve the reliability of the statistics produced by Côte d'Ivoire, MINADER's Department of Statistics and Documentation suggested the following actions:

- Define a strategy for the production and publication of agricultural statistics, including the human, financial and material resources required;
- Establish a discussion platform between the institutions involved in the collection, production and management of data;
- Promote producer or user workshops to ensure that the data collected meets the information required to develop the agricultural sector;
- Cover all relevant areas in the production of agricultural statistics;
- Promote the dissemination of agricultural statistics;
- Harmonize the concepts, definitions and methods used by the producers of agricultural statistics;
- Include ICT, new technologies (electronic data collection, Geographic Information Systems – or GIS – and Electronic Data Transfer) and use of the area sampling frame in the production and dissemination of agricultural statistics.

Box 5.4.3.5. Lao People's Democratic Republic

Since the country's administrative reporting system is known to be more prone to higher non-sampling errors (ADB, LSB & MAF, 2015) that cannot be estimated, the MAF and the Asian Development Bank (ADB) agreed that an audit sampling procedure should be introduced to improve the quality of data collected.

The proposed strategy to improve the administrative reporting system entails the performance of objective audit sampling and the rendering of results that could be used to assess the quality of data originating from the administrative reporting system.

Objective audit sampling requires that each sampling unit or, in this case, a village, be given a chance to be audited or selected. Operationally, this implies that villages should be selected using probability sampling techniques, such that the data from all sampled villages can be aggregated to provide provincial- and national-level estimates using appropriate survey weights. The staff of the Centre for Agricultural Statistics (CAS) of the MAF retrieves a copy of the sample village-level questionnaire or form and consolidates these copies into a data file, from which estimates at the national and provincial levels can be derived and compared with the results from the administrative reporting system. The CAS staff then closely examines the areas in which large discrepancies appear, and works with the relevant provincial offices to reduce specific measurement errors. To provide a sound basis for comparison with data from the administrative reporting system, the audit sampling procedure should also control for non-sampling errors.

To achieve these objectives, the following activities should be undertaken:

- Design a simple village probability selection approach that will constitute the core of the audit sampling. An effective sampling strategy is critical to the introduction of a sound mechanism for validating results from the administrative reporting system.
- Design a village questionnaire or standard reporting form that will allow for the better documentation of data gathered from the administrative reporting system.
- Determine the data items to be included in the questionnaire; to this end, major data users both in government and the private sector can be consulted.
- Apply standard definitions, standards and concepts to ensure consistency across time and space.
- Plan data collection activities that will control for non-sampling errors and will properly implement the probability sampling design. The major change expected to have a significant effect on the quality of estimates is the digitization of village-level data.
- Digitize the village-level data file; it should be noted that the questionnaire adopted must be designed for such a task, and that a data processing system that includes data validation and preliminary data analysis must be planned.
- Train the staff involved in data collection, processing and data validation.
- Train the respondents and key informants in sampled villages.
- Carry out an in-depth evaluation of the reliability of the audit survey, as well as a comparative analysis between the audit survey and the results of the administrative reporting system.

The report of the comparative analysis must be disseminated to the administrative reporting hierarchy, policymakers and relevant government agencies to gain support for efforts to make further improvement to the system. By raising public awareness of the "audit survey" and the use of probability sample surveys, there could be support within the Government to transition or combine survey results and the administrative reporting system for the ultimate purpose of improving the quality of the country's agricultural and rural statistics.

6 Uses of administrative data

Section 4.1 above reviews the many potential advantages of administrative data compared to survey or census data. In summary, administrative data are often collected at a high temporal frequency and in granular geographic detail. Because administrative data are collected naturally for non-statistical purposes, the statistical agency incurs relatively low data collection costs when using administrative data. These characteristics enable administrative data to be used for multiple purposes.

The review conducted by GSARS (2015a, 2015b) found that the uses of the administrative data can be classified into two broad categories: indirect uses and direct uses. In indirect uses, administrative data are used in forming or improving a statistical product that also utilizes survey or census data. Direct use refers to situations in which administrative information is used as the final statistical product for substantive purposes such as government planning. The statistical offices of developed countries make both direct and indirect uses. Developing countries are more likely to make direct use of such data, particularly when funding limitations motivate the use of administrative data as a substitute for survey or census data.

Because of the issues discussed in Section 4.2 and Chapter 5, the quality of an administrative data source should be carefully evaluated prior to use. Effective use of administrative data requires understanding multiple dimensions of data quality and "fitness for use" in the production of official statistics. These methodological issues should be addressed in practice when using administrative data as a direct source of information, or indirectly, in improving the overall statistical product.

This chapter details examples of direct and indirect uses of administrative data and provides methodological tools for addressing the quality issues introduced in Section 4.2. In Section 6.1, the uses of the administrative data in forming the statistical product are investigated. In Section 6.2, the concept of direct use is elaborated. Methodological tools to address quality are discussed in Section 6.3.

6.1 Uses in forming the statistical product

This section shows how administrative data can be used to improve the statistical product. Administrative data can be employed at all stages of the survey or census process, from sample design to estimation. Administrative data can be used to construct a sampling frame, identify ineligible units, or as auxiliary information in sample design. Use of administrative data as auxiliary information in estimation can improve the efficiency of estimators based on survey or census data. Although these uses are more common in developed countries than in developing countries, the concepts are generally applicable. To illustrate how these ideas transfer from developed to developing countries, this section provides examples of indirect uses of administrative data in both developed and developing countries.

6.1.1 Frame construction or improvement

Administrative data are often intrinsically linked to the identity of the individual unit in the target population. Many administrative sources are constructed pursuant to selective processes that define specific populations. Taxation data, for example, results from the process of gathering taxes and applies to the population of taxpayers. A single administrative data source can be used to define the frame.

A better approach than using a single data source to define the frame directly involves using the administrative data from multiple sources to construct or improve frames. This results in the improved coverage of sample surveys and censuses (Carfagna & Carfagna, 2010).

Box 6.1.1.1. ***Examples from developed countries***

Sweden. Statistics Sweden uses several sources of taxation information to analyse the coverage of their business register. These sources include administrative data stemming from VAT payments, "gross pay and preliminary tax based on statements of income," and "gross pay, payroll taxes, and preliminary tax from employers' monthly tax returns" (Berg & Hall, 2007).

Canada. Canada has a centralized statistical system in which Statistics Canada is responsible for the collection and dissemination of statistical information related to demographics, business, agriculture and other sectors. Canada's Statistics Act helps to facilitate the transfer of data from administrative agencies to Statistics Canada. At Statistics Canada, administrative lists have helped in the development of frames covering farms with small land area, such as chicken, egg, pig, fruit and vegetable farms. Such farms are difficult to capture in the absence of administrative lists (Trant & Whitridge, 2000).

Box 6.1.1.2. ***Examples from developing countries***

India. India and many other developing countries extensively use administrative records and other forms of administrative data to develop the sample frames for a wide range of activities, such as: small-, medium-, large-scale or commercial and institutional farms; livestock data, such as slaughterhouse records and vaccinations; agricultural inputs dealers or manufacturers; and exporters and importers. The earliest and perhaps most important form of administrative record use in Indian statistics regards land-use data, that are generated on a regular basis by the state land revenue administration. These data are compiled from village land records maintained by the village *patwari* (accountant). The land-use records are central to the entire process of agricultural production estimates for India. They are used as sample frames to determine where crop-cutting experiments should take place. The 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 (Sen, undated).

Malawi. As in many other African countries, the economy in Malawi is based largely on agricultural activities. Therefore, sound agricultural statistics are crucial in developing policies that can improve the country's economic health. However, reductions in the budgets of statistical offices limit the possibilities for conducting agricultural surveys with adequate sample sizes and regularity. This prompted an increased interest in the use of administrative data for producing agricultural statistics in Malawi. In Malawi, the introduction of a farm input subsidy expanded the options available for using administrative data in the context of agricultural statistics. Harmonizing variables on the survey with variables collected in administrative data sources has the potential to reduce respondent burden. Additionally, village-level officials "collect information on strategies, challenges, and weather parameters of the area, which is crucial information for designing national initiatives (Pangapanga et al., 2013)." The results of a focus group involving important stakeholders in the field of agricultural statistics in Malawi indicated that relative to surveys, administrative data typically fare better in terms of timeliness, improved funding predictability, reduced respondent burden, and lower costs.

6.1.2 Survey design

Efficient sample designs rely on information on the structure of the population of interest. Administrative sources are often critical in providing the external information necessary to design efficient samples. Two examples of sample designs that utilize auxiliary information are probability-proportional-to-size (PPS) sampling and stratified sampling. In PPS sampling, a size measure is defined for all units in the frame, and the selection probability is proportional to the specified size measure. If the size measure is correlated with the response of interest, then PPS sampling is more efficient than simple random sampling. Likewise, in stratified sampling, the population of interest is divided into groups called strata. If the strata boundaries explain variation in the population, then stratification can lead to efficiency gains. Fuller (2009) and Sarndal, Swensson and Wretman (2005) provide thorough discussions of the role of auxiliary information in sample designs. In the

examples given in Box 6.1.2.1., the necessary information on the population for the purpose of sample design is provided by an administrative data source.

Box 6.1.2.1. Statistics Sweden's use of tax data exemplifies the role of administrative data in survey design. Statistics Sweden uses tax data to define strata for a survey of the shares and assets of businesses. The population of interest is highly skewed, with a small number of units accounting for a large percentage of the population totals of the variables of interest. The stratification of the survey follows the total amount of the shares and assets recorded on the tax data (Berg & Hall, 2007).

6.1.3 Model-assisted calibration estimators

Auxiliary variables using information "encapsulated" in administrative data are often used in estimation as well as in design. The rationale underlying the use of administrative data in estimation is that administrative data may not meet the standards required of statistical data in some aspects; however, they have a sampling variance of zero and are often highly correlated with the quantity of interest to the survey. In calibration, the weights for sampled units are modified so that appropriately weighted sums of the auxiliary variable are equal to the administrative control. The term "control" is used to denote the fact that estimates of subcategories must match a predetermined total when combined, and this predetermined total is derived from sources external to statistical surveys or censuses. The stronger the correlation between the variable recorded on the administrative file and the survey variable, the greater the efficiency gain from calibration (Deville, Sarndal & Sautory, 1993). Thomsen and Holmoy (1998) provide examples and a discussion related to Statistics Norway's use of administrative data in calibration.

In some cases, administrative data do not provide information on exact quantities, but rather on ranges and inequalities. For example, an administrative total that represents a combination of more detailed categories provides an upper bound for the total of any one of the contributing categories. In such instances, the survey weights can be constructed to preserve inequality constraints or range restrictions, as determined by the administrative source. The example in Box 6.1.3.1. illustrates the use of customs data to define an inequality restriction.

Box 6.1.3.1. The United States Bureau of Land Management partners with the USDA/NRCS to obtain estimates of rangeland conditions through rangeland surveys. The 2012 rangeland survey aimed to assess the conditions of the greater sage grouse's habitat on bureau rangeland under three domains: greater sage grouse priority habitat, ecoregions and Western Association of Fish and Wildlife Agencies zones. At the estimation stage of the survey, administrative data on the area of rangeland in 13 western states were used as calibration controls in constructing weights.

The following administrative data was used:

- GIS layers defining the boundaries of bureau-managed land in survey-eligible states, from the Bureau of Land Management;
- GIS layers representing the joint work of the NRCS and the Bureau of Land Management by combining information on the spatial distribution of greater sage grouse breeding densities with the NRCS Common Resource Area geographic database;
- the United States Environmental Protection Agency's designation of ecoregion classes, based on Omernik (1987) level II and level III ecoregions; and
- GIS layers delineating sage grouse management zones developed by the Western Association of Fish and Wildlife Agencies, that reflected ecological and biological issues and similarities rather than political boundaries.

The estimation procedure began with the construction of weights for all points in the sample to obtain estimates of the acreage of bureau-managed rangeland in each combination of state, sage grouse habitat and

non-habitat, ecoregion and zone. Subsequent weighting involved the application of raking and successive ratio adjustments to preserve the three sets of control totals – state-by-type strata, ecoregions, and Western Association of Fish and Wildlife Agencies zones. At the end of the calibration, the final analysis weights are added to the administrative acres of bureau-managed rangeland in each Western Association of Fish and Wildlife Agencies zone.

6.1.4 Nonresponse adjustments and imputation

In surveys and censuses, the surveyed units may complete only part of the questionnaire or may refuse to respond to the survey. If the characteristics of non-respondents are systematically different from the characteristics of respondents, then estimators constructed with only the complete data may be biased due to the underlying population parameters of interest. Consider a survey intended to provide information on the average erosion rates of cropland. If farmers who employ conservation practices have higher response probabilities, then the estimates of mean erosion based only on the complete data are likely to be biased.

Administrative data may be available for both respondents and non-respondents. If a variable from an administrative database is observed for both respondents and non-respondents, and is related to the response variable of interest, then the auxiliary information from the administrative source may be used to evaluate and reduce the bias due to nonresponse. Comparisons between the means of the auxiliary variable for respondents and non-respondents may provide insight into the nature of the nonresponse. If the quantity recorded by the administrative source is correlated with the outcomes of interest, then the administrative data may be used as auxiliary information in constructing estimators that account for the nonresponse bias. Two broad methods to adjust for nonresponse bias, imputation and weighting, are discussed below.

a) Imputation

One mechanism that may be applied to adjust for nonresponse is the imputation of missing data. Imputation is especially useful for item nonresponse, i.e. when units complete only part, but not all, of the survey. Once a completed data set is created by means of a plausible imputation method, then it is possible to conduct several types of statistical analysis on the complete data. Kim and Shao (2013) and Sarndal and Lundstrom (2005) provide thorough accounts of the theory and methods of imputation.

To describe the imputation method, assume that a vector of study variable and an auxiliary variable (y_i, x_i) is collected in a survey. Let $\mathbf{A}$ be the set of sampled units, indexed by $\mathbf{A} = \{1,2,\cdots,n\}$ and δ_i be a response indicator that takes a value of one if unit *i* responds, and zero otherwise. Here, the auxiliary data $\mathbf{x}$ are obtained from an administrative data and are available for both respondents and non-respondents. For the sake of brevity, we assume a missing-at-random (MAR) condition on the response mechanism, such that

$$\mathrm{f}(\mathrm{y}|x, \delta = 1) = \mathrm{f}(\mathrm{y}|x, \delta = 0) = f(\mathrm{y}|x). \qquad (6.1)$$

The MAR condition (6.1) states that the responses are independent of the study variable $\mathbf{y}$ given the auxiliary variable $\mathbf{x}$.

Assuming that the sampling procedure adopted is non-informative (Fuller, 2009), the MAR condition holds for both the population and samples thereof. Once a conditional distribution of $\mathbf{y}$ given $\mathbf{x}$ and $\boldsymbol{\delta} = \mathbf{1}$ is estimated using the survey's respondents, plausible values for the non-respondents may be generated on the basis of the estimated imputation model given by $\hat{f}(\mathbf{y}|x, \delta = 0)$.

In practice, there are two population approaches when conducting imputation. Fractional imputation (Kim & Shao, 2013) provides a singly completed data set with multiply imputed values on each missing unit. Multiple imputation (Rubin, 2004) generates multiply completed data sets, where each complete data set contains a singly imputed value for each missing unit.

b) Weighting

An alternative nonresponse adjustment mechanism involves modifying the weights to account for nonresponse. When the final estimator is fixed or predetermined, weighting may be more efficient than imputation. Calibration and propensity scores are two techniques for determining weights to adjust for nonresponse. In calibration, the weight is determined so that the mean of the auxiliary variables across sampled units is equal to the mean based on the administrative data. A propensity score is an estimate of the probability that unit *i* responds. Both methods require auxiliary information (Lundstrom & Sarndal, 2005), which may be derived from administrative sources.

Box 6.1.4.1. Geuzinge, Rooijen and Bakker (2000) describe the use of administrative records in constructing calibration weights to reduce nonresponse bias in household surveys. In one application, administrative registers of jobs and social security benefits were used to weight the respondents to the 1995 Netherlands Health Interview Survey. The theory was that individuals with greater health problems and greater use of medical resources were more likely to respond to the survey as a result of greater interest in health care processes. The concern of the statistical agency was that without adjustment, an estimate of medical use based on the complete data would overestimate the true cost of health care. Estimates of days in hospital that incorporate the administrative data were lower than corresponding estimates based only on the unweighted complete survey data. The weights were also applied to obtain estimates of education levels. The weighted estimates of the proportion of individuals educated beyond higher secondary level were lower than the unweighted estimate. The reason for this result was thought to be that individuals with higher education levels had higher response probabilities because they had a better understanding of the usefulness of the survey and greater trust in the Government.

6.1.5 Model-based small area estimation and forecasting

Many statistical procedures to obtain estimates for small areas or to forecast a future outcome are based on explicit models. For the case of small area estimation, population information at the level of the small domain of interest is critical for improving the efficiency of estimates. If the objective is forecasting or improving the timeliness of estimates, auxiliary information that reflects a more recent time period or changes over time has potential to reduce the mean squared errors of forecasts. In the construction of the 1997 National Resources Inventory (NRI) estimates, administrative data on transportation was used to create small area estimates of the area of roads (Nusser & Goebel, 1997; Wang & Fuller, 2003).

This section briefly discusses how administrative data can be used as auxiliary information in a classical small area estimation. See Technical Report 7 for a detailed illustration based on Namibia's agricultural survey and administrative data.

Rao (2003) classifies small-area estimation methods as either unit-level models or area-level models. The former model was initially introduced by Batteese, Harter and Fuller (1988), the latter by Fay and Herriot (1979). We consider the area-level model

$$\hat{y}_k = y_k + e_k, \qquad (6.2)$$

where k denotes the small area of interest, $\hat{y}_k$ is the survey-based direct estimator, y_k is the unknown true quantity, and $e_k \sim N(0, \sigma^2_{e,k})$ is a sampling error with the known variance $\sigma^2_{e,k}$. Efficiency gains are possible by

specifying a model relating to the true quantity y_k to the auxiliary controls such that

$$y_k = x_k'\beta + u_k, \quad (6.3)$$

where x_k is the auxiliary information obtained from administrative data and $u_k \sim N(0, \sigma_u^2)$ is an area-specific random effect with the unknown common variance σ_u^2.

Combining the two models in (6.2) and (6.3), an estimator of the best linear predictor of y_k is

$$\tilde{y}_k = x_k'\hat{\beta} + \hat{\gamma}_k(\hat{y}_k - x_k'\hat{\beta}),$$

where $\hat{\gamma}_k = \left(\sigma_{e,k}^2 + \hat{\sigma}_u^2\right)^{-1}\hat{\sigma}_u^2$, and $\hat{\beta}$ and $\hat{\sigma}_u^2$ are estimates obtained, for example, from the maximum likelihood or the restricted maximum likelihood.

The ratio of the mean squared error of the direct estimator $\hat{y}_k$ to the mean squared error of the predictor $\tilde{y}_k$ is approximately equal to $\hat{\gamma}_k$. The lower $\hat{\gamma}_k$ is, the greater the efficiency gain from the prediction model. In other words, if the true quantity y_k is already adequately explained by the administrative data, then large efficiency gains may be achieved by incorporating the administrative data in small area estimation.

Box 6.1.5.1. Battese, Harter and Fuller (1988) use satellite data as auxiliary information for small estimation of crop area and yield. In their estimation of the area of corn and soybeans in 12 Iowa counties, the auxiliary information in the small area model was the number of pixels in the county that were classified as corn or soybeans. The satellite data were highly correlated with the survey data; therefore, the small area models led to reductions in the mean squared error in the small area predictors.

Box 6.1.5.2. The use of small area estimation models to obtain subnational estimates of crop area in Namibia was explored. The domains were those administrative regions of Namibia that are primarily involved in communal agriculture. The 2013/2014 Namibia Census of Agriculture (NCA), a sample of communal agricultural holdings, provides the survey data. Data from Namibia's Crop Assessment Checklist, a routine reporting system administered by the MAWF, served as auxiliary information. The table below displays the survey estimates from the NCA, the estimates based on the MAWF's crop assessment checklist, and the minimum mean squared error (MMSE) predictors based on the small area model. The estimated CVs for the survey and small area models are also provided. Such estimated CVs for the MMSE predictors cannot be greater than the estimated mean squared errors for the survey estimators due to the additional information contained in the small area model and the MAWF's auxiliary data. The efficiency gains are modest at best, because the estimated variances for the NCA estimators at the regional level are relatively low.

Region	NCA (C.V.)	MAWF	MMSE (C.V.)
Zambezi	15,904 (0.128)	19,384	16,823 (0.122)
Kavango	51,302 (0.090)	21,588	49,999 (0.088)
Omusati	109,673 (0.051)	78,030	109,492(0.051)
Ohangwena	81,337 (0.051)	79,828	81,649 (0.051)
Oshana	40,021 (0.198)	35,100	41,600 (0.177)
Oshikoto	68,481 (0.043)	58,080	68,568 (0.042)

The annual agricultural surveys conducted by Ethiopia's Central Statistical Agency provided crop-wise area estimates at regional and zone levels; however, district-level estimates were not available. Meanwhile, the Ministry of Agriculture and Rural Development generated area estimates using an approach that was very similar to that adopted in 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 (United Republic of Tanzania, National Bureau of Statistics, 2012).

6.1.6 Cut-off surveys

In a cut-off survey, all or part of the questionnaire is not administered to a portion of the population. Instead, the required information is obtained from an external source. Census or administrative data often provide the auxiliary information necessary for successful implementation of a cut-off survey. Examples of the types of external data sources that have provided the necessary information for cut-off surveys include tax data and information from private corporations. In many countries, statistical agencies have applied cut-off strategies in surveys related to business establishments or energy, for example.

The design of cut-off surveys relies on an auxiliary variable that is known for the full population. A common approach to cut-off surveys begins with ordering the population of interest with respect to a measure of size. The size measure associated with a unit is often indicative of the importance of the unit to the overall estimate. For example, in surveys of business establishments or agricultural operations, the size measure may be related to total employment or farm area, respectively. In typical applications of cut-off sampling, units with a size measure lower than a specified "cut-off" value are not included in the sample. Cut-off sampling may be viewed as related to probability-proportional-to-size sampling, in which the size measure associated with certain units in the population is zero.

In cut-off surveys, reliable auxiliary information is also critical at the estimation stage. It is necessary to obtain surrogates for the responses to target questions of interest for units in the population that were not part of the data collection. These target variables are often derived from administrative sources that collect similar or related information. Concepts measured in the external data source may differ from the target variable of interest to the survey, due to differences in reference periods, coverage or definitions. In such cases, models may be required to calibrate the variables available in the administrative file to the survey's target concepts of interest.

Box 6.1.6.1. In the late 1990s, Statistics Canada used a cut-off survey design to reduce the burden on small business respondents. Businesses that were too small to contribute substantially to the overall estimate were placed in a "take-none" stratum, and tax data were used to produce estimates for these units (Yung, Rancourt & Hidiroglou, 2007).

Box 6.1.6.2. The statistical office of Slovenia uses a cut-off survey to improve the timeliness of estimates of monthly turnover indexes. In the population of businesses, the largest 3 percent of units accounts for more than 50 percent of total turnover. A classical questionnaire is administered to estimate turnover for the 3 percent, and tax data are used to estimate monthly turnover in the remaining units in the population (Seljak, 2007).

6.1.7 Use of administrative data in assisting data collection for surveys and censuses

Administrative data can also help to facilitate data collection processes in surveys or censuses. This is especially true when the specific characteristics or identity of the sampled units are unknown until contact with the sampled unit is established. One important case, especially for agricultural surveys, occurs in surveys that are based partly or entirely on area frames. In the example below, administrative sources provide lists with names and addresses that are useful in contacting units that were originally sampled from an area frame, rather than a list frame. Area maps are also useful in assisting data collectors that conduct surveys related to agriculture and natural resources.

Box 6.1.7.1. In some area frame surveys, the address of the sample unit is not available in the frame, and locating

sampled units for data collection may pose a challenge. An example is the Conservation Effects Assessment Project (CEAP), a series of surveys intended to measure different kinds of soil and nutrient loss from crop fields. FSA data have been used to identify potential farm operators, reducing the time and effort required to search for the operator associated with a given sampled point. The reliability of these data varies geographically. In parts of the country where these data are judged to be less reliable, information from additional sources was incorporated into the process to facilitate the contacting of sampled units.

6.1.8 Use of structural measurement error models to combine multiple measurements of related quantities

A popular way to form a single, improved estimate combining multiple data sources is to use a structural measurement error model. This approach specifies measurement models and structural models to describe relationships between several data sources. Given the model formulation, the parameters can be estimated jointly and predictors of quantities of interest can be constructed. Technical Report 5 covers the details of the methodology adopted for several data structure patterns.

This section briefly introduces a factor analysis (Fuller, 1987), an approach that can be used when three data sources are available. Factor analysis is discussed in further detail in the context of crop area data for Namibia in Technical Reports 6 and 7. Let the three observed data sources be denoted with Y_{i1}, Y_{i2}, and X_i. Assume that the three data sources satisfy the models

$$Y_{i1} = \beta_{01} + \beta_{11} x_i + e_{i1},$$

$$Y_{i2} = \beta_{02} + \beta_{12} x_i + e_{i2},$$

$$X_i = x_i + u_i, \qquad (6.4)$$

where x_i is the true quantity of interest and (e_{i1}, e_{i2}, u_i) are random error terms. The assumed mean and variance of the distribution of the true value x_i are functions of a lower dimensional parameter vector. Note that X_i is assumed to be an unbiased measurement of x_i, while Y_{i1} and Y_{i2} may be biased. The biases of Y_{i1} and Y_{i2} are represented in the regression parameters. These biases may arise from the issues discussed in Chapter 4, such as a change in the reference period or a subtle difference in the phrasing of a question. In a typical setting, X_i is obtained using a probability-based survey sample conducted under controlled conditions. The other two observations may be obtained from external data sources such as administrative data. Assuming that the covariance matrix of the vector $(x_i, e_{i1}, e_{i2}, u_i)$ is diagonal, the model parameters in the measurement model (6.4) are identifiable and estimable.

Box 6.1.8.1. As part of the pilot project conducted for Tasks 5-7, the feasibility of using a structural measurement error model to estimate crop area using several related data sources in Namibia was investigated. Namibia has two primary data sources on the planted area of major crops (maize, sorghum and millet). One source is a set of survey-based estimates obtained from the Annual Agricultural Surveys (AASs). The second source is administrative data obtained from the MAWF. The AAS survey estimates are treated as the unbiased measurement (X_i) of the true planted area. Two MAWF estimates, one for commercial agriculture and the other for communal agriculture, serve as the biased measurements, represented by Y_{i1} and Y_{i2} in the measurement error model above. We apply the measurement error model to obtain a single estimate of planted area that incorporates the information in the three sets of observations. This example illustrates that a further advantage of this approach is that the use of the model allows estimating a measure of uncertainty associated with the predictors. Details are available in Technical Report 6 (FAO, 2015f), while a related example is provided in Technical Report 7 (FAO, 2015g).

6.2 Uses as final product

Administrative data can be used directly as the final statistical product. In this case, the information contained in the administrative data source is used directly for substantive purposes such as policy, management, business decisions and farming decisions. Table 6.1 summarizes the uses of administrative data for statistical purposes in 13 African countries that participated in a non-probability survey related to the use of ADSAS. Most countries use administrative data for direct tabulation, frame construction and improvement, survey design, and crop forecasting. Only two countries use administrative data in formal statistical estimation procedures, such as calibration and imputation. This survey demonstrates that direct uses, such as direct tabulation and crop forecasting, are more common in many developing countries. This section discusses various ways in which administrative data are used directly as the final statistical product.

Table 6.1: Administrative uses of ADSAS: uses in constructing statistics

	BURUNDI	EGYPT	GHANA	LESOTHO	LIBERIA	LIBYA	MAURITANIE	MAURITIU	SOUTH SUDAN	SOUTH AFRICA	SUDAN	UGANDA	ZAMBIA	Total/13
Statistical uses														
Direct tabulation	0	1	1	0	1	1	1	1	0	0	1	0	1	8
Frame construction/improvement	0	1	1	1	0	1	0	1	0	1	1	1	1	9
Survey design	0	1	1	1	0	0	0	0	1	0	1	1	1	7
Model-assisted calibration estimators	0	1	0	0	0	0	0	0	0	0	1	0	0	2
Nonresponsive adjustments (weighting)	0	1	0	0	0	0	0	0	0	0	1	1	0	3
Imputation for missing survey data	0	1	0	0	0	0	0	0	0	0	1	0	0	2
Small area estimation	0	1	0	0	1	0	0	0	0	0	1	1	1	5
Forecasting	0	1	1	1	0	0	0	1	0	1	1	1	1	8
Survey data integration	0	1	0	0	1	0	0	0	0	0	1	0	1	4
Further reporting	1	1	0	1	0	0	0	0	0	0	1	1	1	6

Source: Survey of ADSAS 2014. 1=Yes, 0=No.

6.2.1 Direct tabulation

If administrative data are of sufficiently high quality, they may be used directly for the statistical product (Brackstone, 1987; Wallgren & Wallgren, 2010). Based on practices followed at Statistics Canada, Brackstone (1987) describes direct tabulation as the processes of counting units in files, cross-classification by attribute, and the aggregation of quantitative variables associated with each unit. Published estimates on vital events, such as births, deaths, immigration and emigration, are often obtained from administrative sources (Trant & Whitridge, 2000). Such events may refer to people or businesses. An example of a vital event in agriculture is the birth of a new farm operation. Customs documents providing information on imports and exports can serve as the basis for statistics on agricultural production (Trant & Whitridge, 2000). The USDA/NASS routinely publishes information on the imports and exports of agricultural products (Harris & Clark, 2013).

Direct publications of administrative data are often based on a register or register system, defined by Carfagna and Carfagna (2010) and, similarly by the UN (2011), as a systematic collection of uniquely identifiable unit-level data with an updating mechanism. A register that is populated from multiple administrative sources may

have a better coverage and completeness than a single source. Among the most frequently used types of registers are population registers and business registers. Business registers, as discussed by Wallgren and Wallgren (2010), have the potential to provide a basis for agricultural statistics because a farm operation is a type of business enterprise. Wallgren and Wallgren (2010) discuss the use of the IACS for direct tabulation. For some crops receiving government subsidies, IACS is considered highly reliable and is therefore directly tabulated to obtain aggregated area statistics, thus providing an example of the direct use of administrative data for the statistical product. Combining the IACS database with the business register leads to further improvements (Wallgren & Wallgren, 2010).

The availability of administrative databases for the purpose of direct tabulation of agricultural statistics depends on the administrative processes of particular countries. In the last 12 years, many developing countries have created their own agricultural subsidy programs to compete with prices in the United States and European Union (Clay, 2013). Subsidy programs in Brazil, Russia, India, Indonesia and China have grown the fastest (Clay, 2013). Farm insurance programs furnish a different potential source of administrative data in developing countries (Roberts, 2005). Currently, insurance programs are concentrated primarily in developed countries. However, crop insurance in developing countries is expanding due to increasing commercialism of agriculture, new insurance products based on weather indexes and by international trade policy developments (Roberts, 2005). Although the coverage of such programs in developing countries may not currently be sufficient for direct tabulation, if they continue to grow, these programs may be leveraged in the future register-based agricultural statistics.

Box 6.2.1.1. NASS publishes administrative information on hog slaughter (Harris & Clark, 2013) obtained from inspections by federal and state officials. The data from NASS hog and pig inventories should align with published slaughter data.

Box 6.2.1.2. Statistics Canada uses tax records to estimate farm expenses with a view to reducing the burden on respondents. The use of administrative data instead of survey data could improve data quality because farmers are thought to overstate expenses and understate sales in surveys (Trant & Whitridge, 2000).

6.2.2 Crop forecasting

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

Box 6.2.2.1. Zambia. There are two key sources of agricultural statistics in Zambia: the Central Statistics Office (CSO) and the Ministry of Agriculture and Cooperatives (MACO). In addition, a number of actors collect data and provide information to the CSO or MACO for use in various types of reports and analyses. However, the existence of many actors other than the CSO that collect information and use it to make preliminary forecasts before the official CSO results become available contributes to management difficulties. MACO's field staff continues to conduct a monthly monitoring of local livestock and crop development trends using non-survey techniques.

The 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's survey-based numbers from the crop forecasting survey and, later, a post-harvest survey are available (Kelly & Donovan, 2008).

Box 6.2.2.2. Mali. Legislation describing the mission of the *Direction Nationale de la Statistique et de l'Informatique* (DNSI) states that the DNSI is responsible for developing methods for data collection and analysis of national surveys, setting statistical norms, collecting and analysing administrative statistics, coordinating the

National Statistical System and approving requests for official surveys, statistical and computer science training, promoting cooperation with national and international statistical services, publishing statistical information, developing computer applications, developing and supervising the implementation of a national plan for statistical and computer services, and harmonizing standards for computer technologies and software. The DNSI publishes three annual reports: a provisional forecast of expected harvests, due in October; a report of the final harvest results for rainy season production and a provisional estimate of cereal balances for the coming months, due in March; and a report of the entire year (all rainy season and dry season production), due in June. The first two annual reports are used extensively by Mali's Famine Early Warning Unit (*Système d'Alerte Précoce*, or SAP) and the Food Security Office (*Commissariate à la Sécurité Alimentaire*) in assessing the food security situation, developing food balance sheets, and designing food aid interventions if necessary.

The SAP collects and analyses information on crop forecasts, satellite imagery, price trends and potential threats due to climate or pests, to provide early warnings of impending food crises and make recommendations for actions to ameliorate the situation. The SAP conducts on-the-ground assessments of potential food security hotspots, with a focus on high-risk areas in northern Mali. (Kelly & Donovan, 2008).

6.2.3 Planning

The resources required to collect administrative data are predictable and usually met by the administrative process itself. In brief, administrative data help the government machinery established for planning purposes and are available annually (Pangapanga et al., 2013).

Box 6.2.3.1. Through the national statistical system, Malawi's National Statistical Office is now promoting the use of administrative data for statistical purposes among its partners within the national statistical system (Malawi, 2012b).

Box 6.2.3.2. In Zambia, a food balance sheet used in crop forecasting represents the most widely circulated use of agricultural statistics in the country. It is used extensively by policymakers when considering subsidies, import/export regulations and key food security issues (Kelly & Donovan, 2008).

6.2.4 Small area estimates

Due to sample size constraints, the results from agricultural annual sample surveys are unlikely to be available at lower levels (such as the district level). Ancillary information from the regular agricultural returns can be used to scale down the highest-level estimates available in sample surveys to district-level estimates, using small-area estimation techniques. These techniques are currently used in several countries.

While the administrative data are used as auxiliary information in the model-based small area estimation carried out in Section 6.2.5., the administrative data are directly used to calculate estimates for small areas.

Box 6.2.4.1. Ethiopia. The agricultural annual surveys conducted by the Central Statistical Agency had been providing crop-wise area estimates at regional and zone levels; however, district-level estimates were not available. Meanwhile, the MARD 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 MARD as an auxiliary variable (National Bureau of Statistics, Tanzania, 2012).

6.2.5 Integration of administrative data into agricultural surveys

Administrative records can be used to cover data gaps, because the agricultural surveys in developing countries are conducted irregularly due to budget constraints (Pangapanga et al., 2013).

Box 6.2.5.1. As discussed previously, administrative data play an increasing role in agricultural statistics in Malawi. Since the last agricultural survey was conducted in fiscal year 2006/07, the Ministry of Agriculture and the Government generally has been developing programmes based on administrative data (Malawi, 2012). Moreover, several ad hoc programmes require information from the Ministry of Agriculture. Administrative processes are used to obtain this information, as they can supply it in a more timely manner compared to agricultural surveys.

6.2.6 Non-probabilistic procedures to integrate the 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 the 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 & Chin 2013). Country assessment and in-country presentations at regular meetings often cite budgetary constraints as a factor behind the failure to conduct frequent agricultural surveys. To fill the resulting data gaps , 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 the use of administrative sources, eye estimates by local informants, expert opinion or assessment, windscreen surveys, rapid appraisal methods and focus group discussions. However, among the limitations of these methods are their lack of representativeness, the inherently high levels of subjectivity and the inability to calculate a sampling error or confidence interval (Keita & Chin 2013; Galmés, 2013).

6.2.7 Use of the final statistical product by non-statisticians

Table 6.2 shows that information from ADSAS is important for policy formulation, implementation and monitoring in most countries where survey responses were received. The information is also used in supporting investment decisions, food security planning and monitoring, providing information to users for a number of various uses, and for measuring the progress towards international agreements and goals.

Table 6.2. Administrative uses of ADSAS: the uses of final statistics.

Non-statistical uses	BURUNDI	EGYPT	GHANA	LESOTHO	LIBERIA	LIBYA	MAURITANIA	MAURITIUS	SOUTH SUDAN	SOUTH AFRICA	SUDAN	UGANDA	ZAMBIA	Total/13
Policy formulation implementation and monitoring	1	1	1	1	1	0	1	1	1	1	1	1	1	12
Supporting investment decisions	1	1	1	1	0	1	1	1	0	1	1	1	1	11
Food security planning and monitoring	1	1	1	1	1	0	1	1	0	1	1	0	1	10
Providing information to users	1	1	1	1	1	1	1	1	0	0	1	0	1	10
Measuring progress of international agreements and goals	1	1	1	1	1	0	1	0	0	0	1	1	1	9
Attainment of efficient markets	0	1	1	0	0	0	0	1	0	0	1	0	0	4

Source: Survey of ADSAS 2014. Note: 1=Yes, 0=No.

In terms of users, Table 5.7 in GSARS (2015c) shows that ADSAS serve a broad spectrum of clients, notably governments, researchers, farmers, donors and traders. The ADSAS' objectives and aims may influence some parts of the performance, such as the information provided and the frequency of such provision. For example, although information on area, production and yield is very useful in government policy formulation and food security monitoring and planning, the frequency at which policymakers need it is lower than that required, for example by traders.

These survey results demonstrate that administrative records provide a major source of information to facilitate decision making for the agricultural sector. With regular reporting, policy-makers and implementers at both national and local government levels will be equipped with the data required to make meaningful decisions. Agricultural statistics are essential for service delivery and monitoring of development within the sector. Indeed, most of the routine administrative agricultural statistics are collected for monitoring the developing plans for the agricultural sector.

Data users must be capable of understanding the data and the nature of the data collection processes, so that appropriate inferences and conclusions may be drawn. The provision of thorough and understandable metadata is essential to protect against misinterpretations and incorrect uses. Statistics Canada's Integrated Meta-database is a good example of a formal system for the dissemination and storage of complete information on the nature of the data characteristics and collection processes (Dion, 2007).

To be useful, data must also be accessible, thus rendering critical issues of data access, storage and dissemination. Table 5.8 in GSARS (2015c) summarizes the frequency of use and accessibility of ADSAS in the countries participating in the Survey of ADSAS. Administrative data are most commonly accessed on an annual basis and through open-access forums, namely the Internet. Although data access on an annual basis may appear to be infrequent compared to daily or weekly access rates, it must be recalled that annual data collection by means of censuses is generally impractical, and developing countries may not have the resources to conduct annual surveys of agriculture. Administrative data, therefore, may be the primary source of annual information on agricultural activity in developing countries.

Box 6.2.7.1. India.

(a) Crop and land use statistics. Planners and policy-makers use administrative data to ensure efficient agricultural development and to make decisions on procurement, storage, public distribution, export, import and many other related issues. With the increasing decentralization of planning and administration, these statistics are needed with the greatest possible degree of disaggregation as possible, down to the level of villages.

(b) Crop forecasts: The Timely Reporting Scheme (TRS) has the principal objective of reducing the time lag in making available the area statistics of major crops, in addition to providing the sampling frame for selecting the crop-growing fields in which crop-cutting experiments are to be conducted. Under the TRS, for the preparation of advance estimates of the area under major crops, the *patwari* is required to complete the *girdawari* on a priority basis in a 20-per-cent random sample of villages and to submit the village crop statements to higher authorities within a stipulated date. The advance estimates are used in the framing of crop forecasts. This provides the Government with advance estimates of production, which are crucial for various decisions relating to pricing, distribution, export and import.

(c) Forestry statistics. Reliable forestry statistics are required for planning, policy-making, analysis and decision making on forestry investment and development programmes.

(d) Agricultural inputs statistics. For a comprehensive appraisal of the agricultural economy, information on inputs is as important as the data on production. The Directorate of Plant Protection, Quarantine and Storage (PPQ&S) within the Ministry of Agriculture advises and assists the Union Government on all matters relating to plant protection, including international obligations; in addition, assists the state governments in their plant

protection activities.

(e) Use of market information. Farmers can make use of market information for the following purposes:

i) Negotiate better prices
ii) Decide the location of sale
iii) Check prices
iv) Decide whether to store
v) Decide whether to grow "out of season"
vi) Decide whether to grow different crops
vii) Decide whether to add value through processing
viii) Work with other farmers to bulk up commodities
ix) Decide timing of commodity sales

Box 6.2.7.2. Uganda. An important user of market information was FOODNET, which developed methods, information and interventions that led towards greater market efficiency and value-added processing in the agricultural sector. FOODNET mainly focused on market analysis studies, market information, agro-enterprise development and related business development support.

6.3 Methodological tools to address quality issues

As administrative data are often collected to meet the specific aims of government programmes, rather than to produce estimates of the characteristics of a given target population, they are not necessarily directly applicable to the objectives of statistical offices. Iwig et al. (2013) provide guidelines for engaging in coordination with administrative offices, to best understand how quantities obtained through administrative processes relate to characteristics of interest to statistical agencies.

This section focuses on methodological issues rather than on those pertaining to structure, conduct and performance. Methodological issues for quality control, multiple data sources, data collection, and data storage and dissemination/diffusion are covered, considering the harmonization and consistency among various data sets to ensure international comparability.

6.3.1 Quality control

Quality control procedures for evaluating measurement errors and coverage problems are necessary to protect against bias. Measurement errors are associated to the data collection process and the coverage problem may be significant when the target population of the administrative data is different from the population of the survey data.

Measurement errors in administrative data arise from multiple sources. Conceptual differences often exist between the quantities collected through administrative processes and the 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 the varying motives in administrative processes. For example, farmers may underreport acres in subsidy programmes to guard against the consequences of inadvertent over-reporting (Carfagna & Carfagna, 2010). Measurement errors in identifying variables can occur if establishments change; however, the identifier, such as the street address, does not. Changes in the nature of administrative processes can also lead to changes in collected data over time that makes it difficult to conduct a consistent longitudinal analysis.

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 conducted 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, and find that differences in coverage properties are related to different forms of coverage error.

Box 6.3.1.1. The IACS database is maintained for the purpose of managing farm assistance programs for the EU. In maintaining the IACS, which is 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 & Carfagna 2010). ESSnet-ISAD (2008b) present a case study that illustrates the use of decision trees to harmonize the 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.

Box 6.3.1.2. 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.

6.3.2 Record linkage and reconciliation for multiple data sources

a) Record linkage

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 should be applied for many-to-one or one-to-many matches (Wallgren & Wallgren 2010).

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 can be used to unify disparate data sources. Several software tools for probabilistic record linkage have been developed to this end (ESSnet-ISAD 2008c). The detailed list is given by GSARS (2015a).

Box 6.3.2.1. Statistics Sweden makes extensive use of record linkage to create rich register systems based on several administrative databases. One mechanism that facilitates the 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. Wallgren and Wallgren (2011) recommend that statistical institutes should invest in register maintenance processes that preserve the integrity and confidentiality of the identification number.

Box 6.3.2.2. In a multiyear experiment conducted in Italy, administrative records are linked with interview surveys to improve the quality of data on income. Due to confidentiality rules, the Italian tax agency performs record linkage to provide the Italian Statistical Institute (ISTAT) with the desired tax information for each sampled individual.

b) Reconciliation

If the differences between data sources are systematic or can be understood, then inconsistencies across data sources do not necessarily prohibit the use of multiple sources of information. Many developed countries have used a probability record linkage method to combine data with inconsistent identifying variables.

However, data reconciliation can be achieved in various ways. Since the probabilistic record linkage method may be due to the equipment of the software and the lack of experts, many developing countries have used other reconciliation methods to combine the survey or census data with other administrative data. Table 5.2. in GSARS (2015c) illustrates the methods of data reconciliation used in developing countries that have participated in the 2014 ADSAS survey. One method is to fit the administrative data into the overall GDP growth; the other is to compare the routine agricultural administrative data collected with the census data.

Detailed examples of the use and reconciliation of multiple data sources in India, Uganda and Mozambique are provided in Appendix A to this final Technical Report.

6.3.3 Data collection

a) Data collection methods

Employing best practices in the collection of administrative data can reduce the number of quality issues to be addressed at the estimation stage. Various methods and equipment must be used to measure land and crop areas, to estimate production and yield, etc. Furthermore, structured questionnaires are used with enumerators or through a wiki approach (users send SMS or update a website designed for the purpose). These issues are discussed in this subsection.

Data collection methods vary depending on the parameter of interest. The methods for collecting production and area estimates sometimes differ from those used in obtaining price information. Table 4.1 in Technical Report 3 (GSARS, 2015c, p. 58) shows that self-administered questionnaires and routine reporting are the main methods used to collect administrative data reported in selected ADSAS.

Box 6.3.3.1. In the United Republic of Tanzania, data forms completed by the VAEWs are compiled by the WAEWs, who check for completeness and seek any clarifications required. To ensure that completed forms are received, the submission of ward-level data to districts and the receipt of blank VAEO or WAEO forms is usually linked to the distribution of salaries to VAEOs or WAEOs from the District. At the district level, data entry is done in Excel, using the village or ward data collection form. Training is provided to equip the district officers with skills in data entry and analysis and provision of feedback to the data collectors. Technical backstopping is provided for the district officers consisting of a team of two competent M&E Technical Working Group members and one regional officer.

Provision of backstopping is important to ensure that district officials are fully conversant with ARDS operation. All LGA officers gather in regional towns and report their progress and challenges.

Box 6.3.3.2. In Uganda, estimates made by the Department of Agriculture were guesstimates extracted from the annual and monthly reports compiled by district agricultural officers, who will have obtained the data from the agricultural extension staff at the lower administrative units – usually sub-counties. The information collected was on the area, yield, production, prices, marketing, etc. 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, groundnut, sim-sim, castor and vegetables.

Two methods were used in the estimation of annual crop areas (Uganda Bureau of Statistics, 2007):

(a) the "*Buganda Method*": randomly selected villages would be supposedly completely enumerated (in the "*Mitala*

Survey") in respect of the areas of all important crops. The area under each crop obtained by pacing and/or eye-estimation or pure guesswork would be aggregated and then divided by a "refined" number of taxpayers belonging to the sample villages. The average thus derived would then be multiplied by the total number of taxpayers in the entire district, to obtain the estimated area under the given crop in the district.

(b) the "*Outside Buganda Method*": returns of plot counts would be carried out by chiefs and compiled for the two major seasons in the year. These plot counts would be aggregated and multiplied by a general plot mean size, supposedly derived from pacing by the extension staff, to obtain the district crop area totals. A variation of this, which was much more widely applied than the "*Buganda Method*" was to require extension staff, rather than chiefs, obtain both plot counts and mean plot sizes. Production was then estimated as the product of area and yield. The yield estimates were always obtained subjectively by the relevant district agriculture officer with the help of his or her staff, led by an assistant agricultural officer. Unfortunately, these systems broke down in the late 1970s. Currently, most data is drawn up at the district headquarters by the extension staff without any consultation with the chiefs. The methodology used did not yield reliable estimates (Uganda Bureau of Statistics, 2007). The complete demise of the system previously in place completely broke down, thus exemplifying the issue of sustainability.

b) Forms, questionnaires and instruction manuals

Ensuring that the questionnaire used is as short as possible can greatly improve the quality of the collected data. A questionnaire that is too lengthy may inflict an excessive burden on the respondent, thus leading to nonresponse and poor data quality. This widely known principle was confirmed by the pilot project conducted in the United Republic of Tanzania, as it became apparent that even the questionnaire used in such project could have been revised or shortened. All instructions should be included in the instructions manual, and separate enumerators' and supervisors' manuals provided. This would, *inter alia*, make the questionnaire less bulky. In any case, when using a tablet, the instructions can be programmed within the instrument, and easily opened separately

c) Technologies used in data collection

The use of certain technologies in administrative data collection leads to increased data quality, especially in terms of timeliness. For example, the use of GPS equipment in distance and area measurements increases accuracy while reducing the time required for data collection, although it might also increase the costs of information collection (thus affecting the ADSAS's sustainability). There is likely to be a trade-off, with the reduction of time required for data collection leading to less time for the enumerators. On the other hand, GPS equipment are generally more expensive than other types of measurement equipment. Table 4.3, in GSARS (2015c), shows that in Africa, the use of modern technologies such as GPS equipment, PDAs, computer-assisted telephonic interviews and scanning questionnaires remains low. The most common technologies mentioned are personal interviews (used by 10 out of 13 countries) and manual data entry into computers (by 8 out of 13 countries). It is recommended that African countries increase their use of new technologies.

Box 6.3.3.3. Crop Area Statistics: In India, since 1990, the centrally sponsored Crop Acreage and Production Estimation (CAPE) scheme has sought to use RS technology to estimate crop areas and land use. The objective of CAPE is, among others, to provide state-level crop area estimates, meeting a 90/90 accuracy goal using the RS data that cover mainly the states' crop-growing parts.

d) Data collection and transmission using CAPI

Modern but cheap and readily available technology, such as tablets, can improve timeliness in these respects. Data entry would be performed in the field by data collectors (chiefs or village extension workers).

If electronic devices are used for data capture, the data is captured directly in the database, which is accessible at district, regional or provincial and national levels. Rights of access should be made such that supervisors are capable of checking and verifying data, and then forwarding them to the district officials who in turn verify and clear for access by the higher levels. As noted during the pilot project conducted in Côte d'Ivoire, hard copies of the data should be kept in case devices fail.

The software should be designed such that reports can be automatically generated once the data is compiled at district level. These reports can be used as a basis for monitoring the system, as well as for forwarding data to the higher levels and district planning purposes.

During the pilot project, it was demonstrated that a **Crop Card** is capable of improving the estimation of production, particularly for mixed, perennial and continuously harvested crops. During the pilot, it was also proposed to program the Crop Card used in the estimation of crop production on mobile telephones, so that farmers can report regularly and electronically. However, this may only apply to mobile telephones using the Android operating system.

There was a view that the Crop Card should be limited to continuously harvested crops.

e) Data transmission and submission

As long as a district or region enjoys a reasonable level of connectivity, data can be immediately transmitted to the central server as soon as it is entered into the CAPI application on the tablet. Although data can be transmitted at any time as long as an Internet connection is available, data and reports should be considered to be officially submitted only after they have been approved by the supervisors.

When an Internet connection is available, data is transmitted to the central server as soon as it is entered. If there is no such connection, data is temporarily stored on the tablet and sent to the server as soon as the connection is restored.

6.3.4 Data storage and dissemination or diffusion

This covers the ICTs used in data storage and dissemination or diffusion, i.e. hand-held equipment, telephones, traditional ICTs (e.g. radio, television and fax); modern ICTs (e.g. e-mail, Internet and SMS); PDAs; etc.

As discussed in GSARS (2015a), metadata are vital for informing both producers and users on data quality. It is recommended that metadata be present at all stages. Incoming data should be accompanied by sufficient metadata to enable their full comprehension, and to ensure that values are correctly allocated to the relevant variables. Metadata are at the heart of the management of the interpretability indicator. An example was previously given of the Integrated Metadata Base as (IMDB) Statistics Canada's sole source of metadata information describing surveys and programs. The quality of the information on the IMDB must be monitored regularly to ensure completeness and accuracy. It was stressed that it is important for statistical agencies to publish good metadata because in so doing, they demonstrate openness and transparency and thus foster trust among data users (Dion, 2007).

When introducing new data storage or dissemination technologies, agencies are advised to consider the associated merits and risks. A new technology may bring benefits in certain dimensions, while creating costs in

other areas. The issues to consider with respect to reliability, accessibility, timeliness, and sustainability are the following:

Reliability: Can the technology improve the accuracy of the information diffused? The use of some technologies, such as SMS or e-mail) reduces diffusion errors.

Accessibility: Some illiterate users may not be able to read SMS and e-mails.

Timeliness: Some technologies may enable the faster transmission of administrative data and information (e.g. prices may be more quickly sent via SMS compared to the updating of a website).

Sustainability: This criterion concerns the costs involved with the use of a technology to store or disseminate information. Some technologies may be fast but expensive, e.g. iPads, which are associated with high fixed costs. Some technologies may also not be feasible if electricity and security are issues.

Box 6.3.4.1. In the United Republic of Tanzania, a number of routine reports were produced at all administrative levels (village, wards, districts and regions) even before the improved ARDS was introduced. The VAEO Report (which is produced on monthly, quarterly and annual bases) is compiled at village level (on paper). These reports are based on VAEO or WAEO formats and are canvassed by VAEOs (or WAEOs, if the village does not have a VAEO). The Integrated Data Collection Report (compiled on a quarterly and annual basis) is produced at district level (and stored in LGMD-2 format). The Integrated Regional Report (compiled quarterly and annually) is produced at regional level (and stored in LGMD-2 format), while the Integrated National Report (also compiled quarterly and annual) is produced at national level (and also stored in LGMD-2). Thus, integrated national-level quarterly and annual reports are expected to be provided by the Improved ARDS.

Box 6.3.4.2. In Uganda, databases are owned by the agencies which generate the data. However, the data are shared with the Uganda Bureau of Statistics according to need, especially when data is required for the Annual Statistical Abstract. Data is disseminated in hard copies (publications) and CD-ROMs; summaries are posted on the website http.://www.ubos.org.

7 Legal issues

7.1 Legal and political constraints

The access to administrative data may be limited by legal and political constraints, which may be in place for good reason – for example, to protect the confidentiality of the individuals in the population (Brackstone, 1987). A detailed MOU explaining the objectives of the statistical office and the data required of the administrative office to meet those objectives is often necessary to establish a flow of data from the administrative agency to the statistical office. Developing a high-quality MoU is often an expensive and time-consuming process (Prell et al., 2009).

7.2 Basic law

7.2.1 Frameworks for legislation and policy

The UN (2011) discusses the value of laws and policies in ensuring that statistical offices can access the necessary administrative data. In many cases, legislation exists that explicitly provides for access to administrative data. For example, the Statistics Acts of Ireland and Norway establish permission for the national statistical office to access administrative data. An extract from the Irish Statistics Act of 1993 states that "for the purpose of assisting the [statistical] Office in the exercise of its functions under this Act, the Director General may by delivery of a notice request any public authority to – (a) allow officers of statistics at all reasonable

times to have access to inspect, and take copies of or extracts from any records in its charge, and (b) provide the Office, if any such officer so requires, with copies or extracts from any such record, and the public authority shall, subject to subsection (2) of this section, comply with any such request free of charge" (UN, 2011). Because opportunities to pass legislation are scarce and effectively impacting legal frameworks requires substantial effort, statistical offices are advised to propose legislation with a long-term strategy in mind (UN, 2011). In an example from Statistics Canada, Brackstone (1987) notes that a government tax reform is an opportune time for the statistical office to engage with policy-makers and strive to shape data collected through legal structures in a way that satisfies the needs of the statistical agency.

When amending legal frameworks is impractical, policies may be formulated to facilitate access or changes to administrative data. Policies are easier to change than laws and tend to evolve more dynamically over time (UN, 2011). One example of a policy framework involving administrative data is Principle 5 of the UN's *Fundamental Principles for Official Statistics*, which emphasizes the cost-effectiveness of administrative data and promotes the use of such sources in the interest of making efficient use of the information available (UN, 2011).

7.3 Agreements between institutions

An MOU is often helpful in facilitating collaborations between institutions. MOUs clearly state the roles and obligations of each participating institution involved in the agreement. The pilot project for this research, conducted in collaboration with the Bureau of Statistics of the United Republic of Tanzania, required an MOU. This MOU is provided as an example in Appendix 1 to this final Technical Report.

Given the legal and policy frameworks required to permit the use of administrative data, written agreements are often necessary to detail and facilitate the transfer of knowledge and data (UN, 2011). These written agreements are often in the form of an MOU that specifies the objectives of the statistical office in using the administrative data and the information required to meet those objectives (Prell et al., 2009). Brackstone (1987) draws attention to the success of Statistics Canada in forming "bilateral committees", with participation from both the statistical office and administrative agencies, in developing the necessary organizational and technical infrastructure.

Prell et al. (2009) analyse seven case studies involving written agreements that establish or expand relationships with administrative agencies. They identify four distinguishing characteristics of a successful MOU: (1) vision and support by agency leadership; (2) narrow but flexible goals; (3) infrastructure; and (4) mutual interest. The case studies indicate that these "elements for success" enable agencies to work through many of the challenges that arise in the process of establishing a written MOU.

Iwig et al. (2013) provides an outline to guide interactions between the statistical office and the administrative office in forming a data-sharing relationship. Their "Data Quality Assessment Tool for Administrative Data" is shaped around the quality dimensions of relevance, accessibility, coherence, interpretability, accuracy and institutional environment. For each quality dimension, Iwig et al. (2013) recommend several questions that the statistical office should ask the administrative agency. For example, in the interest of ensuring the coherence of concepts, classifications and data collection methods over time and across geographic domains, the following query should be made: "Please describe any classification systems used for categorizing or classifying the data".

Organizational agreements also have the potential to overcome the restrictions associated with preserving confidentiality. If both statistical data and administrative data are deemed to be confidential, an MOU as discussed in Prell et al. (2009) may provide a legal mechanism for data transfer. In some cases, expansion of the definition of the national statistical system may enable a more liberal circulation of administrative data among government offices, including statistical agencies. As discussed by Wallgren and Wallgren (2007), Statistics Sweden receives regular deliveries of administrative data from the agencies responsible for government programs and regulations.

Administrative data should be collected with a notification as to the uses to which the information will be put, so that the circumstances are clear as to when the administrative records should be considered private information and treated confidentially.

Data sharing among agencies refers to those methods whereby agencies can obtain access to one another's data on individuals, sometimes immediately but nearly always, in any case, on a timely basis. Data sharing offers a number of benefits. If different agencies collect similar data on the same person, the collection process is duplicative for both the agencies and the person. Data sharing therefore can increase efficiencies by reducing the paperwork burden for the government and the individual, as basic information on clients only needs to be obtained once. It may also be possible to improve the response rate.

Although data sharing has many benefits, it raises issues regarding privacy and confidentiality. Who should have access to these data? How can confidentiality and privacy rights be protected while achieving the benefits of linking program data?

A detailed set of frameworks is necessary to facilitate access to administrative data for statistical purposes. These frameworks typically have several dimensions: legal, policy, organizational and technical. It is necessary to reach agreement in all of these areas before the benefits of the use of administrative data can be attained.

International standards are also of assistance in terms of providing guidance, and should therefore be referred to wherever possible in discussions with administrative departments. It is essential to establish a legal framework that permits the use of administrative data for statistical purposes. The other frameworks mentioned are also useful for assuring a smooth flow of data and minimizing any problems or misunderstandings that may arise between the data supplier and the statistical organization. For this reason, it is helpful if they are reflected in written documents that are agreed upon by all parties.

Legal frameworks are normally constructed at the national level and are specific to national sources and circumstances. In some cases, however, there may also be relevant legislation at the international level. In these cases, there may be two or more alternative legal possibilities to the use of administrative data.

Most national statistical organizations have legal frameworks defining their roles and responsibilities, typically in the form of a statistics act. In some countries, these legal texts have been revised in recent years and now include specific provisions enabling access to administrative data. Countries that have not made such amendments should proceed to do so, as they are a necessary step.

National historical, political and institutional factors strongly influence these legal frameworks. As a result, national differences may arise and result in legal frameworks that are not particularly harmonized or even consistent between countries. It is therefore important to improve the comparability of statistics that have been derived wholly or partly from administrative sources. The relevant regional economic commissions could then establish an international legal framework governing access to administrative data.

In addition to enabling access to data from administrative sources, legal frameworks should also establish limits to such access and to the possible uses of administrative data. Often, there are restrictions according to which data can only be used for specific statistical purposes, or the confidentiality of individual records should be maintained. There may also be specific restrictions on the use of data.

7.3.1 Policy frameworks

Many countries have general policies on data sharing within government bodies, which will influence the right of access to administrative data for statistical purposes. Policy frameworks also encompass voluntary codes of practice, the most important of which, for statistical purposes, is the UN's *Fundamental Principles of Official Statistics*.

Codes of practice should also be published at the national level to reassure the public that data will only be used for specific purposes. To have any real value, it is important that these codes of practice be made available to the general public.

Once the legal and policy frameworks are in place to permit the use of administrative data, it is necessary to consider the organizational arrangements to facilitate data flows. Typically, this takes the form of a written agreement or MOU.

Certain key features should be noted in all such agreements:

- Legal basis. Reference should be made to the legislation permitting the access to the administrative source for statistical purposes, and to any legislation that imposes restrictions on such access.
- Names of the persons transferring and receiving data, The names and contact details of the key people involved in the supply of data in both administrative and statistical organizations should be recorded.
- Detailed description of data covered. This will include information identifying the data set and the variables contained within it.
- Frequency of data supply. This will specify when and how frequently the administrative organization will supply the required data.
- Quality standards. These set the parameters for the quality of the data supplied. Examples include the requirements for a maximum proportion of missing or erroneous variables, to ensure that the data received are fit for purpose. The priorities assigned to different variables, and hence the effort made towards quality assurance, will often differ between administrative and statistical organizations; therefore, agreeing on common standards is of paramount importance.
- Confidentiality rules. It is important to expressly state the uses that may be made of the data, the rules and procedures in place to prevent disclosure, and the circumstances in which the data can be passed on to clients of the statistical organization.

7.3.2 Technical standards

This dimension involves the following aspects:

- Provision of metadata. It is important that data flows be accompanied by the relevant metadata, which may include dates, descriptions for any codes used, information on the units used, etc.
- Provisions on payment for data supply. Data transfers between government departments or agencies are generally free of charge, although in some cases, the statistical organization may be required to contribute towards the costs of extracting and transferring the data. Data from private-sector organizations may be charged for at market rates, although it may be possible to negotiate discounts, particularly if there are several users of a private-sector data source within government. In some cases, it may be possible to offer statistical analyses or expertise as a form of payment for the data received.
- Period of agreement. Agreements will normally be for a fixed period, but should include provisions for renewal or extension if necessary.
- Contingencies for changes in circumstances. It is important for the statistical organization to receive advance warning of changes affecting the administrative source. The agreement should specify that any proposed changes are to be communicated to the statistical organization as soon as possible, to allow the impact of the changes on statistical outputs to be minimized.
- Procedure for resolving disputes. The agreement should specify the method to be adopted in resolving any disputes which may arise between the statistical and administrative organizations; these may envisage the involvement of senior managers or possibly even relevant ministers.

7.3.3 Technical frameworks

The technical frameworks are the mechanisms by which data are transferred, as well as any relevant data or metadata standards. The data transfer mechanism adopted must take into account the technical possibilities available to both the sending and the receiving organization.

8 Cost of setting up the ADSAS

The costs associated with the ARDS and TAD are considered in Sections 8.1 and 8.2, respectively. Because the ARDS is an explicit data collection program, we discuss the costs associated with starting-up or initiating and maintaining an ARDS. For TAD, the data are collected naturally through the administrative process. Rather than discuss the costs associated with collecting the administrative data, we focus on the costs incurred by the statistical office in using the administrative data in Section 8.2. In many developing countries, the lack of necessary financial resources presents a major obstacle. In Section 8.3, we discuss possible sources of funding for ADSAS.

8.1 Reporting system

Considering the great costs associated with setting up an ARDS, it is best to build it in stages. In addition, governments at both national and lower levels should include the operational costs of the ARDS system in their budget.

The findings during the pilot project and other assumptions have been used to create a budget framework, which is attached to Technical Report 6 as Appendix 2. The total estimated budget for the first year (to set up the system) slightly exceeds 7 billion Ugandan shillings (approximately US$2.1 million). However, other considerations also arise, as discussed below.

a) The following items pertain purely to data collection:

1. Advocacy and communication efforts, to ensure sensitization and raising awareness.
2. Data compilation equipment.
3. Training for administrative data compilation at various levels.
4. Field transportation (data collection and monitoring).
5. Data analysis and report writing, and
6. Data dissemination and policy engagement.

b) For some items, only a percentage is attributable to administrative data collection

1. Procurement of vehicles (vehicles and motorcycles)
2. Procurements of data compilation equipment
3. Computers, printers and scanners
4. Salaries and wages

For these items, the percentage of the time spent on data collection by the extension staff could be used. Such percentage was assumed to be 30 per cent; thus, 30 per cent of the above costs can be attributed to data collection.

c) Assumptions on costs incurred in subsequent years

i) Recurrent costs

1. Advocacy and communication: lower than the first year
2. Training for administrative data compilation at various levels – refresher courses and some new staff
3. Salaries and wages
4. Field transportation (data collection and monitoring)
5. Data analysis and report writing
6. Data dissemination and policy engagement

ii) Capital costs – mostly replacements. Therefore, only a percentage of the initial cost is considered

1. Procurement of vehicles (vehicles and motorcycles)
2. Procurement of (data compilation equipment)
3. Computers with printers and scanners

Finally, it should be noted that for ease of comparison between different studies or surveys, it is best to estimate the cost per unit, e.g. by holding or agricultural household. For example, Uganda's 2014 Population and Housing Census cost approximately 31,650 Ugandan shillings per household, i.e. under US$10.

d) Budgeting for agricultural administrative data analysis and management

In many developing countries, the main difficulties encountered in the production of agricultural statistics consist in the following:

i) Inadequate or absent material and resources required to collect statistics;
ii) Unskilled human resources in agricultural data processing and analysis – often, the data is not fully analysed and not related to policy;
iii) Lack of capacity-building programs for the staff involved in data generation;
iv) Lack of information required (especially in disaggregated form) and poor organization of data collection, analysis and archiving;
v) Lack of national classifications to guide the generation of statistics; and
vi) Failure to fully exploit the data (see Technical Reports 1 and 3),

However, it has been noted that in some cases, the increased use of administrative data transfers costs from data collection to data analysis. The costs associated with data analysis and management are a major component and their estimation is very important. One consideration is that, since the data are being collected for administrative purposes regardless, what is required are the additional costs to convert the data into more useable information. This includes direct and indirect costs; improved equipment, additional training, the increased supervision, data analysis, dissemination, etc.

With costs diffused throughout the data collection process, the goal is to estimate the costs at each level of the process.

Taking the experiences of the pilot project, the budget items should specifically include:

i) Printing and photocopying (paper, toner, etc.) costs and other stationery;

ii) Data entry (equipment and personnel) – entering data by means of CAPI could reduce data entry costs;
iii) Data transmission and submission;
iv) Data management and access to the Internet, e.g. establishing and maintaining a webpage;
v) Data analysis and utilization;
vi) Budget for technical support by district/regional IT staff, in case of computer malfunctioning;
vii) Report production and workshops for discussing and disseminating results.

8.2 Other sources of administrative data

The use of TAD in statistical programs is often recommended due to the cost savings that it enables. The data are already collected through the administrative process; therefore, the statistical office does not incur the cost of data collection, as it does with surveys and censuses. Actually, the cost structures associated with administrative data are more complex. The statistical office must consider costs associated with data management, quality control, analysis and advocacy. In some cases, the statistical office may have to pay the administrative office a fee to access the data. Because the relationships between administrative agencies vary by country, it is difficult to obtain a specific cost estimate that applies generally. However, costs can be classified into two categories: (1) activities for which costs can be shared between administrative data management and other activities; and (2) costs associated specifically with the management and use of administrative data.

i) Shared costs

1. Salaries

2. Computers and technical equipment, including data compilation equipment

3. Measures to ensure the security of confidential data

ii) Costs specifically pertaining to administrative data

1. Data analysis and report writing

2. Fees for data sharing arrangements

3. Advocacy

8.3 Sources of funding and sustainability strategies

Many ADSAS receive limited or insufficient funding, which entails the late or irregular collection of information, an inability to hire well-trained staff, and, in many cases, unsustainability. This in turn may lead to poor quality, in terms of timeliness. Table 3 (taken from Technical Report 3) displays the sources of funding for ADSAS for a non-probability sample of 13 African countries. All of these ADSAS receive funding from their respective governments and four receive additional funding from donors. One ADSAS is funded by charity organizations, and another in, South Africa, by the private sector. By default, where data is collected by an MDA, the funding is from the government (although a contribution could of course also be obtained from donors) and is more certain than in cases where a private organization, survey or census is the source of funding. However, the FOODNET market information programme activities in Uganda are an example of an ADSAS funded by a consortium of donors, which included USAID (through ACDI-VOCA); the Government of Uganda through MAAIF and NAADS; the CTA; and RELMA. An option that did not emerge from the studies and literature review is the possibility of raising funds through subscription fees and information sales by ADSAS. This may be due to the fact that ADSAS are mostly government departments that must provide information as a "public good".

Table 8.1. Sources of funding of ADSAS: results taken from interviews with the leadership of statistical offices for a non-probability sample of 13 African countries.

	BURUNDI	EGYPT	GHANA	LESOTHO	LIBERIA	LIBYA	MAURITANIA	MAURITIUS	SOUTH SUDAN	SOUTH AFRICA	SUDAN	UGANDA	ZAMBIA	TOTAL/13
Funding source														
Government	1	1	1	1	1	1	1	1	1	1	1	1	1	13
Charity organizations	0	0	0	0	0	0	0	0	0	0	1	0	0	1
Donors	1	0	0	0	0	0	0	0	1	0	1	0	1	4
Private sector	0	0	0	0	0	0	0	0	0	1	0	0	0	1
Farmer or trader organization	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Source: Survey of ADSAS 2014. Notes:1=Yes, 0=No, Blank

9 Start-up operational strategy for using and improving existing administrative data sources

The following sections contain suggestions on building an effective ARDS. Due to resource constraints, it is best to start with a small system and scale it up gradually. The short-term strategy will focus on leveraging existing data sources, monitoring data quality, and improving data collection and management. The recommendations will be general, with country-specific examples used for illustration.

9.1 Enumerate a complete list of potential data sources

An important first step in building an ADSAS is to identify all possible sources of administrative data. Examples are given of the administrative data sources used by the Global Strategy for Improving Agricultural Statistics to produce statistics related to core data for the – land cover, crop production, and livestock (World Bank, 2010). While not every country will have access to all of these sources, this summary provides a comprehensive overview of the vast range of administrative data available. This section reviews the sources covered in more detail in Chapter 3 above.

i) Summary of main sources

The specific sources covered in Chapter 3 may be aggregated into the following main groups: data from government assistance and regulatory programs; data from monitoring and record-keeping; taxation data; private-sector data; and information from reporting systems. Examples of government assistance and regulatory programs that can provide administrative data include crop insurance and subsidy programs, information from veterinary inspections, and livestock vaccination campaigns. Such data sources provide information on crop area and livestock numbers, for example. Examples of monitoring and record-keeping include land registration and cadastral records, as well as soil surveys – administrative operations that fully document the soil types of a specified region. Statistical offices in Europe, Australia, Canada and the United States of America make extensive use of administrative data in producing agricultural statistics. Canada has even outlined plans to directly use taxation data to measure farm expenses in the 2016 Census of Agriculture. Examples of private-sector organizations that may provide administrative data relevant to agriculture include the following: licensing or regulatory bureaux, grain associations, commodity associations, cooperatives, factories, slaughterhouses and distributors of agricultural inputs. In routine data collection (ARDS), administrative data are collected through subjective assessments by local experts. ARDS are used in developed countries (e.g. in the NASS monthly crop report) and are extensive in developing countries, as highlighted by our pilot project in the United Republic of Tanzania.

ii) Linking sources to core data items

The sources discussed above cover many of the Global Strategy's core data items. Subsidies provide information on crop areas, tax data provide information on farm expenses, slaughter and vaccination records provide information for forecasting and estimating hog inventories, and data from distributors provide information on dairy production (see Table 9.1 below).

Table 9.1. Sources of administrative data for selected core data items of the Global Strategy.

Core data items	Administrative data type	Example
Crops	Farm subsidies	IACS contains crop areas for the crops enrolled for subsidies

• planted area, harvested area, yield, yield, storage, labour, prices • maize, barley, wheat, sorghum, rice, cotton	Grower associations	The Ontario Grain Association provides information on prices
Livestock • cattle, sheep, pigs, goats, poultry • inventory, births, production, prices	Animal health regulations	SANITEL in Belgium is populated with data from animal health regulations and supplements surveys
	Cattle tracing systems	Cattle tracing systems populate the European Union Bovine Register
Forestry • area of woodlands and forests, quantities removed, prices	Forest cover area	United Kingdom: the Forestry Commission records complement statistical surveys in estimations of forest area and woodland prices in the country
Land cover • classification of coverage of a country • categories: cropland, wetland, grassland	Land registration and cadastral records	India: the country's land registration system supports estimates of areas in various land-cover categories

9.2 Assess the possible uses of administrative data

Administrative data may be used directly, as the final statistical product, or indirectly, in forming the statistical product. Developed and developing countries make direct and indirect use of administrative data. Developing countries are often required to use administrative data as the final statistical product due to lack of funding for surveys or censuses. Examples of direct and indirect uses of administrative data are provided in Chapter 6. The following subsections note issues to consider when determining appropriate uses of administrative data sources.

i) Direct use of administrative data

In this case, tabulations of administrative data are published directly as the final product. Below, conditions are provided under which direct tabulations of administrative data may be appropriate or necessary. These issues are revisited in Section 9.3. If the conditions below are not met, the statistical office should consider conducting a survey or census.

- The administrative data source measures the quantity of interest to the statistical agency.
- The administrative data source completely covers the target population of interest.
- Reporting errors in the administrative data source are negligible.
- Conducting a survey or census of the item of interest is impractical or too expensive. This may occur for very rare items that are difficult to reach through a survey but are available in administrative data sources.

ii) Indirect use of administrative data

As discussed in Chapter 6, statistical offices often use administrative data to support survey or census programs. Administrative data play a role in all stages of the survey process: frame construction, data collection and

estimation (i.e. calibration, small area estimation and imputation). Using administrative data as part of a survey or census program is particularly advantageous when administrative and survey data have complementary strengths and weaknesses. In a common situation, administrative data provide essentially complete information (i.e. a census) on a quantity that is correlated with but different from the item of interest to the statistical office. The survey, in contrast, obtains accurate information for a subsample. The list below notes a few ways in which administrative data may help to improve the estimators obtained from surveys or censuses.

- Data collection – administrative data, such as maps, images, and addresses, may be integrated in data collection tools to help data collectors locate the sampled agricultural unit.
- Frame construction and maintenance – administrative data may enrich a frame with auxiliary information and more timely information on births and deaths, and the eligibility of agricultural operations.
- Improving the efficiency of estimators – administrative data known for the whole population may be used in calibration procedures or small area estimation, to improve the efficiency of estimators from surveys.
- Reducing non-sampling errors – administrative data known for the whole population can be used to form imputation models or weighting adjustments that aim to reduce bias due to selection errors, such as under-coverage or non-response.

9.3 Quality assessment

Statistical agencies and academic journals have developed several lists of criteria for evaluating the quality of a statistical product. The UN (2011) defines "quality" as the degree to which an administrative source meets user demands. Administrative data have various roles in the production of official statistics; therefore, an evaluation of the quality of these sources must be specified in relation to the intended use.

i) Quality measures

The following list of quality measures for administrative data is based on the UN (2011) definitions, and is also discussed in Section 6 of Technical Report 1.

- *Relevance* – the extent to which an administrative source aligns with current and future uses. Because administrative data sources have many potential uses, relevance must be considered with respect to specific applications.

- *Accuracy* – the closeness of the match between the measurement or estimate and the value of interest. This applies to estimates of aggregate quantities such as means or totals, and to microdata.

- *Timeliness and punctuality* – the time elapsing between the occurrence of the event of interest and the date on which the administrative data are available for use in official statistics. It is important to factor in editing, modelling and processing times. An administrative source should not be considered ready for use until these processes are complete. Punctuality reflects the ability of an administrative agency to meet deadlines, being the time between the date on which data are released and their announced release date.

- *Accessibility* – the ease with which users can obtain administrative data. Confidentiality conditions and policy and legal constraints may hinder access to administrative data.

- *Clarity and interpretability* – the ease with which users can understand administrative data. Information on the administrative data source may improve the clarity and interpretability of the source, and metadata such as maps, summary tables or documentation on collection procedures can improve users' ability to interpret the data.

- *Coherence and consistency* – the compatibility between administrative data sources or between administrative and survey data. Several data sources may be capable of measuring a target concept; however, different processes can lead to differences among means.

- *Comparability* – the validity of comparisons between estimates based on related administrative sources; this aspect is related to coherence and consistency and may be affected by methodological differences in administrative or survey sources and differences in the definitions applied therein. Assessing comparability involves determining how the observed differences result from differences between the true underlying concepts of interest, or from the differences in observed values arising from the measurement processes.

ii) Quality with respect to use

Administrative data have various uses in official statistics, and a single administrative source may have numerous applications; the dimensions of quality discussed above may have different interpretations depending on the intended use. Laitila, Wallgren and Wallgren (2011) define three forms of quality in relation to three classes of use: (i) in "**output quality**", the administrative source directly supplies a statistical product of interest; an example is direct tabulation ; (ii) in "**input quality**" the administrative data source is used only after refinement or in conjunction with other sources; an example is the use of administrative data as a partial substitute for survey data; (iii) in **"production process quality"**, the administrative data source is used to improve the quality of estimates based on sample surveys or other administrative sources; calibration and small area-estimation are processes in which production process quality measures should be considered. Each of these three qualities are discussed below.

If an administrative source is to be used for the direct tabulation of official statistics, the output quality is pertinent, and the quality indicators listed above (which generally apply to surveys and censuses) will apply to the administrative source.

Administrative data may only be suitable for generating final statistical products after they have been refined or integrates with other data. Laitila, Wallgren and Wallgren (2011) consider two cases: (i) the administrative source is used independently for the statistical process after cleaning and editing; and (ii) microdata from the administrative source are combined with microdata from other registers and surveys, and the data from the completed register system are used to form the administrative database. The input data quality of the administrative source should be evaluated in both cases. The criteria for evaluating input data quality are less restrictive than those for output data quality, because fewer demands are placed on the administrative source. Coordination with the statistical agency may be necessary to improve accessibility. Linking variables may have to be converted to a form compatible with the statistical production system, and units may have to be redefined or aggregated, a process which is known as "profiling" in the literature on business registers. If the coverage of an administrative source is thought to be incomplete, the source may be added to a register system of administrative files from other agencies, censuses or GIS. A critical component of input data quality is that of

"linking variables": a unique personal identification number, for example, can improve the feasibility of linking processes (Wallgren & Wallgren, 2010).

The last dimension is production process quality, comprising, for example, the ways in which an administrative source may be used to improve estimates based on surveys. For calibration and area-level small-area models, the requirement for microdata is reduced, which in turn reduces the need for compatible identifying variables. Because relationships between means based on administrative sources and means obtained from surveys are estimated, the administrative data do not need to provide an unbiased measure of the true underlying quantity of interest. This reduces the amount of coherence and accuracy required in the data source relative to direct tabulation (Laitila, Wallgren & Wallgren, 2011). Timeliness, punctuality and accessibility are important in process quality, because obtaining auxiliary information in a timely manner is essential to give time for model building and evaluation. Analysts are often obliged to choose between correlated variables from administrative sources, even for model-assisted estimators. Thus, an evaluation of trade-offs between bias and variance also require attention.

9.4 Managing and improving the quality of administrative data

Statistical agencies in developed countries have mechanisms to ensure data quality that can be applied in developing countries. Engaging with administrative offices and with the public can help to align the definitions required for statistical purposes with those used by administrative agencies, mitigate the effects of administrative changes on the usability of the data for statistical purposes, and address concerns about privacy and security (Brackstone, 1987; UN, 2011). Approaches that combine administrative databases with information from surveys can reduce the problems associated with measurement error, enable reconciliation between definitions in different sources and improve coverage (Wallgren & Wallgren, 2010). Audits and sampling of administrative data are used to check for errors and evaluate coverage (Carfagna & Carfagna, 2010). Adopting best practices for quality control and assurance can help to manage errors in administrative data. These recommendations are explored in more detail below.

i) Combining multiple data sources

Statistics and numerical summaries resulting from different data collection systems often differ substantially, even if they are intended to measure related concepts. Use of a single administrative data source may lead to biased estimators as a result of undercoverage or reporting errors that may derive from incentives in the administrative process. Combining data from disparate sources can help to compensate for errors in a single source, and using multiple data sources can improve the coverage of a population or increase the available variables. Combining information from several sectors can improve the coherence of estimates in different subject areas: for example, a business register can be combined with a farm register. Synthesizing several data sources with different error properties may lead to the formulation of estimators having better properties than an estimator based on a single source. An administrative file may be used to update a list frame to be later used in combination with an area frame in a multiple-frame survey. External data may be used to reduce non-sampling errors through error checking or imputation of missing data, and administrative data may be used as controls in calibration or as auxiliary information in model-based estimation.

The implications of the differences between administrative and survey data depend on the intended use of the administrative data source. If it is to be used to tabulate statistics directly, and the resulting statistics contradict parallel statistics based on a survey, the disagreement between the two sources may generate confusion.

However, if the administrative data source is to be used as an input in a model-based or model-assisted estimator, the lack of consistency among the sources is not necessarily problematic. Indeed, if part of the model-fitting process involves an estimation of the bias associated with the administrative data, the process of constructing the final statistics accounts for the differences between the administrative and survey data. In combining multiple data sources, it is often necessary to designate one source as the standard unbiased measurement. This source may still have errors, such as sampling errors, if it is a survey; however, designation as the unbiased source means that the average of this source aligns with the target concept or quantity of interest. When determining data reconciliation procedures, it is important to understand the nature of each data source. Understanding input data sources guides the choice of data source as the unbiased standard source. Incorporating sufficient metadata is one way of ensuring that the properties of the administrative data source are transparent to data users.

ii) Audits and embedded samples

Audits and embedded sample surveys are recommended to correct for errors in administrative data. Carfagna and Carfagna (2010) explain that in the IACS database, "… sample surveys must be performed or the statistical system will produce biased results." To identify erroneous cases, a survey or audit that flags unusual records may be conducted. A slightly different design is required if the objective is to estimate the extent of error in an administrative source.

iii) Systems and best practices for quality assurance and control

Standard best practices for quality control and assurance also apply to administrative data; the standardization of variables, for example, is valuable for data integration and interpretation. This section reviews mechanisms for quality management related to data management, collection and dissemination. Establishing conceptual and technical infrastructures for maintaining administrative data can help to ensure data quality. Technical Report 1 proposes use of the ADSAS for maintaining and verifying the quality of administrative data related to the core agricultural data items in the Global Strategy.

An alternative is to borrow from the Health Matrix Network, a tool designed to assist developing countries in planning health information systems (African Development Bank et al., 2007). The reader is referred to Technical Report 1 for further detail. Biemer et al. (2014) describe the System for Product Improvement, Review and Evaluation, a framework for quality evaluation that "provides a comprehensive framework for systematically evaluating all dimensions of quality with the primary focus on accuracy." The concept, which covers a range of quality dimensions, was developed to evaluate the collection of statistical products generated by national statistical offices.

Detailed monitoring of the collection of administrative data can help to improve their quality. Galmes (2013) provides guidelines for improving the quality of collected data. Although developed for the context of expert assessment, these guidelines apply to data collection and include the following advice: (i) use of a standard format for collecting information; (ii) preparation of manuals containing clear definitions of activities; (iii) provision of periodical training to data collectors; and (iv) supervision of all operations. Implementing such quality controls in data collection can improve the transparency of data collection procedures and help statistical analysts and consumers of statistical products to use the data appropriately.

To apply quality controls, it is important to possess sufficient information on the administrative data. Information on collection processes may indicate the reliability and accuracy of the resulting administrative data. Accurate definitions of administrative data sources are essential for understanding the data items and utilizing them appropriately. When transferring administrative data between administrative and statistical agencies, administrative data offices should provide sufficient information to statistical agencies to ensure appropriate use and interpretation of the data (Dion, Chartrand & Murray, 2010).

In their approach to evaluating administrative data, Daas et al. (2012a) define quality as "statistical usability", which is different from usability as a final product. They found that Statistics Netherlands had a well-developed approach to evaluating metadata, but that there were no corresponding guidelines to evaluate "the quality of the data in administrative sources" (Dion, Chartrand & Murray, 2010). The approach to routine evaluation of administrative data for statistical purposes developed indicators for five quality dimensions (see Example 7.4.2). The tools recommended for evaluating the quality dimensions include graph techniques, which are useful for visualizing relationships between data sources and identifying errors. Daas and Fonville (2007) emphasize the value of metadata in evaluating and improving an administrative register. They recommend that before an administrative data source is used, the metadata be evaluated in a quality checklist to determine whether a source can be used and the extent to which it can be used. They advocate a sequential approach: (i) review the metadata; (ii) evaluate coverage and overall reliability; and (iii) evaluate the data in terms of timeliness and continuity.

Statistical offices apply established best practices to verify the quality of incoming administrative data. The 2013 special issue of the Journal of Official Statistics was devoted to "Systems of Architectures for High-Quality Statistics Production". Concepts ranging from quality control and assurance frameworks developed for survey data apply to administrative data, and frameworks for evaluating and maintaining the quality of data used and produced by statistical offices contain concepts valuable within the ADSAS, as outlined in Technical Report 1.

iv) Methodological and technical tools for data integration

The operation of data integration presents computational and methodological challenges. Statistical software can reduce the computational burden of merging large data sets. Statistical methods such as profiling and probabilistic record linkage can handle any lack of standardization in the definitions of units or identifying variables. Depending on the objectives and the available data, the integration of multiple sources can be done at the level of an individual unit or for an aggregated group of units (i.e. a region or the country). Below, examples are reviewed of unit-level and aggregate-level integration methods. Additional details are available in Technical Reports 2 and 7.

a) Unit-level integration

● *Profiling*

One challenge associated with integrating multiple sources of information is posed by the fact that different data sources can have different definitions of units. The UN (2011) explains that "… converting administrative units to statistical units can be quite difficult conceptually and often involves some form of modelling." The term "profiling" is used in business surveys to describe this process, although the concept applies in other contexts as well (UN, 2011).

Profiling may be manual or automated (UN, 2011). Standard rules based on attributes or on the nature of links between units can help to overcome differences between administrative and statistical units. Statistical households, for example, can be derived on the basis of relationships between the individuals living in a building; indeed, this approach is a component of the register-based population census method used in Nordic countries. Even with clerical profiling, the disaggregation of units may require subjective determinations, and a single correct solution may not exist. In automated processes, which are cheaper and faster than clerical profiling, standard rules regarding the nature of links are applied uniformly.

An alternative to rule-based profiling involves the specification of statistical models. Relationships between administrative and statistical units may be established for a subset of a population, for example through a survey; and parameters of models describing relationships then estimated and applied to the full population. An example is the case where the administrative unit is a "job" and the statistical unit is a "person" (UN, 2011). In an estimate based on a survey, each person has 1.15 jobs on average; this estimate can be used as a global adjustment factor to determine estimates of employment from the number of jobs. The variability in the survey-based estimate of the ratio would have to be incorporated in subsequent employment analyses.

● *Deterministic record linkage*

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 & Wallgren, 2010).

● *Probabilistic record linkage*

Many applications of administrative data used to produce official statistics involve multiple sources of information, such as administrative and survey files. In many cases, it is desirable to link records from at least two files at the level of the individual unit in a population. Consider, for example, the use of administrative data to check for errors in survey data. A comparison of the marginal distributions of the administrative file to the corresponding marginal distributions from the survey may be informative; however, a comparison of the alternative data sources at the unit level opens greater possibilities: a unit-level linking operation, for example, enables the evaluation of records with relatively large differences in the values recorded in the two different sources. Merging the files at the record level presents many challenges. The identifying variables may differ across data sets, and even if a unique identifier exists, the identification variable may be missing or incorrectly recorded for some units. Duplicate records may exist in one or more files, and large data sets may demand a substantial computational effort. Probabilistic record linkage is a statistical procedure to determine the probability that two sets of identifying variables represent the same unit in the population.

● *Mass imputation*

Different registers and surveys often contain different "response variables." When data are linked at the unit level, the existence of several versions of related variables provides an opportunity for quality improvement and expansion. Creating a single complete data set in which each record appears once is called "mass imputation". Data are then imputed for all records in the resulting register system. Mass imputation involves complex modelling techniques, and computational challenges arise as a result of the enormous volume of data (Guigo, 2008).

As an example, Statistics Canada's survey of employment payroll and hours provides monthly estimates of status and trends in 10,000 establishments. Statistics Canada also has access to the complete file of payroll deductions remittance forms from the customs and revenue agency. These administrative sources provide the number of employees and gross monthly payroll variables. Using this data, regression models can predict missing survey variables using the administrative variable as covariates. In many instances, mass imputation of the survey response variables for all units in the administrative file is possible (Grondin & Lavallée, undated).

b) Aggregate-level data integration

● *Small area estimation*

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 district 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 currently used in several countries. In Africa, Ethiopia is one of the countries that employs small area estimation. The agricultural annual surveys conducted by the Ethiopian Central Statistical Agency had been providing crop-wise area estimates at regional and zone levels, but not at the district level. Concurrently, MARD generated area estimates through an approach 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 MARD as an auxiliary variable (National Bureau of Statistics, Tanzania, 2012). For a more mathematical treatment of small area estimation, see Technical Report 2.

● *Measurement error modelling*

Agricultural variables are almost always subject to measurement error. If the measurement error is large, ignoring it may lead to biased and inconsistent estimates that may, in turn, result in spurious conclusions. Measurement error models are statistical approaches that combine multiple sources of information in a multi-level model to obtain a single unified statistic, and an associated measure of uncertainty. In Technical Report 6, a measurement error model is applied to estimate the area planted to maize in Namibia. Three sources of information on the maize planted area are used. One of the estimates is obtained from the Annual Agricultural Survey (AAS) conducted by the NSA. Two are obtained from the MAWF. These different sources of information are then combined to obtain a more precise estimate of the true planted area. Further details on the model and estimation procedure are available in Section 5.3.1 of Technical Report 6.

c) Linking the ARDS with statistical surveys and censuses

It is useful to link the development of the ARDS with the cycle of censuses (i.e. the 2010 and 2020 PHCs and agricultural censuses). For example, in 2010, the PHCs of several countries included an agricultural module. This is useful in developing an agriculture sampling frame and presents a great deal of potential for small area statistics. The administrative data in agriculture could be based on the NSO or CSO enumeration areas used in the PHCs or agricultural censuses.

Further, the AAS or annual livestock census could be combined with the ARDS estimates on the number of plots or livestock households. The average plot sizes and crop yields, plus the average number of livestock owned by livestock households, etc.; could be taken from the annual surveys and censuses especially in the upcoming round of the World Census of Agriculture (WCA).

To make crop production estimates, in the ARDS, one would only count the plot numbers for countries that conduct an AAS. In countries that do not conduct an AAS, a subsample can be used to estimate the average plot sizes and average yield using GPS equipment and Crop Cards, respectively, as done during the pilot.

Box 9.4.1. Uganda's experience (1970s): the **"**Outside Buganda Method" (discussed in Box 6.3.3.2. above)

- Chiefs carried out plot counts and compiled returns thereof for the two major seasons in the year.
- Extension staff:
  - aggregated the plot counts;
  - derived a general plot mean size by pacing; and
  - multiplied the aggregated plot counts by the general plot mean size to obtain district-level crop area totals.
- The district agricultural officers **subjectively obtained yield estimates** in their respective districts, with the help of assistant agricultural officers.
- **Production** was then estimated by the district agricultural officer as the **product of area and yield.**
- **The yield estimates could be computed objectively, by using a Crop Card**

It is also recommended that in future WCAs, developing countries should endeavour to also carry out ARDS, to enable comparisons and imputations (calibrations). The ARDS data collection could also be carried on a sample basis, in the same areas where the census data are collected. Currently, few countries carry out ARDS during the agricultural census year.

Finally, it will also be necessary to conduct annual censuses, or at least surveys, for large and institutional farms to collect data on production either by the NSO or farmers' associations/groups and ministries of agriculture. These can easily be covered under the censuses or surveys of business establishments covered by NSOs. A systematic approach must be used to develop and maintain a unified directory of large and commercial farms that can be used for all surveys. Sufficient resources should be allocated for updating this directory over time. Eventually, it would be ideal for this directory to be maintained by the NSO as a component of the business register.

Each district agricultural office should compile a list of all the large-scale or commercial and institutional farms in the district, as defined in terms of a minimum farm size or number of livestock.

9.5 Institutional arrangements

Institutional arrangements include the structures, staffing and linkages with other sources of agricultural data. A critical component to establishing an effective ADSAS is ensuring that data collection, analysis and dissemination are coordinated and shared between different agencies. This typically requires the conclusion of a

formal agreement or MOU specifying the obligations of the participating institutions. Formal coordination and technical committees must be formed, with clear terms of reference as outlined below.

i) The Coordination Committee

The institutional arrangement begins at the top, with a Coordination Committee composed of key stakeholder institutions. This is a national body formed to oversee the formulation of and compliance with the policy guiding the system, as well as of the system overall. This committee should also be responsible for ensuring the standardization of concepts, definitions and processes to enhance comparability with other agricultural data sources.

The institutions involved should include:

1) Government ministries and departments active in the field of agriculture
2) The ministry responsible for local governments
3) Parastatal bodies handling agricultural produce
4) Private-sector institutions handling agricultural produce, inputs, etc.

ii) Technical Committee

The Technical Committee should be responsible for managing the system. It should include experts in different aspects of the implementation and management of administrative agricultural routing data systems. The Technical Committee should be responsible for:

- Establishment of the system, starting small and expanding as resources become available and as expertise is improved;
- Overall management of the system down to the grassroots level;
- Regular monitoring and periodic evaluation of the system;
- Promotion of the utilization of the data generated from the system;
- System maintenance, revision and improvement;
- Data capture, cleaning and verification;
- Data storage and dissemination;
- Maintenance and improvement of data collection tools, materials, methods and other relevant documents; and
- Capacity building of technical staff from the national level to the lowest level, where extension officers collect data at the grassroots.

iii) District Technical Committee

The Technical Committee should have branches at district level to oversee, monitor and provide support and guidance to the technical staff in their respective assignments. Such staff should be constituted by technical staff from the department of agriculture, veterinary, apiary and other related departments, and district planners and statisticians. The Technical Committee should provide technical guidance to field staff, which include field supervisors and the field extension workers who will be collecting data.

Meetings should be held on a quarterly basis to receive and review reports from their areas of operation, so that feedback can be provided that continually improves system processes. Teams should visit the field staff to ensure that all tasks are being adequately accomplished.

Depending on the size of the country and its administrative hierarchy, a regional or provincial technical committee could be established above the district level, that would be responsible for the districts in the region or province, as appropriate.

The roles of the various institutions involved are outlined in Table 9.2.

Table 9.2. Institutions involved and their respective roles.

Institutions	Roles
NSO/CSO	• Overall coordination and guidance; • Standardization of tools, concepts and definitions; • Supervision or provision of support to supervision of all activities – a specific official or department should be designated • Auditing the system • Cross-check on the supervision of the MOAFS, MLFD and other agencies
Ministries of Agriculture and Forests (MOAFS) and Ministry of Livestock and Fisheries Development (MLFD) for crops and livestock, respectively – specific officials will be designated	• Recruitment/designation and training of field staff • Supervise data collection, flow, analysis and dissemination. • Maintain data banks
District and regional offices	Data collection
Donors	• Provide technical assistance or funding, especially at the beginning • Assist in the designing and initial supervision of the field activities

In Canada, components of the Statistics Act that mandate certain forms of data-sharing between federal agencies has facilitated the transfer of administrative data between administrative institutions and Statistics Canada. The formation of bilateral committees involving the statistical office and the administrative agency have helped to ensure the success of such data-sharing arrangements in Canada, as noted by Brackstone (1987).

When forming data-sharing arrangements, participating institutions are advised to consider the guidelines provided by Prell et al. (2009) and Iwig et al. (2013). Based on an analysis of seven case studies, Prell et al. (2009) identify four characteristics that underlie successful collaborations: (i) vision and support by agency leaders; (ii) precise but flexible goals; (iii) infrastructure; and (iv) mutual interest. These "elements for success" enable agencies to work through the challenges that may arise when an MOU is being established. Iwig et al. (2013) develops a data quality assessment tool to help guide administrative and statistical offices through the process of transferring the data. These principles and guidelines are discussed in further detail in Chapter 7 above.

9.6 Costs and advocacy

i) Costs of maintaining an ARDS using TAD

In many developing countries, a major limiting factor is a lack of financial resources. Ostensibly, administrative data present an attractive solution because the statistical office does not incur data collection costs, as it does instead with surveys and censuses. However, the statistical office must still consider the costs associated with integrating the administrative data source into the statistical system. Examples of costs that the statistical office will likely have to absorb include costs of advocacy, ensuring data security and quality control.

One characteristic that distinguishes routine reporting systems from TAD is that routine reporting systems are implemented with the explicit purpose of collecting agricultural information from specified units. The costs of establishing and maintaining an ARDS will therefore be considered below. Certain costs can be shared across multiple activities (i.e. vehicles, computers, data compilation equipment, salaries and wages), while others will be devoted entirely to the ARDS (advocacy, training, transportation, data analysis and dissemination). Many costs can be shared across levels of government – national, regional or sub-regional. When establishing the initial budget, it is important to consider that capital costs (such as vehicles and equipment) will amortize over time, while recurrent costs will require funding on an annual basis. When evaluating the cost of an ARDS relative to a survey or census, it is informative to consider the cost per unit (i.e. agricultural holding). The ARDS may be considered more efficient from a cost perspective if the data collection cost per unit is lower than that of a survey or census.

For TAD, the data are gathered naturally through the administrative process. However, the statistical office is often responsible for costs with associated data management, analysis and quality control. These are often nontrivial because the objectives of the administrative institution differ from those of the statistical office. As for the ARDS, some costs can be shared across activities, while other costs will be devoted specifically to the use of administrative data.

Government agencies often incur costs associated with managing administrative data for the purpose of agricultural statistics. Developing countries are advised to consider other possible funding sources, which may include charity organizations, donors, the private sector, or farmers' or traders' organizations.

ii) Advocacy: proactive involvement in administrative processes and engagement with the public

Engagement with the public and participation in decision making as to the nature and accessibility of administrative data can help to improve usability. Legislation permitting statistical offices to influence and access administrative data can promote standardization and prevent time lapses between collection and release (UN, 2011; Brackstone, 1987). Bilateral committees have fostered productive working relationships between administrative and statistical offices. Publicizing the measures used to ensure the confidentiality of administrative data and the public benefits derived from utilizing administrative data may help to minimize scepticism.

9.7 Country examples

This section illustrates how some of the issues discussed above are manifested in specific cases. The countries considered are India, Namibia, the United Republic of Tanzania, and Côte d'Ivoire. India is a good example because extensive literature exists describing its unique land records system. Namibia, the United Republic of Tanzania, and Côte d'Ivoire were selected because the pilot projects for this research took place in these countries.

i) India

India has one of the most elaborate systems for the collection and management of data, specifically agricultural statistics and administrative data. It is a decentralized system also within the state governments. The National Sample Survey Organisation is responsible for the planning and operation of the ICS scheme and employs full-time staff for field supervision. It shares the fieldwork with the designated state agencies, which carry out the field supervision in approximately half of the sample villages. The institutions involved are the State Agricultural Statistics Authorities (SASAs), which operate at the state level, and DESMOA, which is responsible for compiling data at the national level.

Administrative data sources include the NHB and the state Directorates of Horticulture and Agriculture. Two main sources generate statistics on the production of horticultural crops: (i) DESMOA, which operates the centrally sponsored "Crop Estimation Survey on Fruits and Vegetables" scheme to estimate the area and production of fruit, vegetable, and spice crops in eleven states; and (ii) the NHB, which compiles and publishes estimates of area, production and prices of the all important fruit and vegetable crops based on reports furnished by the state Directorates of Horticulture and Agriculture. These estimates are based on the informed assessments of local-level officials dealing with horticulture and the reports of market arrivals in major wholesale fruit and vegetable markets.

Estimates of cotton production are collected and published by the Cotton Advisory Board (CAB) and those for oilseeds by the Central Organization for Oil Industry and Trade (COOIT). DESMOA's estimates on yield are based on the *girdawari* (*complete enumeration of all fields*) for area and crop-cutting experiments under the GCES, whereas COOIT's estimates of cotton production mainly depend on the feedback received from important markets on arrivals, crop trends and additional information provided by members of the industry.

Administrative data is also used to determine land usage. The institutions involved are SASAs, DESMOA and the National Remote Sensing Agency (NRSA). Statistics of land use are compiled from the village land records maintained by the *patwari (village revenue agency)*. Land use statistics are also collected through nationwide land use or cover mapping by the NRSA according to a 22-fold classification. The categories are much more detailed and provide useful information for land development programmes. There are many more administrative data sources that provide information on agricultural prices, agricultural inputs (e.g. fertilizers and farm equipment), livestock and fishers. More details on specific agencies can be found in Technical Report 6.

Finally, another advantage for India's agricultural statistics is its infrastructure. The most common weakness in developing countries is the lack of coordination between the NSO and the various administrative agricultural data collection and management systems. India, however, is an exception: most parts of the country have detailed cadastral survey maps, frequently updated land records, and a permanent village reporting agency. Therefore, in addition to collecting a vast quantity of administrative data, India is capable of effectively using the data to generate reliable and timely statistics.

ii) Uganda

In Uganda, MAAIF has a standard template to facilitate regular reporting on agricultural data generated from administrative records. MAAIF collaborates with the local governments at all levels, primarily the district and subcounty levels. The local administrative-level staff collects agricultural data, including crop and livestock-related data, on a regular basis – such as monthly or quarterly. They report to the District Production

Coordinator at district level, where the data is summarized further and utilized when necessary. The District Production Coordinator is supposed to share the agricultural data with MAAIF every quarter.

A number of Commodity Boards operates outside the usual MAAIF structure. These receive policy guidance from MAAIF and collect data (most of it administrative) mainly for their own operations. Some of these boards are: NARO; the CDO; the UCDA; the DDA; NAGRIC & DB; and NAADS.

In terms of managing administrative data, Uganda too suffers from the problems that commonly affect developing countries, mainly that of frequent changes in the administrative structure. Indeed, Uganda has experienced many fluctuations in the number and boundaries of districts. For example, in the 1980s there was a shift from purely administrative chiefs to semi-political local council leaders who were not used to collecting data. Similarly, the decentralization policy meant that the extension staff were no longer answerable to the central government. The effect of these changes was lower quality data, in addition to difficulties encountered integrating administrative data with data from the NSO.

iii) Namibia

Namibia has several sources of statistical and administrative data. Currently, administrative data in Namibia is exploited primarily for direct use – that is, they are published directly as the final statistical product. Because statistical data on agriculture also exists, Namibia has a real opportunity to make additional use of administrative data to improve the statistical product. This initiative's pilot project in Namibia centred on exploring the feasibility of synthesizing administrative data (routine reporting systems and TAD) with statistical data. Below, the work undertaken for the project is summarized, highlighting best practices, areas for improvement and future opportunities.

First, the use of administrative data to improve the efficiency of subnational estimators through small area estimation models is considered. The administrative data were obtained from the Crop Assessment Checklist, a type of routine data collection based on a nonprobability sample of agricultural holds and on the judgement of local experts. The Crop Assessment Checklist provides regional "estimates" of crop area and production each year. Because the Crop Assessment Checklist estimates are based on a subjective assessment, they do not have an associated sampling error. The Crop Assessment Checklist estimates at the regional level are used as auxiliary variables in Fay-Herriot-type models (Fay & Herriot, 1979) for estimating crop area at the regional level. The survey data are from the communal portion of the 2013/2014 Namibia Census of Agriculture. Because this is based on a probability sample of agricultural holdings in the communal sector, it is possible to estimate the associated sampling variance. The modelling resulted in reductions in the estimated CV for some areas. Because regional-level sample sizes based on the Namibia Census of Agriculture are relatively large, the efficiency gains were modest, even though the administrative and survey estimates across regions are highly correlated. For details, see Technical Report 7.

Second, we consider the use of measurement error models to combine three or more data sources. Administrative and survey data often measure parameters that are different but related. Differences in target parameters may occur because of differences in reference time points or subtle differences in the way questions are phrased. Variable patterns in measurement bias and measurement errors may occur because of differences in data collection procedures. Factor models (Fuller, 1987, Chapter 1.5) provide a way to synthesize correlated measures of the same quantity to obtain a single estimate of a parameter of interest. One of the estimators must be assumed to be unbiased for the parameter of interest. This unbiased estimator often comes from a probability-based sample survey conducted under controlled conditions. Biases of the estimators are represented

through fixed regression coefficients. Assumptions that certain variances are equal or uncorrelated are often required to identify the model. If an estimate of the sampling variance of a survey estimator is available, these assumptions can often be relaxed. We considered the use of measurement error models to combine multiple sources of information on crop area in Namibia. For details of the models and the assumptions, see Technical Reports 6 and 7.

The pilot project conducted in Namibia shed light on several best practices. Most prominently, the NSA has concluded an MOU with the MAWF that facilitates data sharing and collaboration. Namibia's Livestock Census and cattle tracing system are administrative data sources maintained by the MAWF that exemplify many best practices in collecting and managing administrative data. For the purposes of disease monitoring and vaccination, the data are collected and disseminated in a consistent and timely fashion. Namibia is the only African country with a complete livestock tracing system.

In light of these best practices, areas for improvement may be identified. One such area would benefit from the archiving of quantitative administrative data in a system with consistent identifying variables. Data from the crop assessment checklist are aggregated to the level of administrative regions through subjective process; however, data are initially collected at the holding level. Archiving the quantitative data at the level of the holding and the constituency would enhance data usability and improve the ability to analyse differences over time and between data sources. Data on communal crops are associated with administrative regions, while data on commercial crops are associated with crop production regions. This creates a modifiable areal unit problem that makes combining commercial and communal crop data prohibitive. Identifying commercial units and communal holdings with both an administrative region and a production region would resolve this issue.

Creation of a consistent archiving system would also enhance opportunities for using administrative data to improve estimators from surveys and censuses. For example, data from the livestock tracing system could be used to maintain frames for agricultural surveys. Commercial agricultural data could be used to adjust for nonresponse in surveys of commercial agricultural operations.

iv) United Republic of Tanzania

In the United Republic of Tanzania, the ARDS was developed to meet the data needs for monitoring and evaluation of the ASDP. The ARDS is very elaborate and the arrangements in place are such that data are aimed to be collected from every village. The structure of the ARDS is well-laid out and its tasks and operations are well-stipulated in the relevant ARDS operation guide. However, it has been difficult to conduct effective monitoring, supervision, planning and policy formulation, because the ARDS has not been functioning properly. A pilot study for the improved version of the ARDS was carried out in Kondoa and Mpwapwa districts in Dodoma Region and in Kilosa and Morogoro District Council in Morogoro Region from 2008 to 2010.

The quality of the ARDS was evaluated using a variety of dimensions, such as relevance of the data, accuracy, accessibility, consistency and timeliness of data reporting. Unfortunately, however, many shortcomings were identified. Only 18 per cent of the districts were able to transmit data to the national level – an indication of difficulties in routine data processing within most districts. When consulted, most people said that they do not have the resources to travel around the area within their jurisdiction. The relevant Operations Guide stipulates that certain data quality checks be made by the district-, regional- and national-level staff. It was concluded, however, that the quality control systems and support supervision systems in place are inadequate, which casts doubt on the quality of the collected data. A more comprehensive assessment of the ARDS is available in Technical Report 3.

v) Côte d'Ivoire

The main findings of the pilot test conducted in Côte d'Ivoire were as follows:

- There is no system for generating crop area and production estimates for the food crops. However, whenever data is collected, it is validated at the regional level between MINADER and ANADER. Meetings are regularly held between the INS, MINADER, ANADER, the ONDR (for rice) and other state institutions responsible for generating statistics on the cash crops, to discuss the reliability of the data generated.
- There is no national agricultural statistics committee. MINADER used to have a Technical Working Group (TWG) to check the validity of the agricultural data under the CountryStat programme but this ceased to function in 2014.
- The Ministry is in the process of revamping the Permanent Statistics Data Collection System and therefore the recommendations from the pilot are eagerly awaited.

10 Long-term strategy for constructing statistical registers from administrative data

The *Global Strategy to Improve Agricultural and Rural Statistics* (World Bank, FAO & UN, 2011) discusses the integration of agriculture into the national statistical system to improve agricultural statistics. Consistent with this view, a long-term strategy is proposed here to construct an integrated system of statistical registers utilizing administrative data sources.

The strategy closely follows the guidelines proposed by Wallgren and Wallgren (2017). Because Wallgren and Wallgren exclude expert judgements and subject estimates in their definition of administrative data, this strategy focuses on combining TAD, such as farm, population and business registers.

It is sought to provide a plan to create an integrated statistical farm register that can be regularly updated with multiple administrative data sources. Prerequisite conditions for a register-based system are discussed in Section 10.1, and step-by-step strategies in Section 10.2.

10.1. Modernization for the integrated register-based system

a) Microdata with identities

Administrative registers comprise *identifiers.* Identity numbers play an important role in the construction of an integrated register-based system: they are capable of assessing data quality and connect multiple data sources, using techniques such as deterministic record linkage. Therefore, the starting point for an integrated system is a register-based survey or census combined with administrative data linked using identifiers.

b) Improving the administrative system

Administrative data should have the same level of quality as survey data. Key variables such as identity numbers, registration of births and deaths, and migration are also important to link data across sources.

c) Protection of confidentiality

Identity information should be made anonymous for individuals and businesses. Ensuring confidentiality is critical if data is to be freely linked with other administrative data sets without risking the exposure of personal information.

d) Centralization, cooperation and legislation

Another key condition is the construction of a centralized statistical system. Wallgren and Wallgren (2017) recommended that the "*national statistical institute in a country should be responsible for all registers that replace the population and housing census and all registers that are used for the National Accounts*". In addition, cooperation between NSIs and other organizations is necessary for administrative systems to operate effectively.

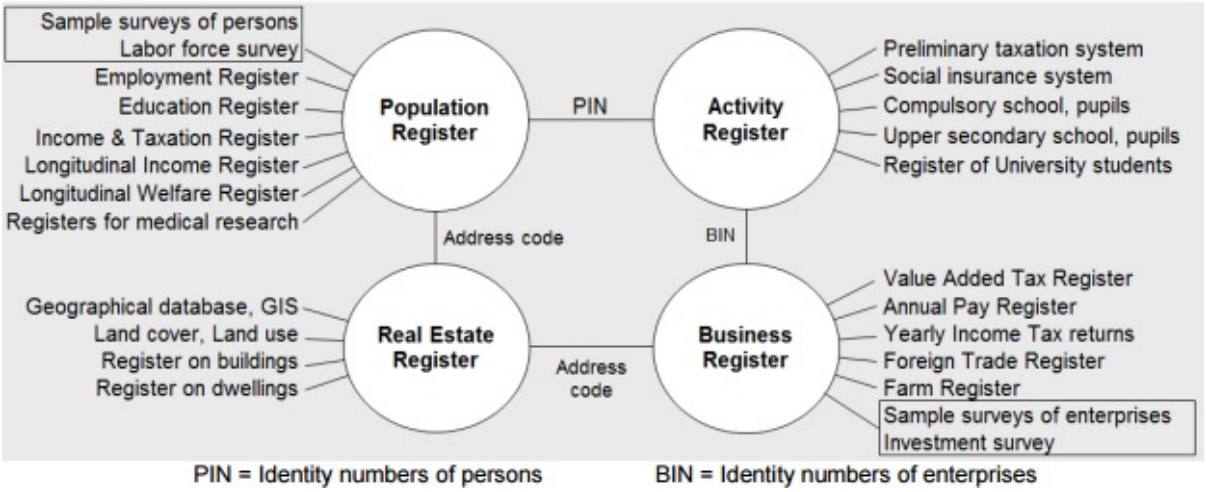
e) Quality assessment

As discussed in Chapter 5, administrative data frequently have methodological, sampling, and other data quality issues. In this respect, register data is similar, often presenting, for example, coverage problems and measurement errors. Thus, administrative data should be evaluated prior to use using a variety of quality metrics. A systematic check of input data quality is discussed in Wallgren and Wallgren (2014).

10.2. Creating the integrated register-based system for agricultural statistics

A necessary condition for modernization is the existence of an integrated statistical system, upon which a consistent register is built. Figure 10.1 illustrates a general statistical production system with base registers, other statistical registers and sample surveys. The term "integrated system" is used to emphasize that populations and variables are consistent, and estimates coherent.

Figure 10.1. A register-based production system, the register system and sample surveys.



Source: Wallgren and Wallgren (2017)

The creation of new statistical systems depend on the construction of a registration system. Details of the system procedures proposed by Wallgren and Wallgren (2017) are given below:

a) Step 1: Create a national registration system with good *identity numbers*. These personal identity numbers should be used across the various administrative systems.

b) Step 2: Develop a *statistical population register*, which may be based on the administrative population register and supplemented with other sources to improve the coverage and quality of residential addresses.

c) **Step 3:** Develop the *real estate register or cadaster*, the *business register* and the *farm register*. The employment and education register are also essential in a register-based system.

Wallgren and Wallgren's plan (2017) is considered to be long-term because the necessary conditions and building steps require significant time and cost. The small-scale nature of agricultural production in many developing countries makes the production, let alone maintenance, of farm registers very difficult at best. The most viable option could be to institute registers of institutional and large-scale farms. Unfortunately, in many developing countries, these account only for small proportions of production. Another drawback of this strategy is certain administrative data, such as expert judgements and eye estimates, are not handled in an integrated register-based system. However, if financial resources and institution support are available, a register-based system is an ideal method for collecting agricultural data.

A number of developing countries, especially in Asia, have developed and even digitized their land registers. This improves the quality of agricultural administrative data and is a good example to follow in those developing countries without land registers.

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*. Paper presented at the Fifth International Conference on Agricultural Statistics, Kampala. 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, 4/12/2017.

Administrative Data Liaison Service (ADLS). 2017. *Administrative data introduction*. Available at: http://www.adls.ac.uk/adls-resources/guidance/introduction/, 4/12/2017.

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*. AfDB Publication: Tunis.

Asian Development Bank (ADB). 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*. ADB Publication: Manila.

Barboza, W.J. & Harris, M. 2009. *Utilizing an Alternative Sampling Frame to Produce Agricultural Survey Indications*. Paper prepared for the 2009 Research Conference of the Federal Committee on Statistical Methodology. NASS/USDA Research and Development Division. Available at http://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/reports/conferences/FCSM/fcsm2009_paper.pdf, 4/12/2017.

Benedetti, R., Espa, G. & Lafratta, G. 2005. *Tree-based approach to forming strata in multipurpose business survey.* Discussion paper No. 5. Publication of the Università Degli Studi Di Trento – Dipartmento di Economia.

Biemer, P., Trewin, D., Bergdahl, H. & Japec, L. 2014. A System for Managing the Quality of Official Statistics. *Journal of Official Statistics*, 30: 381–415.

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

Bycroft, C. 2010. *A register-based census: What is the potential for New Zealand?* Statistics New Zealand Publication: Wellington.

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. & Piersimoni, F. (eds), *Agricultural Survey Methods*. John Wiley & Sons: Chichester, United Kingdom.

Chambers, R. 2009. *Regression analysis of probability-linked data.* Official Statistics Research Series, vol. 4, Statistics New Zealand.

Clay, J. 2013. "Are agricultural subsidies causing more harm than good?" *The Guardian*, 8 August 2013. Available at: http://www.theguardian.com/sustainable-business/agricultural-subsidies-reform-government-support, 4/12/2017.

Daas, P.J.H. & Fonville, T.C. 2007. *Quality control of Dutch Administrative Registers: An inventory of quality aspects.* Paper prepared for the Seminar on Registers in Statistics – methodology and quality, Helsinki, 21-23 May 2001.

Daas, P.J.H., Ossen, S.J., Tennekes, M. & Burger, J. 2012a. *Evaluation and visualisation of the quality of administrative sources used for statistics*. Paper prepared for the European Conference on Quality in Official Statistics, 29 May – 1 June 2012, Athens.

Daas, P.J.H., Puts, M.J., Buelens, B. & Van den Hurk, P.A.M. 2012b. *Big Data and Official Statistics*. Available at http://www.pietdaas.nl/beta/pubs/pubs/NTTS2013_BigData.pdf, 4/12/2017.

Deville, J.C., Sarndal, C.E., & Sautory, O. 1993. Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88: 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: R. Benedetti, M. 97
Bee, R. Espa & F. Piersimoni (eds) *Agricultural Survey Methods*. John Wiley & Sons: Chichester, United Kingdom.

Elliott, M. & McDonnell, P. 2007. *Request from the United Kingdom to Use the Bovine Registers of Great Britain and Northern Ireland in Replacement of Statistical Surveys*. Publication of the Department for Environment Food and Rural Affairs Surveys, Statistics and Food Economics Division: London.

European Communities (2003) Methodology of animal statistics, Office for Official Publications of the European Communities, Luxembourg

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.

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.

Goebel, J.J. 2009. *Statistical Methodology for the NRI-CEAP Cropland Survey*. USDA/NRCS Report. USDA Publication: Washington, D.C.

Global Strategy to improve Rural and Agricultural Statistics (GSARS). 2015a. *Reviewing the Relevant Literature and Studies on the Quality and Use of Administrative Sources for Agricultural Data*. Technical Report 1. GSARS Technical Report: Rome.

_____ 2015b. *The Role of Administrative Data in Developed Countries: Experiences and Ongoing Research.* Technical Report 2. GSARS Technical Report: Rome.

_____ 2015c. *Critical Analysis of Agricultural Administrative Sources Being Currently Used by Developing Countries.* Technical Report 3. GSARS Technical Report: Rome.

_____ 2015d. *Analysis of Agricultural Administrative Data Gaps and Ways of Improving the Quality and Use of Administrative Data Sources for Agricultural Statistics*. Technical Report 4. GSARS Technical Report: Rome.

_____ 2015e. *Strategy and Methodology for Improving the Use of Administrative Data – A Protocol for in-Country Testing. Technical Report 5. GSARS Technical Report: Rome.*

_____ *2015f. Expert Meeting on Improving Methods for Estimating Crop Area, Yield, and Production under Mixed and Continuous Cropping, 15-16 April 2015. FAO Headquarters, Rome.*

_____ 2016a. *Findings from the Field in-Country Testing*. Technical Report 6. GSARS Technical Report: Rome.

_____ 2016b. *Strategy and Methodology for the Improvement of the Collection and Management of Administrative Data in an Agricultural Statistics System*. Technical Report 7. GSARS Technical Report: Rome.

Grondin, C. & Lavallée, P. 2001. *Survey of Employment, Payroll and Hours: An Update*. Statistics Canada Internal Document: Ottawa. Available at: http://www.oecd.org/std/30036134.pdf, 4/12/2017.

Guigo, M. 2008. *Micro-integration of different sources: Introduction and preliminary issues*. In *Proceedings of the ESSnet Statistical Methodological 98 Project on Integration of Survey and Administrative Data: Report of WP2. Recommendations on the use of methodologies on the integration of surveys and administrative data*. http://cenex-isad.istat.it/, 4/12/2017.

Hamer, H. 2013. *Weekly Crop Progress and Condition at USDA-NASS*. Proceedings of the *Sixth International Conference on Agricultural Statistics*, Rio de Janeiro, Brazil, 23-25 October 2013.

Hidiroglou, M.A., Rancourt, E. & Yung, W. 2007. *Administrative Data in Statistics Canada's Business Surveys: The Present and the Future*. Statistics Canada Internal Report.

Holt, D. 2007. The Official Statistics Olympic Challenge: Wider, Deeper, Quicker, Better, Cheaper. *The American Statistician*, 61(1): 1-8.

Republic of India. 2013. *Report of the Committee on Statistics of Agriculture and Allied Sectors.* Publication of the National Statistical Commission Ministry of Statistics and Program Implementation: New Delhi.

Government of India Planning Commission. 2001. *Report of the Working Group on Agricultural Statistics*. *Government of India Planning Commission Publication: New Delhi*

ISTAT, CBS, GUS, INE, SSB, SFSO & EUROSTAT, 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, 4/12/2017.

_____2008b. *Report of WP2: Recommendations on the Use of Methodologies for the Integration of Surveys and Administrative Data.* http://cenex-isad.istat.it/, 4/12/2017.

_____2008c. *Report of WP3: Software Tools for Integration Methodologies.* http://cenex-isad.istat.it/, 4/12/2017.

Isad, E. 2008c. ESSnet Statistical Methodology Project on Integration of Survey and Administrative Data Report of WP3: Software tools for integration methodologies. http://cenex-isad.istat.it/, 4/12/2017.

Iwig, W., Berning, M., Marck, P. & Prell, M. 2013. *Data Quality Assessment Tool for Administrative Data*. Available at: http://www.bls.gov/osmr/datatool.pdf, 4/12/2017.

Keita, N. & 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. Michigan State University Publication: East Lansing, MI, USA.

Kim, G. & Chambers, R. 2012. Regression Analysis under Probabilistic Multi-Linkage. *Statistica Neerlandica*, 66(1): 64-70.

Kim, J.K. & Shao, J. 2014. *Statistical Methods for Handling Incomplete Data.* CRC Press: Boca Raton, FL, USA.

Kizito, A. 2011. *The structure, conduct, and performance of agricultural market information systems in Sub-Saharan Africa.* Michigan State University, East Lansing, MI, USA (Ph.D. Dissertation).

Laitila T., Wallgren A. & Wallgren, B. 2011. *Quality Assessment of Administrative Data Research and Development*. Statistics Sweden Publication: Stockholm.

Republic of 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, 4/12/2017. National Statistical Office Publication: Lilongwe.

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 in Honour of Daniel Thorburn* (pp. 205–223). Available at: www.nordbotten.com, 4/12/2017.

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.

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.

Prell, M., Bradsher-Fredrick, H., Comisarow, C., Cornman, S., Cox, C., Denbaly, M., Wilkie Martinez, R., Sabol, W. & Vile, M. 2009. *Profiles of Success of Statistical Uses of Administrative Data*. Report of a subcommittee of the Federal Committee on Statistical Methodology, U.S. Office of Management and Budget. U.S. Office of Management and Budget Publication: Washington, D.C. Available at http://www.bls.gov/osmr/fcsm.pdf.

Rao, J. 2003. *Small Area Estimation.* John Wiley & Sons: Chichester, UK.

Sen, P. undated. *Challenges of Using Administrative Data for Statistical Purposes: India Country Paper*. Available at: https://unstats.un.org/unsd/dnss/docViewer.aspx?docID=2426.

Smith, J., Beaulieu, M., Kumar, E. & O'Neill, L. 2013. *Recent developments in the production of agriculture statistics.* Paper prepared for the Sixth International Conference on Agricultural Statistics, 23-25 October 2013. Rio de Janeiro, Brazil.

Srivastava, A. Undated. *Agricultural Statistics System in India*. Available at: http://www.iasri.res.in/ebook/TEFCPI_sampling/AGRICULTURAL%20STATISTICS%20SYSTEM%20IN%20INDIA.pdf, 4/12/2017.

United Republic of Tanzania. 2011. *Monitoring and Evaluation Framework; Revised Final Draft, March 201. Publication of the Agricultural Sector Development Programme (ASDP) M&E Thematic Working Group: Dodoma.*

_____ Bureau of Statistics. 2012. *Assessment of the Improved Agricultural Routine Data System (ARDS)*; December, 2012. Bureau of Statistics Publication: Dodoma.

_____ *Bureau of Statistics. 2012. Assessment of the Agricultural Routine Data System in Tanzania. JICA, December, 2012.* www.resakss.org/2014conference/docs/Tanzania_JSR_Assessment.pdf, 4/12/2017.

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

Trant, M. & Whitbridge, P. 2000. *Integration of Administrative Data with Survey and Census Data.* Agriculture and Rural Working Paper Series, 42. Statistics Canada Publication: Ottawa. Available at http://www5.statcan.gc.ca/olc-cel/olc.action?ObjId=21-601-M1999042&ObjType=46&lang=en, 4/12/2017.

Turtoi, C. Akyildirim, O. & Petkov, P. 2012. Statistical Farm Register in the EU Acceding Countries – A Conceptual Approach. *Economics of Agriculture* 1/2012.

Uganda Bureau of Statistics (UBOS). 2007. *The Development of the Agricultural Sector Strategic Plan for Statistics: A Data Collection Plan for Agricultural Statistics in Uganda*. Final Report to the Uganda Bureau of Statistics by the National Consultant: February 2007. UBOS Publication: Kampala.

United Nations (UN). 2011. *Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices*. Available at http://unstats.un.org/unsd/EconStatKB/KnowledgebaseArticle10349.aspx, 4/12/2017.

United Nations Economic and Social Council. 2007. Seminar on Increasing the Efficiency and Productivity of Statistical Offices: Pros and Cons for using Administrative records in Statistical Bureaus, *Fifty-fifth Plenary Session - Conference of European Statisticians*, 11-13 June 2007. Geneva.

United States Department of Health and Human Services (HHS). 2002. *Studies of Welfare Populations: Data Collection and Research Issues.* Publication of the U.S. Department of Health and Human Services: Washington, D.C.

_____ 2002. *Studies of Welfare Populations: Part II: Administrative Data.* Publication of the U.S. Department of Health and Human Services: Washington, D.C.

_____ 2002. *Studies of Welfare Populations: Access and Confidentiality Issues with Administrative Data*. Publication of the U.S. Department of Health and Human Services: Washington, D.C.

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.* In Proceedings of the American Statistical Association Section on Survey Research Methods, pp. 357-365. American Statistical Association Publication: Alexandria, VA, USA.

Wang, J. & Fuller, W.A. 2003. The mean squared error of small area predictors constructed with estimated area variances. *Journal of the American Statistical Association*, 98(463): 716-723.

World Bank, FAO & UN. 2010. *Global Strategy to Improve Agricultural and Rural Statistics.* World Bank Publication: Washington, D.C.

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

Appendices

Appendix 1: COSTS OF COMPILING ADMINISTRATIVE DATA

Assumptions:

1. In the local government, the levels that matter in the compilation of administrative data are the district, subdistrict and village levels.

1.1 The administrative data is generated at village level and consolidated at subdistrict level to facilitate planning at that level.

1.2 Mobile data collection devices are utilized at village level and the data is transmitted to the higher levels.

2. A standard district is assumed to have 10 subdistricts, and each subdistrict 50 villages. From every village, 20 households (10 households for crop farming and another 10 for livestock farming) are chosen. Hence the figure of 10 x 50 x 20 = 10,000 HHs for administrative data collection in a district.

3. Villages and subdistricts are each assigned an extension worker, while at district level there are three officers (agricultural extension worker, livestock extension worker and statistician).

4. Three motorcycles are required at district level and one at subdistrict level

5. The period of estimation is one year of the Administrative Data Compilation System

6. It was not possible to proceed to comparison with a survey or census budget, because this budget presents the cost of compiling administrative data in the launch of the administrative data collection. It is clearly cheaper to compile data in subsequent years, except for equipment replacements or updates and training.

1 Advocacy and communication: for sensitization and awareness creation

Sub-item	Description/ particulars	Comments	Quantity	Frequency: months/days/number/ times	Unit of measurement	Unit cost (Ugandan shillings)	Total amount in US$
1	Development of the sensitization materials	Consultations and coming up with appropriate materials	1	30	Days	194.03	5,820.90
2	Facilitation to Officials	National and local leaders	20	15	Days	104.48	31,343.28
3	Transport facilitation	Transport	10	10	Vehicles	20.90	2,089.55
4	Mobilization of local communities		10	2	Persons	14.93	298.51
5	Radio announcements	6 announcements for 3 days	60	3	Days	11.94	2,149.25
6	Worship places announcements	Once for 6 venues of worship	500	6	Days	14.93	44,776.12
7	Talk shows	Once for two radio stations	10	2	Days	179.10	3,582.09
8	Fliers and stationery	Consolidated	1	1		2,985.07	2,985.07
9	Driver	Drivers	10	10	Days	29.85	2,985.07
	Subtotal						**96,029.85**

2 Procurement

2.1 Procurements of Vehicles

Sub-item	Description/particulars	Comments	Quantity	Frequency/months/ days/number/times	Unit of measurement	Unit cost (Ugandan shillings)	Amount in US$
1	District vehicles	3 motorcycles	3	1	Times	2,388.06	7,164.18
2	Sub-district vehicles	1 motorcycle per sub-district	10	1	Households	2,388.06	23,880.60
	Subtotal						**31,044.78**

2.2 Procurement of data compilation equipment

Sub-item	Description/ particulars	Comments	Quantity	Frequency/months/ days/number/times	Unit of measurement	Unit cost (Ugandan shillings)	Amount in US$
1	Procure tablets/smart phones	A tablet to be procured for each village	500	1	Times	283.58	141,791.04
2	Procure Crop Card and other questionnaires (hard copies)	Every household in the village to obtain one questionnaire per year	500	200	Households	1.49	149,253.73
3	Procure computers with printers and scanners	Every district and subdistrict level should have 1 computer for data management	11	1	Computer	1,194.03	13,134.33
4	Procure assorted computer software	Software to run on the mobile data equipment and desktop computers	1	1	Unit	2,089.55	2,089.55
5	Procure internet	Tablets and computers	500	12	Months	14.93	89,552.24

	airtime/data						
6	Procure programming services	System design and support	10	2		74.63	1,492.54
	Sub-total						**397,313.43**

3 Training for administrative data compilation at various levels

Sub item	Description/ particulars	Additional information	Quantity/ number	Frequency/months/ days number/times	Unit measurement	Unit cost (Ugandan shillings)	Amount in US$
1	Trainees' allowance	Training for 3 days twice a year	560	6	Days	29.85	100,298.51
2	Training materials		780	1	Persons	8.96	6,985.07
3	Trainers' allowance	Training at sub-district level	10	8	Trainers	358.21	28,656.72
4	Training venues		10	6	Days	149.25	8,955.22
5	Transport costs		10	2	Times	104.48	2,089.55
6	Communicati on expenses		10	2	Times	4.48	89.55
	Subtotal						**147,074.63**

4 Data collection/compilation of administration data at various levels

4.1 Salaries and wages

Sub item	Description/ particulars	Additional information	Quantity/ number	Frequency/ months/days number/times	Unit of measurement	Unit cost (Ugandan shillings)	Amount in US$
1	Village-level compilation	Salary for village extension workers	500	12	months	208.96	1,253,731.34
2	Subdistrict supervisor	Salary for subdistrict extension workers	10	12	months	313.43	37,611.94
3	District-level Supervisors	Salary for district extension workers	3	12	months	417.91	15,044.78
	Subtotal						**1,306,388.06**

4.2 Field transportation (data collection and monitoring)

Sub item	Description/ particulars	Additional information	Quantity/ number	Frequency/ months/days number/times	Unit of measurement	Unit cost (Ugandan shillings)	Amount in US$
1	Transport costs subdistrict level	24 trips	10	24	trips	29.85	7,164.18
2	Transport costs district level	12 trips	10	12	trips	104.48	12,537.31
3	National-level monitoring visits	12 trips	10	12	trips	417.91	50,149.25
	Subtotal						**69,850.75**

5 Data analysis and report writing

-

Sub item	Description/ particulars	Additional information	Quantity/ number	Frequency/ months/days number/times	Unit of measurement	Unit cost (Ugandan shillings)	Amount in US$
1	Capacity building	Training, mentoring and attachment	12	1	Training	597.01	7,164.18
2	Assorted stationery		12	1	Stationery	149.25	1,791.04
3	Validation meetings	Meeting to discuss the results attended by subdistrict officials	12	1		746.27	8,955.22

	Subtotal						**17,910.45**
6	**Data dissemination and policy engagement**						-
Sub item	**Description/ particulars**	**Additional information**	**Quantity/ number**	**Frequency/ months/days number/times**	**Unit of measurement**	**Unit cost** (Ugandan shillings)	**Amount in US$**
1	Report production and dissemination	Dissemination through hard copies, internet and workshops	20	2	Times	1,492.54	59,701.49
	Subtotal						**59,701.49**
	Grand total						**2,125,313.43**

Appendix 2: Prototype Memorandum of Understanding

MEMORANDUM OF UNDERSTANDING

Between
Center for Survey Statistics and Methodology (CSSM)
College of Liberal Arts and Sciences (LAS)
Iowa State University
United States
and

National Bureau of Statistics,

United Republic of TANZANIA

for

Guiding the conducting of the Pilot Study on Improving Administrative Data for Agricultural Statistics

The Memorandum of Understanding is entered into on this ……...day of May 2016 between the Center for Survey Statistics and Methodology, College of Liberal Arts and Sciences, Iowa State University, Ames, Iowa, United States of America (referred to as "ISU");

And

The National Bureau of Statistics, Tanzania (referred to as "Bureau").

WHEREAS ISU is desirous of conducting a Pilot Desktop Study on improving administrative data for agricultural statistics using data from Tanzania, and intends to conduct an analysis directed toward combining multiple data sources,

AND WHEREAS the Bureau intends to reach the objective of collaboration through active participation in the partnership

NOW THEREFORE, THIS MEMORANDUM OF UNDERSTANDING WITNESSES AS FOLLOWS:

Article 1

Purpose

The purpose of this Memorandum of Understanding is to facilitate the working relations of the Bureau and ISU as partners in conducting a pilot study on the role of administrative data in an integrated agricultural statistics system in developing countries. These activities are funded by the Global Office of the Global Strategy to improve Agricultural and Rural Statistics, which is hosted by FAO Statistics Division.

Article 2

Responsibilities of the Center of Survey Statistics and Methodology,

Iowa State University (ISU):

a. Work with the Bureau to determine appropriate data sources;
b. Specify a request for data from the Bureau;

c. Document the process of defining the data;
d. Analyse the obtained data;
e. Document the data analysis.

Article 3

Responsibilities of the National Bureau of Statistics, Tanzania:

a. Will coordinate the implementation of the Study in Tanzania especially with the Ministry of Agriculture & Food Security, Ministry of Livestock & Fisheries, and the Ministry of Local Government;
b. Will provide information to enhance the understanding of the Tanzanian Agricultural Statistics System;
c. Will assist ISU in identifying and defining appropriate data sources;
d. Will provide ISU with agreed upon data;
e. Will assist ISU in understanding data elements after data transfer;
f. Will discuss with ISU the Bureau's objectives, such as variables of interest, with respect to administrative data.

Article 4

Commencement and duration

This Memorandum of Understanding shall come into force on the date of signature by all parties for a term of one year. If the responsibilities are not completed at the end of the term, it will be automatically renewed, unless terminated earlier in accordance with the terms.

Article 5

Termination

1 This Memorandum of Understanding may be terminated by any **Partner** upon such breach of this Memorandum of Understanding by any other **Partner,** if the disagreement as **AGREED** cannot be resolved by dialogue. A breach includes but is not limited to disagreement on fundamental principles of the Memorandum of Understanding and/or failure or inability of any **Partner** to fulfil their obligations without reasonable cause;

2 Any **Partner** may terminate this Memorandum of Understanding by giving the other **Partners** One (1) month written notice to this effect;

3 Upon receipt of notice to terminate this Memorandum of Understanding, the **Partner** receiving the notice will take all necessary steps to terminate activities governed by this contract and shall incur no further costs or undertake any commitments;

Article 6

Miscellaneous

1. This Memorandum of Understanding and the obligations hereunder will remain in effect for so long as the Memorandum of Understanding remains effective.
2. No provision of this Memorandum of Understanding may be amended, assigned, suspended, abrogated, terminated or waived without the prior consent of both parties.

Article 7
Amendments

Any alternation to this Memorandum of Understanding shall be made with written consent of all parties.

Article 8

Dispute Resolution

The parties shall use their best efforts to settle amicably any dispute, controversy or claim arising out of this Memorandum of Understanding in the event of failure to have any dispute settled amicably the agreement shall be terminated.

IN WITNESS WHEREOF**,** the parties hereunto have caused this agreement to be executed the day and year above written.

For the Center of Survey Statistics and Methodology, Iowa State University, USA

Signed by:

Name: …….

Signature: ……...

Designation: ……

Date: …….

Signed by:

Name: …….

Signature: ……...

Designation: ……

Date: …….

Witnessed by:

Name: …….

Signature: ……...

Designation: ……

Date: …….

For the National Bureau of Statistics, TANZANIA

Signed By:

Name: …….

Signature: ……...

Designation: ……

Date: …….

Witnessed by:

Name: …….

Signature: ……...

Designation: ……

Date: …….