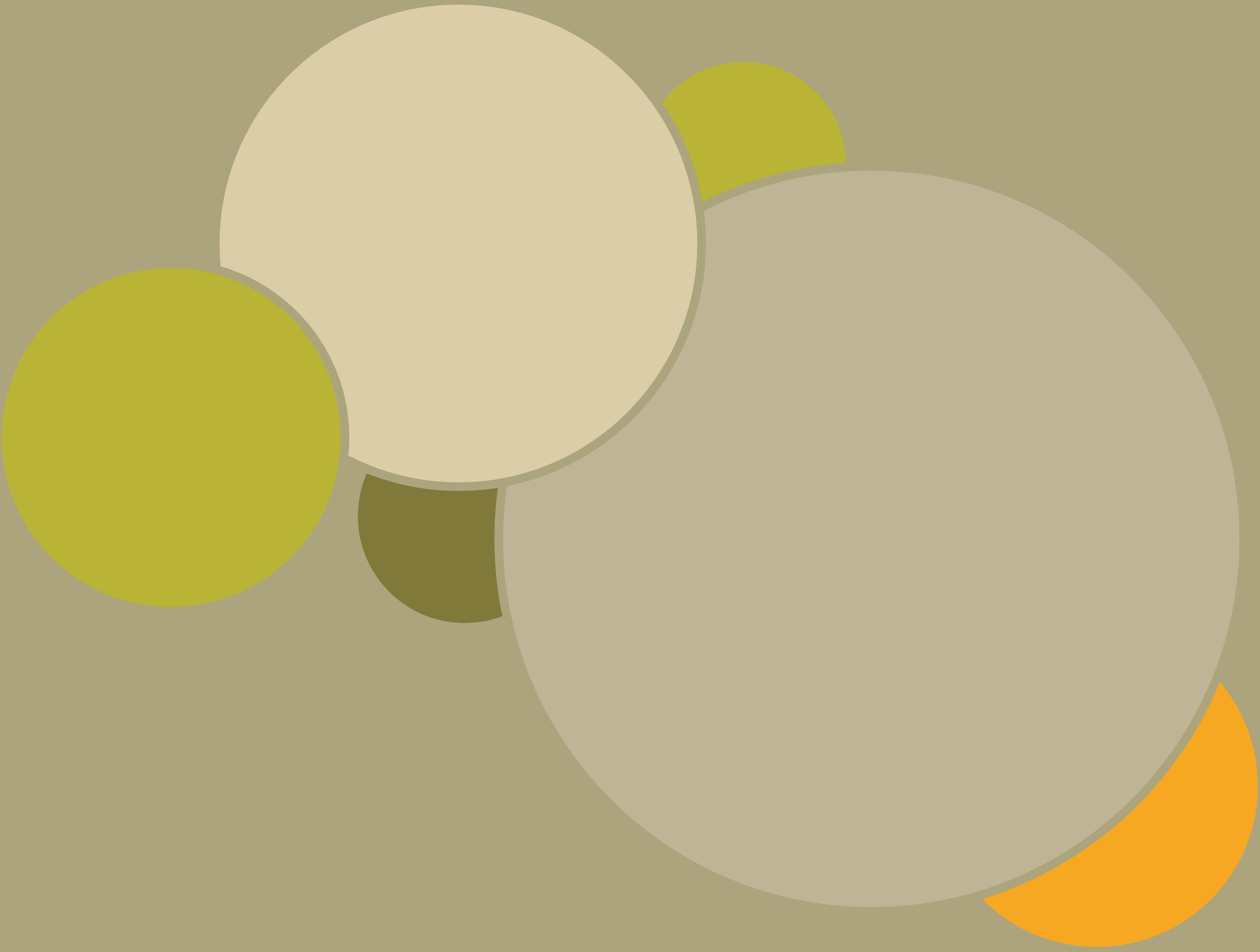
IMPROVING
AGRICULTURAL
& RURAL STATISTICS
Global Strategy

Guidelines on improving and using administrative data in agricultural statistics

Guidelines on improving and using administrative data in agricultural statistics

February 2018



Contents

Tables and figures

TABLES

FIGURES

Acronyms

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
ARSA	Administrative Reporting Systems for Agriculture
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 National 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)
MOU	Memorandum of understanding
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
NASO	National Agricultural Statistics Office
ONDR	*Office National 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
PMORALG	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

Acknowledgments

This document is mainly based on the findings of the ADMIN project of the Global Strategy, implemented by 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 list of key experts from the two institutions involved in the project and all other contributors can be found in the final technical report of ADMIN: GSARS (2017).

The Guidelines were prepared by Dramane Bako, International Consultant – Statistician of the Global Strategy to improve Agricultural and Rural Statistics, under the supervision of Flavio Bolliger, Research Coordinator of the Global Strategy. The following experts of the Global Strategy and FAO have reviewed the document and provided valuable inputs: Naman Keita (Senior Consultant in Agricultural Statistics, Global Strategy), Valerie Bizier (Global Strategy Technical Assistance and Training Coordinator), Christophe Duhamel, (Coordinator of the Global Office of the Global Strategy), Carola Fabi (Statistician, FAO Statistics Division) and Gladys Qacha (Agricultural Statistics Expert, Global Strategy).

This publication was prepared with the support of the Trust Fund of the Global Strategy, funded by the UK's Department for International Development (DfID) and the Bill & Melinda Gates Foundation.

It was formatted by Laura Monopoli, under the supervision of Norah DeFalco.

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 Program, 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.

The Global Strategy has implemented an important line of research on Improving the Methodology for Using Administrative Data in an Agricultural Statistics System (ADMIN), one of the key priorities of the Research Program. As part of ADMIN, eight technical reports have been developed. These are available at http://gsars.org/en/tag/administrativedata/.

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. The ultimate goal of this document is to provide operational guidance to developing countries on how to set up an effective Administrative Data System for Agricultural Statistics (ADSAS), as well as on the improvement, use and integration of administrative data in the national statistical system. The concept of ADSAS refers to the *set of all administrative institutions producing administrative agricultural data that may be used for the purposes of agricultural statistics and providing them to the national institution in charge of agricultural statistics for official use and publication*.

These Guidelines on improving and using administrative data in agricultural statistics have large areas of overlap with the final technical report of ADMIN. The findings conveyed in the other ADMIN technical reports and additional operational inputs from the literature were used to develop the document.

Executive summary

The definition of administrative data adopted for this document describes these 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”[1]. This definition encompasses two classes of administrative data, designated respectively by the term "**Traditional Administrative Data**" **(TAD)** – which includes information collected through taxation, regulatory processes (farm inspections), farm assistance programs (subsidies and insurance), etc. – and **“Administrative Reporting Systems for Agriculture” (ARSA)**, which refer to the data routinely collected by extension officers. The properties of these two types of data are different because of the differences between the data collection methods.

The concept of Administrative Data System for Agricultural Statistics (ADSAS) refers to the set of all administrative institutions producing administrative agricultural data that may be used for the purposes of agricultural statistics, and that provide them to the national institution in charge of agricultural statistics for use and official publication. This institution is usually the National Statistics Office (NSO), Central Statistics Office (CSO) or the department of agricultural statistics in the ministry in charge of agriculture, which in these Guidelines will be identified with the acronym NASO (National Agricultural Statistics Office).

Administrative data have various benefits; as discussed in GSARS (2015a), these include cost savings, reduced respondent burden, and improvements in the efficiency of macro-level estimators and small-area statistics. Therefore, the establishment of an effective ADSAS can contribute to the improvement of the quality, availability and accessibility of agricultural statistics.

The design of the ADSAS begins with the identification of the relevant administrative sources that will be part of the system. In many developing countries, basic agricultural administrative data (that is, on crops, livestock, fisheries and forestry) are collected and managed under the ministries of agriculture, livestock, fisheries and/or forestry. However, in many other countries, parastatal organizations also produce administrative data, especially on commercial or cash crops. Private-sector agencies or organizations also often collect and manage various forms of administrative data, for example on prices, marketing, inputs, etc., especially following the restructuring policies adopted in many of these countries. These agencies may collect and manage the data without any direct participation of the relevant NSO or CSO. 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 aspects such as crop area and livestock numbers. 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. Examples of private-sector organizations that may provide administrative data relevant to agriculture include the following: licensing or regulatory bureau, grain associations, commodity associations, cooperatives, factories, slaughterhouses and distributors of agricultural inputs.

At country level, the identification of potential administrative data sources generating information that is related to the agricultural sector and usable for agricultural statistics may not be straightforward. A practical way to establish a good inventory of agricultural administrative data sources and link them to the core data items may be the organization of a qualitative survey on the agricultural information processed or produced within administrative structures, and then holding a national workshop on the results of the survey. This workshop should lead to a final list of all administrative sources that can contribute to agricultural statistics.

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

The quality assessment of the ADSAS consists in a structural diagnosis of the system and a data quality assessment. The structural diagnosis consists mainly in analysing the main structural issues faced by institutions producing agricultural administrative data, such as the lack of harmonization of concepts and definitions or legal and political constraints. A framework is proposed to assess the quality of administrative agricultural data. The quality dimensions include the following: relevance, accuracy, accessibility, confidentiality and privacy protection, coherence, timeliness, punctuality, and comparability.

Statistical agencies in developed countries have developed mechanisms to ensure data quality that can be applied in developing countries. Engaging with administrative offices and with the public (for example, regarding the sharing of sensitive administrative data) 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. Due to the importance of pooling multiple data sources to overcome undercoverage 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. 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. Audits and sampling of administrative data are used to check for errors and evaluate coverage. Adopting best practices for quality control and assurance can help to manage errors in administrative data.

Administrative data may be used directly, as the final statistical product, or indirectly, in forming the statistical product. The direct use consists of publishing directly tabulations of administrative data as final products. This kind of use is recommended only under a number of conditions, which include the complete coverage of the target population of interest. Statistical offices often indirectly use administrative data to support survey or census programs. Administrative data can play a role in all stages of the survey process: frame construction, data collection and estimation (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 (that is, 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 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. Institutional arrangements are crucial to ensure a better quality of and access to administrative data. They 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. 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.

Considering the potential uses of administrative data in improving the availability and quality of agricultural statistics, the proper integration of these data in agricultural statistics system is recommended. These data should be taken into account in the national system or strategy related to the production of agricultural statistics. A long-term perspective is the development of a register-based agricultural statistical system through the development or improvement of key registers in the country.

1

Designing an ADSAS

This chapter presents the main aspects to consider for the design of an ADSAS at country level. It begins with the definition of administrative data adopted in these Guidelines, after briefly reviewing the many definitions available (see section 1.2, GSARS (2015a) and section 1.1, GSARS (2015b)). The typologies and common data sources of agricultural administrative data are outlined, and the findings from the field tests of the ADMIN research conducted by the Global Strategy on identifying administrative data sources are highlighted. Finally, operational guidance is proposed on how to design an ADSAS for countries.

1.1 DEFINITION OF ADMINISTRATIVE DATA AND ADSAS

1.1.1 Administrative data

Administrative data can be derived from diverse sources, including government records, reporting systems and private organizations. This indicates 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. Perhaps the most comprehensive of the traditional definitions was set out by Gordon Brackstone of Statistics Canada, who identified the following four distinguishing features of administrative data (Brackstone, 1987):

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

The United Nations (UN; see UN, 2011) consider a traditional or "narrow" definition of administrative source that comprises only public-sector non-statistical sources, whereas a wider definition would also include private-sector sources. Thus, under the narrow definition, administrative sources are a subset of secondary sources, while under the wider definition, these terms are synonyms. There is a growing number of reasons for favouring the wider definition, including: (a) increasing privatization of government functions; (b) register-based population statistics (UN, 2011), growth of private-sector data and "value-added resellers"; and (c) user interest in new types of data (Brackstone, 1987). 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.

In these Guidelines, we adopt the definition used by GSARS (2017a) for administrative data in the context of agricultural statistics: *"information collected primarily for administrative (not statistical) purposes by government departments and other organizations usually during the delivery of a service or for the purposes of registration, record keeping or documentation of a transaction"*[1].

1.1.2 The ADSAS

The adopted broader definition of administrative data in the agricultural and rural context encompasses two large classes of these 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. When producing official statistics, well-developed statistical systems make extensive use of this first type of administrative data. For the second type, mainly found in developing countries, an extension officer (when delivering assistance services to farmers), a village chief or other type of agricultural field officer makes a determination based on his or her observations and expert judgment and routinely produces a report on crops or livestock in his or her area of work.

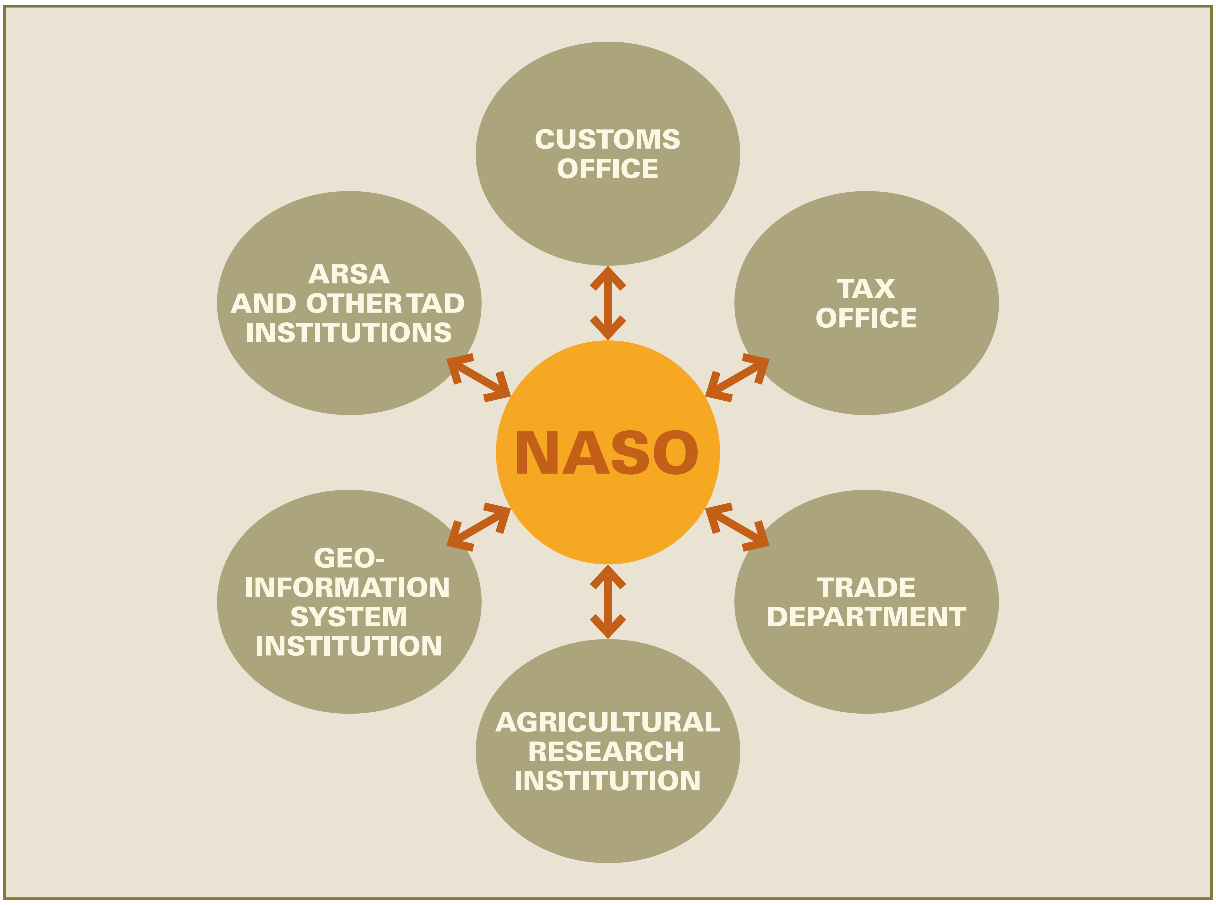
Two terms are introduced to distinguish these two classes of administrative data. The first one is designated by the term "**Traditional Administrative Data**" (**TAD**). In the context of agriculture, TAD include information collected through taxation, regulatory processes (that is, farm inspections), farm assistance programs (subsidies and insurance) and monitoring programs (livestock tracing systems). The term **"Administrative Reporting System for Agriculture" (ARSA)** describes the second class of data. The properties of these two types of data differ because of variations in the data collection methods. A great volume of research and methodological works exist on TAD. However, this is not the case with ARSAs; for this reason, this second class of administrative data is emphasized in these Guidelines, because existing extension programs routinely collect information on agriculture in many developing countries.

The **ADSAS** is defined as the set of all administrative institutions producing agricultural administrative data that may be used for agricultural statistics purposes, and providing them to the national institution in charge of agricultural statistics, for official publication. Depending on the statistical system of the country, this institution may be the National Statistics Office (NSO), the Central Statistics Office (CSO) or the department of agricultural statistics within the ministry in charge of agriculture. For simplification, in these Guidelines, the relevant institution will be identified as the National Agricultural Statistics Office (NASO). Similarly, the NSO and CSO will simply be indicated with the acronym NSO.

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

ADSAS cover both TAD and ARSAs. The NASO is the lead institution of an ADSAS, collecting data from the administrative sources and using them to produce or improve the agricultural statistics in the country. In return, the administrative sources involved in the ADSAS could benefit from the NASO technical supports or specific statistics corresponding to their needs. The NASO may also be charge of ARSA data compilation.

FIGURE 1. REPRESENTATION OF ADSAS.


CUSTOMS OFFICE
ARSA AND OTHER TAD INSTITUTIONS
TAX OFFICE
NASO
GEO-INFORMATION SYSTEM INSTITUTION
TRADE DEPARTMENT
AGRICULTURAL RESEARCH INSTITUTION

1.2 IDENTIFYING AGRICULTURAL ADMINISTRATIVE DATA SOURCES

In most developing countries, basic agricultural administrative data (on crops, livestock, fisheries and forestry) is collected and managed under the ministries of agriculture, livestock, fisheries or forestry. However, in many other countries, parastatal organizations produce administrative data, especially on commercial or cash crops farms. 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, on prices, marketing, inputs, etc. These agencies may collect and manage the data without any direct participation of the relevant NSO.

In the context of agriculture, the administrative sources include (GSARS, 2015a):

- 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;
- lists of agricultural production and inputs from manufacturers and distributors;
- farm registers and other registration or licensing systems;
- records on agro-tourism;
- lists maintained by farmers' associations;
- private businesses' data;
- meteorological data;
- and traceability data, such as traceability livestock data.

This section provides an overview of the main data sources of the two categories of administrative data.

1.2.1 TAD

The main sources of TAD are reviewed below.

Soil information

Topographical maps and maps of soil characteristics are often maintained in administrative processes. For instance, 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 for example Goebel, 2009) and as auxiliary information in constructing estimates.

Crop insurance and subsidy programmes

Government assistance programmes generate administrative data. Subsidies and crop insurance programmes can collect information such as areas planted with particular crops (Carfagna and 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.

Some examples from developed countries include:

The Integrated Administrative Control System (IACS) of EUROSTAT: 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).

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."

Land registration and cadastral records

A cadastre consists of records defining the "extent, value and ownership of land" (Bins and Dale, 1995). Maintained by several developed countries and some developing 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 cadastres and contain a wealth of information on land use, including crop management.

For instance, India has a decentralized statistical system in which tasks are distributed among 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 percent 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-percent subsample instead of all crop areas, is to accelerate data collection (Republic of India, 2013).

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 censuses 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.

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 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. Some examples are provided below:

- An example of a database generated to regulate an industry that is also 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 animal 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).
- 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.

Private sector and associations

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).

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).

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.

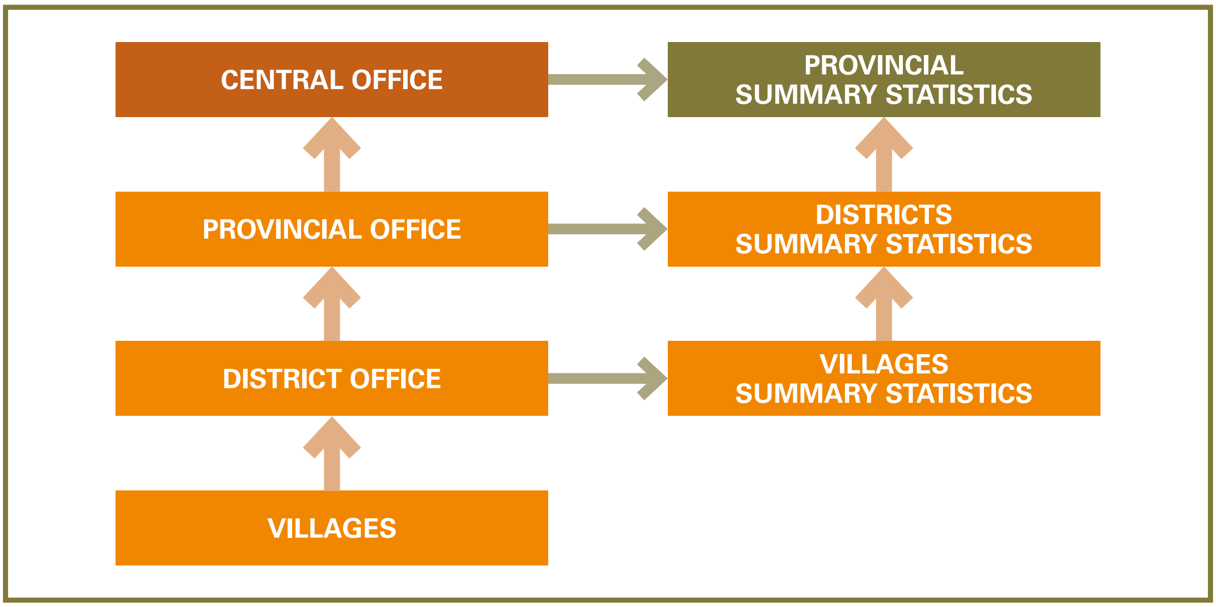
1.2.2 Administrative reporting systems for agriculture

Experts and subject matter specialists 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 and Chin, 2013; Galmes, 2013; Hamer, 2013). These reporters include experts 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 these Guidelines 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 is often administered through the relevant ministries for agriculture, livestock, forestry and fisheries (MALFs) on a regular (weekly, monthly or annual) basis. Often, these administrative reporting systems even provide data on the smallest administrative units, such as districts or villages. Examples of routine systems in developing countries are provided below:

- 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 several regions and districts, with the objective of meeting the data needs for monitoring and evaluation of 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.
- Namibia's MAWF administers a questionnaire called the Crop Assessment Checklist (also known as the Cereal Production Checklist). This checklist enables the gathering of various qualitative and quantitative information: weather and crop conditions, percentage of area planted to cereal crops, estimation of production, etc.
- The statistical system of the Lao People's Democratic Republic is decentralized and involves several institutions, each with its specific assignment. The Agricultural Statistics Yearbook is the annual publication of the Department of Planning and Cooperation of the Ministry of Agriculture and Forestry (MAF); it 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 field personnel assesses crop production by observing harvests and interviewing key informants (generally, farmers and village heads) in their localities.

FIGURE 2. TYPICAL ADMINISTRATIVE REPORTING SYSTEM IN ASIA.


CENTRAL OFFICE
PROVINCIAL SUMMARY STATISTICS
PROVINCIAL OFFICE
DISTRICTS SUMMARY STATISTICS
DISTRICT OFFICE
VILLAGES SUMMARY STATISTICS
VILLAGES

Source: Maligalig, 2017.

1.3 DESIGN OF THE ADSAS

At the country level, the identification of potential administrative data sources working with information related to the agricultural sector and usable for agricultural statistics may not be straightforward. Some examples of sources have been provided in this chapter, with practical descriptions for some countries. However, this list is certainly not exhaustive for many countries, and not every country will have access to these sources. After compiling an inventory of the relevant agricultural administrative data sources, the next important step is linking the sources to core data items and other important national data requirements.

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 livestock inventories, and data from distributors provide information on dairy production (see table 1 below).

TABLE 1. SOURCES OF ADMINISTRATIVE DATA FOR SELECTED CORE DATA ITEMS OF THE GLOBAL STRATEGY.

Core data items	Administrative data type	Example
Crops • planted area, harvested area, yield, yield, storage, labour, prices • maize, barley, wheat, sorghum, rice, cotton	Farm subsidies	IACS contains crop areas for the crops enrolled for subsidies
	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
Fishery • List of large vessels	Administrative records of fishing boats/vessels	European Union: most of the fishery data compile by EUROSTAT come from administrative data, such as national registers of fishing vessels

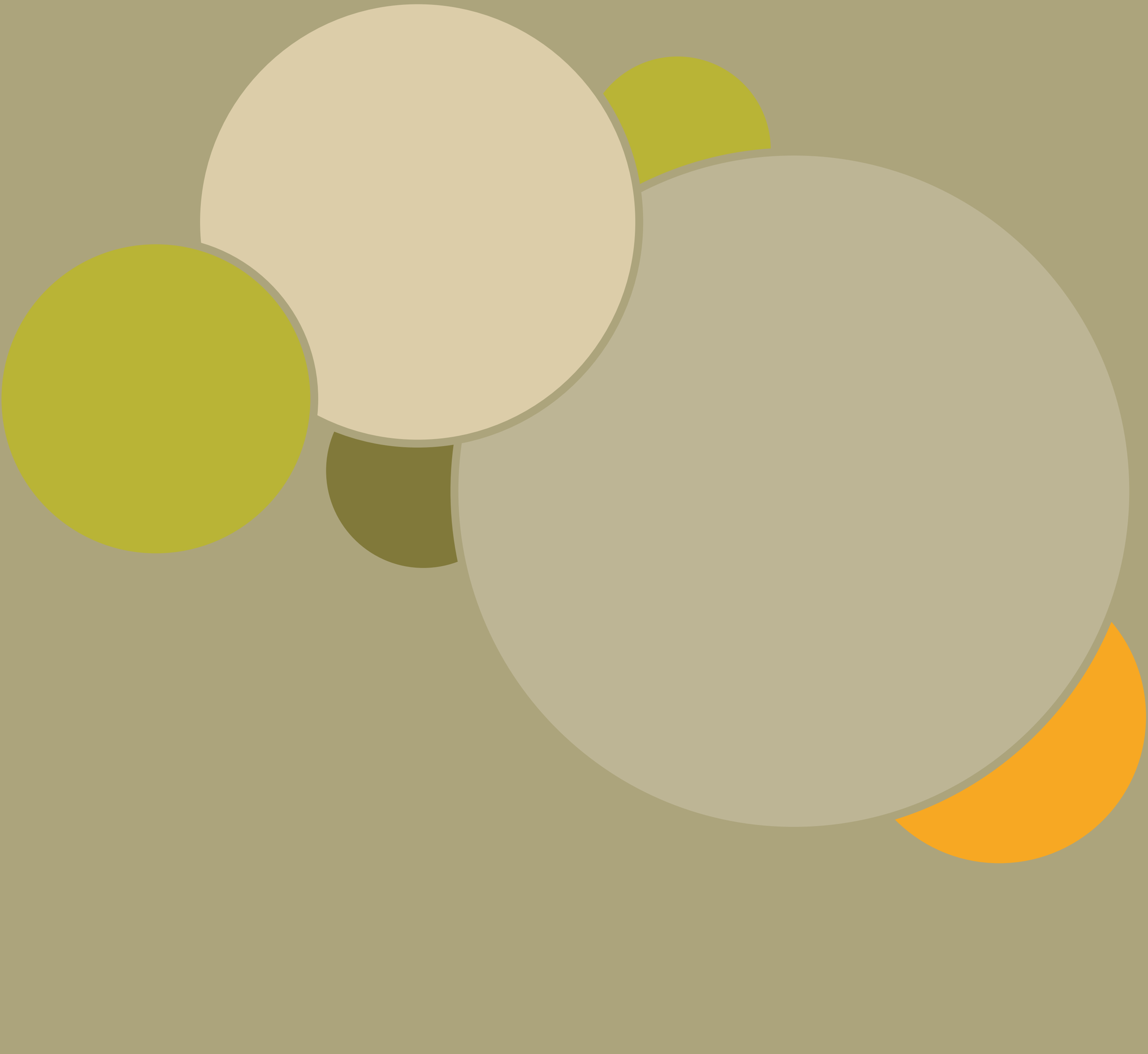
As stated above, the NASO is the lead institution of the ADSAS. Accordingly, this institution should play a central role in designing the ADSAS.

A practical way to establish a good inventory of agricultural administrative data sources and link them to the core data items may envisage the following steps:

- Organize a qualitative survey on the agricultural information processed and/or produced in administrative structures in the public sector (public administration), civil society (NGOs, farmers associations, etc.) and private sector; and then
- Organize a national workshop to illustrate the results of the survey, with representatives of the potential administrative sources identified and their potential users (NSOs, ministries of agriculture, etc.). Such a workshop will enable clarification of the results of the inventory survey and evaluation of their strengths, weaknesses and suitability for use in agricultural statistics, within an integrated and cost-effective agricultural statistical system. This workshop should lead to the elaboration of a final list of all administrative sources that can contribute to agricultural statistics.

SUMMARY

This chapter presents the definitions adopted in these Guidelines for the concepts of administrative data and ADSAS. The common administrative agricultural data sources for both categories of administrative data are illustrated. Operational steps for designing an ADSAS are proposed under the lead of the NASO, starting with an inventory of administrative data sources and linking them to core data items. This can be done through a qualitative survey followed by a national workshop. The next step will consist in assessing the quality of the data produced by these steps.



2

Quality assessment of the structure and data of the ADSAS

Quality assessment is a critical step to identify quality issues affecting the ADSAS in view of the improvement of the country's agricultural administrative data. The results of the assessment should be disseminated to the users of data from the ARSA and the TAD, as well as to the data collectors. This is one way of raising awareness among the various users, including government agencies, on the importance of good-quality data to inform policy-making and policy monitoring (Maligalig, 2017). Consequently, there could be support within governments to transition or combine survey results and ARSA and TAD data, for the ultimate purpose of improving the quality of the country's agricultural and rural statistics.

2.1 STRUCTURAL DIAGNOSIS

The structural diagnosis consists mainly in analysing the main structural issues faced by the agricultural administrative data producers.

Are concepts and definitions harmonized among agricultural administrative data sources?
As discussed in chapter 1, data are produced by various institutions. Often, different concepts and definitions are used, which may lead to the data on the same item being different.

In this regard, it appears necessary to examine all concepts and definitions adopted by the target institutions producing agricultural administrative data to identify and understand potential differences and explore ways of standardization.

Are the staff involved in administrative data production well-qualified?
Many ministries, departments and agencies (MDAs) charged with the collection and management of data have staff at headquarters and in the field (extension staff and village chiefs, or even enumerators). However, in many developing countries, well-qualified staff often cannot be retained, due to poor working conditions and incentives; this results in a shortage of qualified staff. All of these factors contribute to the low quality of the data generated. A second weakness is that the field staff are often poorly supervised.

Thus, an assessment of the qualifications of the human resources involved in working with the agricultural administrative data will be helpful in improving data quality.

How are data collection and processing controlled?
Unlike statistical surveys and censuses, administrative data are gathered for purposes that differ from the objectives underlying a statistical operation. The data may be collected by individuals in non-controlled settings, without enforcement of the strict protocols that dictate the data collection processes of carefully implemented surveys and censuses.

Usually, 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 village chiefs. The statistical 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 and Carfagna, 2010).

An assessment of the processes of data collection and entry is recommended. In particular, the existence of protocols related to these processes and mechanisms of quality control within the administrative sources should be verified.

Are there legal and political constraints?
Access to administrative data may be limited by legal and political constraints, which may be in place for good reason (Brackstone, 1987). One dimension of these constraints is that of confidentiality. The NASO is under an inherent commitment to preserve the confidentiality of the statistical data. The complexity of these requirements increases, however, when administrative data, which are collected and maintained by other agencies, are considered.

For each agricultural administrative data producer, it is important to identify any legal and political constraints regarding access to and the use and publication of the target data.

2.2 ASSESSMENT OF ADMINISTRATIVE DATA QUALITY

A quality assessment framework for agricultural administrative data is proposed. Particular focus is placed on the use of audit sample surveys.

2.2.1 Quality assessment framework

GSARS (2016a) proposes a quality assessment framework for administrative agricultural statistics. The framework can be used to assess the quality of both ARSAs and TAD. As part of the Global Strategy's ADMIN research project, it was used in the in-country testing to assess the quality of the ARDS of the United Republic of Tanzania, which is an ARSA. Table 2 below illustrates the quality dimensions that were considered relevant to administrative data and how each dimension can be measured and assessed. Where possible, efforts were made to make these measures quantitative.

TABLE 2. 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? Is the data relevant, considering the object of measurement? Each administrative data set should be accompanied by metadata on its contents, so that users can assess the data set's suitability for their purposes.
Accuracy and Reliability	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 (such as coverage, sampling, response and nonresponse). The data should adequately represent the entire population (full coverage) and relevant subpopulations (disaggregation).	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, when possible. The administrative data should be compared to survey or even census data, whenever such data becomes available[1].

Dimension	Description	Evaluation method
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 for 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 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 and 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. Therefore, such data may not 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 data with other external sources, e.g. the data from satellite imagery. The use of satellite data for estimating land use statistics has become more prevalent with the increase in the availability of inexpensive satellite imagery (Maligalig, 2017). • Compare the approaches, classifications and methodologies used in administrative data collection and analysis with those used in censuses and surveys.
Timeliness and 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 users when needed.
Comparability	This dimension focuses on the validity of comparisons between administrative sources and census and survey data, 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 and census and survey data. This should also occur within the same source, over time and space.

[1] Trant, 2010.

The results of the quality assessment of the ARDS of the United Republic of Tanzania are presented in table 3.

TABLE 3. RESULTS OF THE ASSESSMENT OF THE ARDS OF THE UNITED REPUBLIC OF TANZANIA.

Dimension	Assessment based on the ARDS in the United Republic of Tanzania
Relevance	***Limitations:*** • Not all villages are covered. • Village diaries are not standardized • National-level technical stakeholders – crop and animal products specialist areas – do not seem keen to use data generated through the ARDS. Crop and animal products specialists continue to collect their own data for planning and policy formulation.
Accuracy and Reliability	• Harmonized reporting formats for all regions. • Reporting frequencies harmonized. • Village diaries not standardized. • Random selection used. • Production data especially are more of estimates than actual production data from the households. The village extension workers are supposed to measure; however, they do not usually do so. The data tend to be merely guesstimates provided by village extension workers. • Inadequate supervision. • No data verification process. • The regional managers certify the data entered at the district level by checking the correctness of the information. • Very few districts are currently reporting (18 percent), because of failure of the software on one hand, and on the other, lack of knowledge by staff at the lower levels of the software used for data capture and compilation. Insufficient field staff and working conditions, such as lack of transportation, also contribute to the low reporting rate. • Data is compiled by village, ward, district, region and national levels.
Accessibility Confidentiality and privacy protection	• The data is available on a website and on CD-ROMs; however, accessibility is still an issue. • The website captures the number of people accessing the website to access information. • Absence of metadata
Coherence and Consistency	Lack of comparability of ARDS data and census and survey data due to differences in the instruments of data collecting, timing of data collection, methodology, reference periods, etc.
Timeliness and Punctuality	• Data is compiled on a monthly, quarterly and annual basis at village, ward and district levels respectively; and compiled on quarterly and annual basis at regional and national levels. • Reports are often either late or not regularly submitted leading to a low percentage of 18 percent received.
Comparability	Lack of comparability between ARDS data with census and survey data due to differences: in instruments of data collecting, timing of data collection, methodology, reference periods, etc.

2.2.2 Auditing administrative data quality through a sample survey

An audit of both ARSA and TAD data quality can be performed through a sample survey, to assess their accuracy and consistency.

Regarding TAD, 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 (GSARS, 2017).

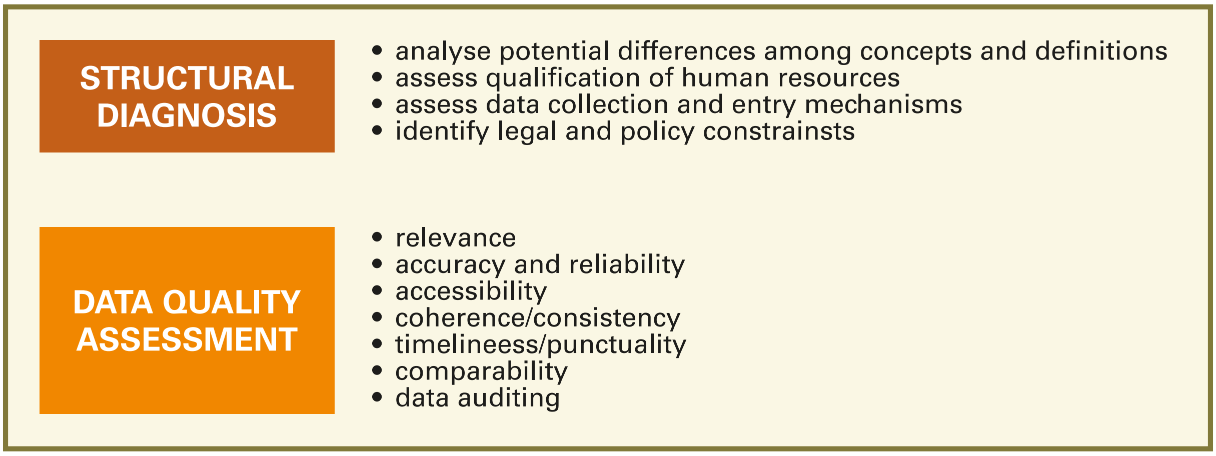
For the ARSA, a parallel sample survey can be used to check the quality of the reporting data at various levels: village, province, country, etc. Maligalig (2017) proposes a sampling procedure to audit the quality of ARSA data:

- Villages can be sampled with probability, so that data from all sampled villages can be aggregated to provide, for example, district-level estimates.
- Staff of the agricultural statistics unit (ASU) at the ministry of agriculture can train the data collectors or respondents of the sample villages on standard concepts and definitions, as well as on the importance of providing accurate and timely data to policy-making and monitoring bodies.
- ASU staff can retrieve a copy of the sample village-level questionnaire or form, and consolidate these into a data file so that district-, provincial- and national-level estimates can be derived using the inverse of the village selection probabilities as weights.
- These estimates can then be compared with the results of the ARS. ASU staff should then closely examine the areas with large discrepancies.

SUMMARY

This chapter discusses the administrative data quality assessment, which has two dimensions: a structural diagnosis and a quality assessment of the data produced by administrative structures. A quality assessment framework is provided, and illustrated with the results of an application on the ARDS of the United Republic of Tanzania. The auditing of the data quality through a sample survey is also discussed.

FIGURE 3. THE TWO DIMENSIONS OF ADMINISTRATIVE DATA QUALITY ASSESSMENT.



STRUCTURAL DIAGNOSIS

- analyse potential differences among concepts and definitions
- assess qualification of human resources
- assess data collection and entry mechanisms
- identify legal and policy constrainsts

DATA QUALITY ASSESSMENT

- relevance
- accuracy and reliability
- accessibility
- coherence/consistency
- timelineess/punctuality
- comparability
- data auditing

3

Improvement of the administrative agricultural data quality

As administrative data are collected to meet the specific aims of the administrative source 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. Therefore, engaging in a rigorous quality control process, enhancing the data collection, entry and processing approaches, and addressing human and financial resources issues, are all key processes to improving administrative data before their use in agricultural statistical systems.

3.1 METHODOLOGICAL TOOLS TO ADDRESS QUALITY ISSUES

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 relating to quality control, multiple data sources, data collection, and data storage and dissemination/diffusion are covered, considering the need to ensure harmonization and consistency among various data sets to ensure comparability.

3.1.1 Quality control

A lack of control over the administrative data collection process may lead to inconsistent data formats. The NASO may not have any control over data collection, and different data collection efforts may use different formats. Furthermore, administrative agencies may change data collection protocols over time or may use paper questionnaires, that may result in processing errors when the information is digitized.

The NASO in charge of agricultural statistics can provide helpful support to agricultural administrative data producers to improve the control of both data collection and data entry. Independent professional statisticians may also contribute, providing diagnoses of quality issues regarding the data collection process and technical advice for improvement. However, the best solution would be the recruitment of statisticians by these administrative institutions.

Administrative Reporting Systems for Agriculture (ARSAs)

In ARSAs, the quality control protocol should be designed to guide the monitoring of the data collection, data compilation and data flow processes through the system from the grassroots up to the national level. For instance, the protocol should establish when the data should be collected; when the supervisors should check on the data collectors and what is their responsibility; when the supervisors should submit reports to the district offices; when the district officers should submit reports to the regional offices; etc. There should be established and observed schedules for consultations or meetings to discuss issues arising from the processes. This will ensure continuity and sustainability.

Efforts to monitor the administrative data collection process more closely can also lead to improved data quality. Galmes (2013) provides recommendations to improve the quality of data obtained from ARSA data collection processes including:

- use of a standardized format for collecting information;
- preparation of manuals containing clear definitions of activities to perform;
- periodic training of data collectors; and
- strong supervision.

Implementing such quality control standards in data collection can also improve the transparency of data collection procedures. This not only enhances the quality of the data; it also aids users (both statistical analysts and consumers of statistical products) in interpreting and understanding how to appropriately use the data.

Traditional administrative data system

Quality control procedures for evaluating measurement errors and coverage problems are necessary to protect against bias. Measurement errors are associated with 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 interest to a statistical agency. For example, administrative processes entail collecting information on the beneficiaries of unemployment insurance, a concept that is related to unemployment; however, the definition of unemployment adopted in such processes may differ from that used by statistical agencies. False reporting can also stem from the varying motives in administrative processes. For example, farmers may underreport areas in subsidy programs to guard against the consequences of inadvertent overreporting (Carfagna and Carfagna, 2010). Measurement errors in identifying variables may 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 make 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 overcoverage and undercoverage. Carfagna and Carfagna (2010) describe studies conducted to examine coverage problems in 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 AF, 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.

Some quality control procedures for measurements and coverage errors are, for example, the following:

- 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 and Carfagna, 2010).
- ESSnet-ISAD (2008b) presents 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.
- Bakker (2012) develops model-based approaches to quantify the bias resulting from measurement error in administrative data. Berka e*t al*. (2012) examine the effectiveness of the Dempster-Shafer theory to quantify the uncertainty in each datum in each of the several registers used in the Austrian census.

3.1.2 Improving ARSA data collection, storage and dissemination

Employing best practices in the collection of administrative data can reduce the number of quality issues to be addressed at the estimation stage. In TAD, data collection is usually performed through well-established procedures and standardized tools. However, in ARSA, various methods and equipment are used in the collection, storage and dissemination processes, and may need to be improved to enhance data quality.

Improving 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 impose 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 (GSARS, 2016a), as it became apparent that even the questionnaire used in that 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. Another way to improve the quality of the data reported is to inform respondents about the value and intended use of the data; this should be covered in the instruction manual.

Use of new technologies in data collection

The use of certain technologies in administrative data collection can contribute to increased data quality, especially in terms of timeliness. 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 as compared to farmer declarations (thus affecting the ADSAS's sustainability).

Another important technology that is increasingly used in data collection is the Computer-Assisted Personal Interviewing (CAPI) software through mobile phones or tablets. This modern but cheap and readily available technology can improve timeliness in data collection. In the field, data entry would be performed by data collectors (such as village extension workers).

If electronic devices with CAPI are used, the data is captured directly in the database, which is accessible at district, regional or provincial, and national levels. Rights of access should be provided such that supervisors are capable of checking and verifying data, and then forwarding them to the district officials who in turn verify and clear them for access by the higher levels. As noted during the pilot project conducted in Côte d'Ivoire, enumerators should be provided with some paper questionnaires to be used in case of device failure (GSARS, 2016a).

The CAPI 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 for district planning purposes.

During the pilot project, it was also demonstrated that providing farmers with a Crop Card can improve 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, it was recommended to limit the use of Crop Cards to continuously harvested crops.

Improving 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 whenever Internet connection is available, data and reports should be considered officially submitted only after they have been approved by the supervisors. If there is no Internet connection, data is temporarily stored on the tablet and sent to the server as soon as the connection is restored.

Improving data storage and dissemination or diffusion

This covers the Information Communication Technologies (ICTs) and media used in data storage and dissemination or diffusion: handheld equipment, telephones, traditional media (such as radio, television and fax); modern ICTs (for example, 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 is the Integrated Metadata Base (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 (for example, 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 technology to store or disseminate information. Some technologies may be fast but expensive, such as iPad tablets, which are associated with high fixed costs. In addition, some technologies may not be feasible if electricity and security are issues.

3.2 ADDRESSING HUMAN AND FINANCIAL RESOURCES ISSUES

3.2.1 Human resources

ARSAs

An assessment of the human resource needs should be made to determine whether there are staffing gaps at any level of the routine reporting system, from the national to the local government level. The finances required for the recruitment and training of staff should be determined and incorporated into the national budget and budget of the administrative institution (parastatal and local governments). Recruitment of the required staff should be done based on the identified staff gaps.

A hierarchical training protocol should be designed to be used at all levels. At national level, officers should be trained. These should train the regional officers who, in turn, train district extension workers. The training guidelines should have details about how to use the data capture tools at the grassroots level, and the data compilation format at district, regional and national levels. There is also a need for clear manuals on the data collection, capture, analysis and dissemination processes. There should also be systematic supervision by qualified staff, including, if possible, staff from the NSO.

EXAMPLE OF GOOD TRAINING PRACTICE: THE ARDS OF THE UNITED REPUBLIC OF TANZANIA

- A training guide is provided for district officers on data consolidation, analysis and feedback in ARDS, which provides guidelines for data handling and analysis at district level.
- Regional officials and district officers are trained on the common reporting formats.
- The district officers in turn train the village/ward agricultural extension officers on the village/ward data collection format.
- Training on Excel and on the Local Governments Monitoring Data Base 2 (LGMDB2) for data management:
  - Regional officials and IT specialists from several neighbouring regions are brought together and trained in same venue on Excel and LGMDB2.
  - The regional officers then train the district officers on Excel and LGMDB2.

Traditional administrative data system

For this category of administrative data, knowledge transfer can be performed. The professional statisticians or independent statistical experts within the NSO could examine the tools and processes in place to collect the information and provide training or recommendations on how to improve them. Exchange of staff is another practice that can be implemented. For instance, Statistics Canada and the Revenue Canada Agency pursued it in an effort to improve the quality of the tax data received.

3.2.2 Financial resources

Institutions in charge of administrative agricultural data production need specific financial support for the costs of data collection. In developing countries, this is generally not straightforward, because of budget constraints. However, it is an issue that can significantly affect the quality of the data and is particularly crucial for ARSAs. Better coordination among the institutions collecting agricultural statistics could minimize the costs of collection, compilation and management processes. This could further be enhanced by closer collaboration with data users. For the specific case of administrative reporting systems, the following recommendations may be helpful:

- Advocacy for best practice – Lobbying governments in some countries in Asia and the Pacific has proven successful in securing access to national budgets for the purposes of the routine data collection system. Governments should therefore be made aware of the importance and benefits of having a functional routine data collection system. Even local governments (administrative units) should be persuaded to commit funds to ensure the sustainability of the routine data collection system.
- Pooling the resources (human and financial) and harmonizing the administrative data activities of the various agencies may cut down on costs for countries. However, in these operations, the roles and obligations of each institution (especially the ministry responsible for agriculture and the NSO) should be clearly spelt out in a memorandum of understanding (MOU) or other type of legal framework.

3.3 STANDARDIZATION OF CONCEPTS AMONG ADMINISTRATIVE AGRICULTURAL DATA INSTITUTIONS

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 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.

The process of concept standardization between ADSAS institutions should address the differences between (i) definitions of units (such as agricultural holdings) (ii) definitions of variables (for example, temporal employees) and (iii) coding systems. This section will emphasize the necessity for this process. Through the organization of specific consultations or technical group meetings, these harmonizations can be performed.

3.3.1 Harmonizing 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 function of the administrative agency in question. Consequently, the administrative agency's definition of a unit may differ, for instance, from the definition endorsed by the NSO. 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 wide-ranging choice of possible units (family, agricultural holding, household, parcel of land, point, etc.) and the dependency on the availability of a quality frame of units are particularly significant issues. The differences between the definitions of units may limit the utility of the administrative source, especially for purposes that involve linking micro data.

It is therefore necessary to harmonize the definitions of units adopted by the institutions of the ADSAS and the NSO.

3.3.2 Harmonizing unit identifiers

It is important that administrative data sources adopt the same identification approach for units. The censuses and surveys implemented in the country should, to the greatest extent possible, use the same identifiers. That will facilitate linkage of the databases for the various possible uses developed in chapter 5.

3.3.3 Harmonizing 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 on 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 and Wallgren, 2010; Carfagna and 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 occur 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 (UNESC, 2007).

In case of changes in definitions, the definitions of variables should be harmonized and systematic communication with the NSO should be adopted.

3.3.4 Harmonizing 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.

SUMMARY

This chapter explores the various approaches that may be adopted to improve the quality of the administrative data. Methodological tools to improve data quality are provided that cover data collection and quality control, as well as data storage and dissemination. Human and financial resources issues and the standardization of concepts among administrative agricultural data institutions are also discussed. It is recommended to introduce quality control checks at all stages, and to standardize concepts by harmonizing the definitions of units and variables as well as the coding systems adopted by the various administrative data sources. With specific regard to ARSA, data collection instruments should be improved and adequate financial and human resources should be made available.

4

Improving administrative data accessibility

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 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 fully attained.

4.1 LEGAL FRAMEWORK

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 relevant country NSO 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 [...] 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 drawn 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. International standards are also of assistance in terms of providing guidance, and should therefore be referred to wherever possible in discussions with administrative departments.

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 NSOs 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. The international comparability of statistics that have been derived wholly or partly from administrative sources may be improved through 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.

4.2 POLICY FRAMEWORKS

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). This principle from a document approved by the UN General Assembly may be used by NSOs to advocate for greater access to administrative data for statistical purposes.

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.

4.3 ADDRESSING ISSUES RELATED TO 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 office (often the NSO) and the administrative office is necessary to enable access to 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.

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. Consequently, 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 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 costs associated with administrative data and limit their accessibility.

Countries facing issues in accessing some administrative data because of confidentiality restrictions could carefully review their statistics legislation and compare their confidentiality provisions with those established in the administrative data producers' regulatory law. If the provisions of the Statistics Act are identical or stronger in terms of confidentiality, the NSO could advocate that obtaining access to the administrative data does not represent a risk of breach of confidentiality.

4.4 AGREEMENTS BETWEEN INSTITUTIONS

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).

4.4.1 Overview and benefits of a MOU

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:

i. **vision and support by agency leadership**: cross-agency data-sharing projects can require significant involvement by agency leadership. However, statistical uses of the agency's data may be considered of secondary importance by the leadership.

ii. **narrow but flexible goals**: it can be important for the MOU to specify goals narrowly, sometimes down to the level of which particular fields in a database will be shared between agencies and how those fields will be used. Narrow goals are also helpful because they facilitate cross-agency discussion of data stewardship issues. Although it is beneficial for goals to be narrow enough to discuss fruitfully in cross-agency discussions, goals must also be flexible – that is, to change as a result of those discussions.

iii. **infrastructure**: this element of success has two components: staffing, and policies and procedures. Cross-agency projects benefit from people who are results-oriented, supportive of the project's goals, experienced with the data, and able to work cooperatively with people in their agency and the partner agency. The quantity of staff time required was frequently a concern for the seven data-sharing projects examined in Prell *et al*. (2009). The second component of infrastructure is the importance of having appropriate policies and procedures in place to support data-sharing activities.

iv. **mutual interest**: to reach a successful conclusion, data-sharing arrangements must benefit each partner to the project.

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 confidentiality and privacy rights can be protected while achieving the benefits of linking program data; etc. All of these issues should be addressed in the design of the MOU.

4.4.2 Key features to be included in an MOU

a. **Legal basis.**
Reference should be made to the legislation permitting access to the administrative source for statistical purposes, and to any legislation that imposes restrictions on such access.

b. **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.

c. **Detailed description of data covered.**
This will include information identifying the data set and the variables contained within it.

d. **Frequency of data supply.**
This will specify when and how frequently the administrative organization will supply the required data.

e. **Quality standards.**
These set the parameters for the quality of the data supplied. Examples include the indication of a maximum acceptable 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.

f. **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.

g. **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 that may arise between the statistical and administrative organizations; these may envisage the involvement of senior managers or possibly even relevant ministers.

h. 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.

An example of MOU can be found in Annex 4 to these Guidelines.

SUMMARY

This chapter discusses the ways to improve both legal and policy frameworks in order to facilitate access to administrative data for statistical purposes. Guidance is provided on how to address issues related to confidentiality and public perception, as well as on how to establish relevant agreements between institutions, notably through an MOU. It is recommended to:

- ensure that adequate legal and policy frameworks are in place;
- address issues of confidentiality; and
- establish agreement between institutions, ensuring that key features are taken into account.

5

Uses of administrative data

Administrative data are often collected at a high temporal frequency and in granular geographic detail. Because administrative data, by definition, are collected for non-statistical purposes, the statistical agency incurs relatively low data collection costs when using such data. These characteristics enable administrative data to be used for multiple purposes.

The review conducted by GSARS (2015a, 2015b) found that the uses of 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 of administrative data. 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.

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 the previous chapters. A common operation that is usually performed before the effective use of administrative data is the record linkage, which serves the purpose of integrating data from different administrative sources or between administrative data and census or survey data. Therefore, the chapter begins with a description of the main tools used for record linkage. In the following sections, the uses made of administrative data in forming the statistical product, as well as the concept of direct use, are investigated.

5.1 METHODOLOGICAL AND TECHNICAL TOOLS FOR RECORD LINKAGE AND 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 (that is, a region or the country). Below, examples are reviewed of unit-level and aggregate-level integration methods. More details are available in GSARS (2015b and 2017b).

5.1.1 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 the links between units may help to overcome differences between administrative and statistical units. The statistical households, for example, can be derived on the basis of the 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 be estimated and applied to the full population. An example is the case in which 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.

5.1.2 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. When unique identification numbers are used in the different files, a deterministic linkage may be performed through a simple merging. Because administrative databases can contain different kinds of units, linkages across databases are not necessarily one-to-one, and procedures are needed for many-to-one or one-to-many matches (Wallgren and Wallgren, 2010).

5.1.3 Probabilistic record linkage

Many applications of administrative data to the production of official statistics involve multiple sources of information – multiple administrative files or administrative and survey files. For many uses of administrative data, linking records from at least two files at the level of the individual unit in the population is desirable. Consider, for example, use of administrative data to check for errors in survey data. While a comparison of the marginal distributions of the administrative file to the corresponding marginal distributions from the survey or census may be informative, a comparison of the alternative data sources at the unit level opens greater possibilities. A unit-level linking operation, for example, permits evaluation of records with relatively large differences in the values recorded in the two different sources.

The operation of merging the files at the record level presents many challenges. The identifying variables may differ across data sets. 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. Large data sets may demand substantial computational effort. Probabilistic record linkage is a statistical procedure for determining the probability that two sets of identifying variables represent the same unit in the population.

Fellegi and Sunter (1969) developed one of the most widely used probabilistic record linkage procedures. In the framework they propose, the latent match status of interest is represented as a latent binary variable, $\boldsymbol{\delta}$, that is 1 if a given pair is a match and is 0 otherwise. The observations are vectors of comparison variables, $\boldsymbol{\gamma} = (\boldsymbol{\gamma_1}, \dots, \boldsymbol{\gamma_K})$, where $\boldsymbol{\gamma_j} = 1$ if variable $\boldsymbol{x_j^A}$ from data set A is equal to variable $\boldsymbol{x_j^B}$ from data set B and is zero otherwise. The distribution used for inference is the conditional distribution of $\boldsymbol{\delta}$ given the observed vectors of comparison variables.

Many extensions to the above procedure have been developed. For example, Larsen and Rubin (2001) develop a procedure that involves alternating between inference based on conditional distributions and manual review. Incorporating clerical review reduces the degree of uncertainty. When data sets are linked through probabilistic linkage models or from incomplete linkage, the estimation procedures need to account for linkage error. Methods such as those proposed by Kim and Chambers (2012) can be employed, which extend they regression analysis techniques of Chambers (2009) to applications with more than two linked data sets. 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.

i. Steps involved in data cleaning and record linkage

The processes of preparing a data set for record linkage and performing the linkage algorithm involve several related components. One involves the selection of variables to use for matching. Then, given a set of matching variables, an operation is often needed to convert representations in different data sets to a standard form. Because comparison of all pairs of records in two files is often computationally prohibitive, methods are needed to reduce the dimension of the comparison space. In comparing two vectors of identifying variables, strict equality is often an excessively restrictive metric. Record linkage algorithms therefore permit the use of different comparison metrics that determine the extent to which two vectors of matching variables agree. Furthermore, the general Fellegi-Sunter probabilistic record linkage paradigm contains several options that a user can modify to suit his or her particular needs. These options involve the method of estimating matching probabilities and the decision rule that determines which records to review manually. ESSnet (2008c) and Day (1994) discuss methodology for multiple aspects of record linkage.

The selection of matching variables often involves both manual and automated steps. The analysts often have prior knowledge about logical matching variables. For example, the social security number and the personal identification number are two useful variables for matching files of individuals in the United States of America and the Netherlands, respectively. Automated algorithms, often termed profiling procedures, have also been developed to aid in the variable selection process. These automated procedures measure both the correlation between variables across data sets as well as the quality of the potential matching variables within a data set. Profiling procedures are particularly helpful in applications with large numbers of potential matching variables with varying degrees of reliability.

Before performing record linkage, the formats of the matching variables need to be standardized. As a simple example, two ways to store the date, "December 8, 1952" are "12-8-1952" and "12/8/1952." A standardization algorithm would convert these two representations of the same date into a format that would allow a matching algorithm to recognize the two representations as equivalent. In practice, standardization algorithms need to operate on more complex character strings that represent attributes such as telephone numbers, addresses, names of people, names of businesses, and names of farming operations. Character strings that represent unique entities may differ due to subtle differences in capitalization, spelling and punctuation, for example. One way to standardize character strings is through a phonetic coding scheme. Such algorithms convert strings that "sound the same" into a unique character format. Soundex and the New York State Identification and Intelligence System (NYSIIS) are two widely used phonetic coding algorithms.

Comparing all pairs of records in two files is often computationally prohibitive. Two methods for reducing the number of comparisons are called "blocking" and "sorted neighborhood". In "blocking," the files are divided into subgroups called blocks and pairs of records are only compared within each block. To illustrate, suppose that two files, each with 5 000 records, are split into 10 blocks of 500 records. Then, the number of comparisons is reduced from 5 000 x 5 000 to 10(500 x 500). For this example, the number of comparisons required for the nonblocked structure is ten times the number required for the blocked structure. In the sorted neighbourhood dimension reduction procedure, records are only compared if they fall in a window that traverses the sorted records.

Because exact matching is often too restrictive, different metrics have been developed to measure the extent to which two vectors of matching variables differ. In the record linkage literature, different metrics are often referred to as "comparison functions". Alternatives to strict equality include the Levenshtein metric for comparing two strings, and the Jaro-Winkler metric, which is specific to comparing names.

In probabilistic record linkage, the user may choose between different methods of parameter estimation and decision rules. The match probabilities, for example, may be estimated by maximizing a likelihood, which is often based on an assumption of conditional independence, and implemented with the expectation-maximization (EM) algorithm. Alternatives to maximum likelihood include frequency-based matching or algorithmic procedures that evaluate patterns of agreement and disagreement. Record linkage algorithms often result in decisions about which records to review manually. Fellegi and Sunter (1969) prove that a particular decision rule minimizes the number of pairs to review for a given error rate. An alternative to the Fellegi-Sunter (1969) decision procedure is a threshold-based rule, according to which any pair with a probability within a specified range is reviewed.

ii. **Software for micro-integration and record linkage**

The challenges associated with combining disparate data sources are not only conceptual; they are also computational. Record linkage often involves managing large quantities of data, and algorithms for data cleaning and standardization are needed. A wide variety of software packages have been developed to perform operations associated with cleaning data and combining multiple data sources. This section first reviews the technical capabilities of existing software tools for performing the steps involved in record linkage discussed above. Second, the software tools are compared along dimensions not strictly related to technical capacity, such as cost, extendibility and transparency.

The discussion below is primarily based on ESSnet (2008c); however, it also contains ideas from Day (1995), Sariyar and Borg (2010), and da Silva *et al*. (2011). ESSnet (2008c) reviews several record linkage software packages from the standpoint of producing official statistics of business data. Day (1995) reviews record linkage software with the specific objective of determining the most appropriate tool for the USDA/NASS. Day (1995) contains an extensive and useful list of questions and criteria for an analyst to consider when selecting the appropriate record linkage tool for his or her needs. The specific computational tools discussed in Day (1995) may be somewhat outdated. However, the suggested criteria and questions to consider remain highly relevant. Da Silva *et al*. (2011) reviews probabilistic record linkage software for the purpose of integrating data from the Brazilian census with data from a post-enumeration survey.

The software packages below are considered: the first nine are reviewed in ESSnet (2008c) while the tenth is an R package discussed in Borg and Sariyar (2010).

1. AutoMatch, developed at the United States Bureau of Census, now under the purview of IBM (Herzog *et al*. 2007, chapter 19).

2. Febrl – Freely Extensible Biomedical Record Linkage, developed at the Australian National University (FEBRL).

3. Generalized Record Linkage System (GRLS), developed at Statistics Canada (Herzog *et al*. 2007, ch. 19).

4. LinkageWiz, commercial software (LINKAGEWIZ).

5. RELAIS, developed at ISTAT (RELAIS).

6. DataFlux, commercialized by SAS (DATAFLUX).

7. The Link King, commercial software (LINKKING).

8. Trillium, commercial software (TRILLIUM).

9. Link Plus, developed at the U.S. Centre for Disease Control and Prevention (CDC), Cancer Division (LINKPLUS).

10. RecordLink, an R package developed by Murat Sariyar and Andreas Borg.

The conclusions and interpretations of software capability should be considered as an indicative guideline, rather than a standard. In deciding which tools to employ for a particular application, an independent comparative evaluation of software may be useful to determine the most appropriate package to meet the needs of the specific application. The overview below aims to provide a useful starting point.

Technical capacity

Table 4 below summarizes the technical capabilities of alternative software packages. These packages are reviewed in more detail in ESSnet (2008c). The column titled "Standardization" indicates the preprocessing and standardization capabilities, if any. The "Profiling" column indicates whether the software has options for automated profiling, and "Space reduction" indicates the blocking methods available. The "Estimation and decision rules" column provides information on the procedure used to estimate matching probabilities (EM algorithm, or other) as well as the type of rule used to decide if a pair of records is classified as a "match," a "non-match," or a "possible match." Because all software packages contain comparison functions, the "Comparison Functions" column indicates the extent of the available comparison functions based on the information provided in ESSnet (2008c).

TABLE 4. SUMMARY OF TECHNICAL CAPABILITIES OF RECORD LINKAGE SOFTWARE PACKAGES.

Package	Standardization	Profiling	Space reduction	Estimation and decision rules	Comparison functions
AutoMatch	NYSIIS, Soundex, other	None	Blocking	Frequency weighting with threshold	Standard
Febrl	Hidden Markov Models & rules-based methods	None	Blocking and sorted neighborhood	Several unsupervised classifiers	Wide variety
GRLS	NYSIIS, Soundex, other	None	Blocking	Agreement/disagreement patterns	Standard
LinkageWiz	NYSIIS, Soundex	None	Not specified	Few details on estimation and decision method	Standard
DataFlux	Tools for business data	Yes	Not specified	Simple, deterministic decision	Wide variety
RELAIS	None	Yes	Blocking and sorted neighborhood	Maximum likelihood estimation, manual review of many-to-many links	Standard
The Link King	None	Yes	Blocking	Ad hoc, iterative procedure for estimation and both probabilistic and deterministic decision rules	Wide variety
Trillium	Extensive	Yes	Not specified	Probabilistic, not Fellegi-Sunter, procedure not specified	Standard
Link Plus	None	None	Blocking	Maximum likelihood, and probabilistic decision	Wide variety
RecordLink	None	None	Blocking	Maximum likelihood with EM algorithm and numerous classifiers for decision rules	Standard

Usability

In addition to the comparison of technical capacity, an understanding of usability is important for choosing the appropriate record linkage software tool. According to ESSnet (2008c), the following indicators of usability are considered:

- Cost. Is the software free or commercial? Does the software require licenses for particular data management or statistical analysis tools?
- Domain specificity. Can the tool handle different languages, or is the software specific to English? Is the tool developed for a specific class of applications or objects, such as business data, human subjects, or health services?
- Transparency. Are the procedures well documented? Can the analyst build an understanding of how the record linkage and data management tools work?
- Extendibility. Can the analyst modify and adapt the procedures to suit his or her specific needs?
- Output Reports. Is the output in a convenient format? Are linked files easy to use and transport to a different system?

Table 5 below, based on ESSnet (2008c), summarizes the usability of the record linkage software tools.

TABLE 5. SUMMARY OF THE USABILITY OF ALTERNATIVE SOFTWARE PACKAGES.

Package	Cost & Requirements	Domain specificity	Transparency	Adoption
AutoMatch	Commercial	English only	Rich documentation	High
Febrl	Free	English only	Source code available	Medium
GRLS	Requires ORACLE	English only	Free, bilingual training course	Medium
LinkageWiz	Commercial but low price	English and French	No precise description	Medium
RELAIS	Open-source, free	No specific domain	Full availability of source code	Low
DataFlux	Requires SAS, but low cost	No specific domain	Documentation available	High
LinkKing	Free	Health and human subjects	Well documented	Medium
Trillium	Commercial	Almost any language or country but specific to marketing applications	Algorithms not precisely defined	Medium
LinkPlus	Free	Cancer registries	No source code but good documentation	High
RecordLink	Free	English and German	Source code and documentation available	Unknown, relatively new R package

Table 6 below, based heavily on ESSnet (2008c), summarizes the strengths and weakness of the alternative software tools.

TABLE 6. PRIMARY STRENGTHS AND WEAKNESSES OF RECORD LINKAGE SOFTWARE PACKAGES.

Package	Strengths	Weaknesses
AutoMatch	User-friendly preprocessing	No error rate estimate, English only
Febrl	Open-source, good preprocessing and tools to compare solutions from different record linkage algorithms	No profiling
GRLS	Good preprocessing and documentation	Only English, requires ORACLE
LinkageWiz	Speed, preprocessing standardization	No profiling or space reduction; black box
RELAIS	Allows user-specified combinations of linkage options; adaptable to a wide range of situations	Low adoption (new); relatively untested
DataFlux	Flexible preprocessing	Deterministic decision
LinkKing	Easy to use	Not flexible, nonstandard estimation
Trillium	User-friendly and language flexibility	Specific to commercial applications, rather than official statistics; limited documentation
LinkPlus	User-friendly and free	No preprocessing, specific to cancer registries, poor handling of non-showing characters in input data
RecordLink	Free, open-source, numerous decision procedures, good documentation	No preprocessing, requires standardized input data in format compatible with package

For use in developing countries, a package with the characteristics of RELAIS may be particularly useful. The software is free and the source code is completely available. The software allows the user to fix any combinations of linkage options, and the software is not specific to a particular subject domain.

5.1.4 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 and Lavallée, 2001).

5.2 USES IN FORMING THE STATISTICAL PRODUCT (INDIRECT USE)

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.

5.2.1 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 and Carfagna, 2010). Examples from both developing and developed countries are provided below:

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 Value-Added Tax (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 and 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 and Whitridge, 2000).

Examples from a developing country: India

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).

5.2.2 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.

For instance, in Statistics Sweden, the necessary information on the population for the purpose of sample design is provided by an administrative data source. 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 and Hall, 2007).

5.2.3 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 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 and 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 of the United States Bureau of Land Management illustrates the use of customs data to define an inequality restriction. This bureau 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 were 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.

5.2.4 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 nonrespondents 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 nonrespondents. If a variable from an administrative database is observed for both respondents and nonrespondents, 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 nonrespondents 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, that is, 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 A be the set of sampled units, indexed by $A = \{1, 2, \cdots, n\}$ and δ_i be a response indicator that takes a value of 1 if unit i responds, and 0 otherwise. Here, the auxiliary data x are obtained from an administrative data and are available for both respondents and nonrespondents. For the sake of brevity, we assume a missing-at-random (MAR) condition on the response mechanism, such that

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

The MAR condition (6.1) states that the responses are independent of the study variable y given the auxiliary variable 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 y given x and $\delta = 1$ is estimated using the survey's respondents, plausible values for the nonrespondents may be generated on the basis of the estimated imputation model given by $\hat{f}(\mathrm{y}|x, \delta = 0)$.

In practice, there are two population approaches when conducting imputation. Fractional imputation (Kim and 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 and Sarndal, 2005), which may be derived from administrative sources.

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.

5.2.5 Measurement error modelling

Agricultural survey 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 multilevel model to obtain a single unified statistic, and an associated measure of uncertainty. In GSARS (2016a), 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 Namibia Statistics Agency (NSA). Two are obtained from the MAWF of Namibia. 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, GSARS (2016a).

5.2.6 Small-area estimation

Many statistical procedures to obtain estimates for small areas or to forecast a future outcome are based on explicit models. In 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 the potential to reduce the mean squared errors of forecasts. In the construction of the 1997 National Resources Inventory (NRI) of the United States of America estimates, administrative data on transportation were used to create small-area estimates of the area of roads (Nusser and Goebel, 1997; Wang and Fuller, 2003).

This section briefly discusses how administrative data can be used as auxiliary information in a classical small-area estimation. See GSARS (2016b) 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 Battese, 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

$$y_k = x'_k \beta + u_k , \qquad (6.3)$$

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

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 $\hat{\gamma}_k = \left(\sigma^2_{e,k} + \hat{\sigma}^2_u\right)^{-1} \hat{\sigma}^2_u$, and $\hat{\beta}$ and $\hat{\sigma}^2_u$ 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 into small-area estimation.

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 planted to corn and soybean 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 soybean. 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.

The use of small-area estimation models to obtain subnational estimates of crop area in Namibia was explored in GSARS (2016b). 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. Table 7 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.

TABLE 7. ESTIMATED VARIANCES FOR THE NCA ESTIMATORS.

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)

Ethiopia

In Africa, Ethiopia is one of the countries that has tested the use of small-area estimation. The agricultural annual surveys conducted by the Central Statistical Agency (CSA) provided crop-wise area estimates at regional and zone levels only. Due to the small sample sizes, estimates at the level of districts (werada) were not available. On the other hand, the Ministry of Agriculture and Rural Development (MoARD) did generating area estimates from ARSA data. The small-area estimation approach was used to develop district-level estimates for crop area from annual surveys, using MoARD data as an auxiliary variable (GSARS, 2015c). For further details on the small-area estimation methodology, see GSARS (2015b).

5.2.7 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 being related to PPS 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 included in 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. Below are some practical examples:

- 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 and Hidiroglou, 2007).
- 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).

5.2.8 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 identities 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.

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) in the United States of America, 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.

5.2.9 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. GSARS (2015e) 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 GSARS (2016a and 2016b). Let the three observed data sources be denoted with Y_{i1}, Y_{i2}, and X_i respectively. Assume that the three data sources satisfy the models

$$\begin{aligned} Y_{i1} &= \beta_{01} + \beta_{11}x_i + e_{i1} , \\ Y_{i2} &= \beta_{02} + \beta_{12}x_i + e_{i2} , \\ X_i &= x_i + u_i , \end{aligned} \tag{6.4}$$

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 issues 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.

As part of the pilot project conducted (GSARS, 2016a), 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 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 enables estimation of a measure of uncertainty associated with the predictors. Details are available in GSARS (2016a), while a related example is provided in GSARS (2016b).

5.3 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 8 summarizes the uses of administrative data for statistical purposes made in 13 African countries that participated in a non-probability survey related to the use of ADSAS (GSARS, 2015c). 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 African countries. This section discusses various ways in which administrative data are used directly as the final statistical product.

TABLE 8. ADMINISTRATIVE USES OF ADSAS: USES IN CONSTRUCTING STATISTICS.

Statistical uses	BURUNDI	EGYPT	GHANA	LESOTHO	LIBERIA	LIBYA	MAURITANIA	MAURITIUS	SOUTH SUDAN	SOUTH AFRICA	SUDAN	UGANDA	ZAMBIA	Total/13
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: GSARS (2015c). 1=Yes, 0=No.

5.3.1 Direct tabulation

If administrative data are of sufficiently high quality, they may be used directly for the statistical product (Brackstone, 1987; Wallgren and 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 and 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 and Whitridge, 2000). The USDA/NASS routinely publishes information on the imports and exports of agricultural products (Harris and 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 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 and Wallgren, 2010).

One approach to assess the relevance of the direct tabulation of administrative data is to compare these data with the estimates from ad hoc surveys over a period. In Great Britain and the United Kingdom, such an exercise revealed a percentage difference of only 4 percent in the cattle population recorded through administrative and survey data from 2003 to 2006 (see table 9 below).

TABLE 9. GB AND UK CATTLE POPULATION AT 1 JUNE AS FROM SURVEY AND ADMINISTRATIVE DATA: 2003 TO 2006.

Total cattle								'000
	Great Britain				United Kingdom			
	2003	2004	2005	2006	2003	2004	2005	2006
June Survey	8,823	8,911	8,727	8,635	10,508	10,588	10,392	10,270
Admin data	9,202	9,300	9,154	8,970	10,946	11,070	10,867	10,657
Difference: Survey-Admin	-380	-390	-428	-335	-438	-482	-475	-386
Percentage Difference	-4	-4	-5	-4	-4	-4	-4	-4

Source: Elliott and McDonnell, 2007.

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 of America and the European Union (Clay, 2013). Subsidy programs in Brazil, the Russian Federation, 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 the increasing commercialism of agriculture, new insurance products based on weather indexes and 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 by register-based agricultural statistics in the future. Below are some examples of direct tabulation of administrative data:

- NASS publishes administrative information on hog slaughter (Harris and Clark, 2013) obtained from inspections conducted by federal and state officials. The data from NASS hog and pig inventories should align with published slaughter data.
- 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 and Whitridge, 2000).

CONDITIONS UNDER WHICH DIRECT TABULATIONS OF ADMINISTRATIVE DATA MAY BE APPROPRIATE OR NECESSARY

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.

5.3.2 Crop forecasting

The administrative data collected on the different aspects of weather (meteorological data, remote sensing information, etc.) are used in a number of countries to forecast crop yield or production for the purposes of food security. Examples in developing countries are:

- **Zambia**: routine data on local livestock and crop development trends collected by the extension officers of the Ministry of Agriculture and Cooperatives are used to establish a preliminary forecast.
- **Mali**: the Famine Early Warning Unit (*Système d'Alerte Précoce, or 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 to make recommendations for actions to ameliorate the situation (Kelly and Donovan, 2008).
- **India**: 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-percent random sample of villages and to submit the village crop statements to the higher authorities within a stipulated date. The advance estimates are used in the framing of crop forecasts. This provides the Indian Government with advance estimates of production, which are crucial for various decisions relating to pricing, distribution, export and import.

5.3.3 Overview of direct uses of administrative data in countries

Sample surveys and censuses that use rigorous statistical methods remain the most reliable source of agricultural data. However, agricultural surveys in developing countries are conducted with irregular frequencies because of budget constraints (Pangapanga *et al*., 2013). Administrative records can be used to cover agricultural data gaps if surveys are totally absent or present an undercoverage of key agricultural variables. For instance, in Denmark, if certain information is available in an administrative register, Statistics Denmark does not include them on the census or survey questionnaire (Jensen and Larsen, 2016). Administrative data help the established government's systems for planning purposes and are available on an annual basis (Pangapanga *et al*., 2013).

In many developing countries, a large proportion of the data produced and disseminated through national, regional or global databases or publications, are from a variety of sources, because of the absence of regular statistical surveys or censuses conducted by countries (Keita and Chin, 2013).

Table 10 below shows that information from an 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 various uses, and for measuring progress towards the implementation of international agreements and goals.

TABLE 10. USES OF ADSAS AS 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 on 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: GSARS (2015c). Note: 1=Yes, 0=No.

SUMMARY

In statistics, administrative data can be used in a variety of ways. A first approach is a direct use as agricultural statistics, under certain conditions. In addition, administrative data can be used to form the statistical product, either during sampling design or during data processing and estimation. Therefore, administrative data should be integrated into the agricultural statistics system through the design and improvement of the ADSAS.

6

Integration of administrative data into national agricultural statistics systems

Considering the potential uses of administrative data towards improving the availability and quality of agricultural statistics, these data should be taken into account in the national system or strategy related to the production of agricultural statistics. A short-term goal is to design and improve the ADSAS (as discussed in the previous chapters) to take advantage of the benefits of the administrative data related to the agricultural sector. Then, it will be necessary to fully integrate the use of ADSAS data into the national agricultural statistics plan, which may be for example the Strategic Plans for Agricultural and Rural Statistics (SPARS), developed and promoted by the Global Strategy. A long-term goal is instead the development of a register-based agricultural statistical system through the development or improvement of key registers in the country.

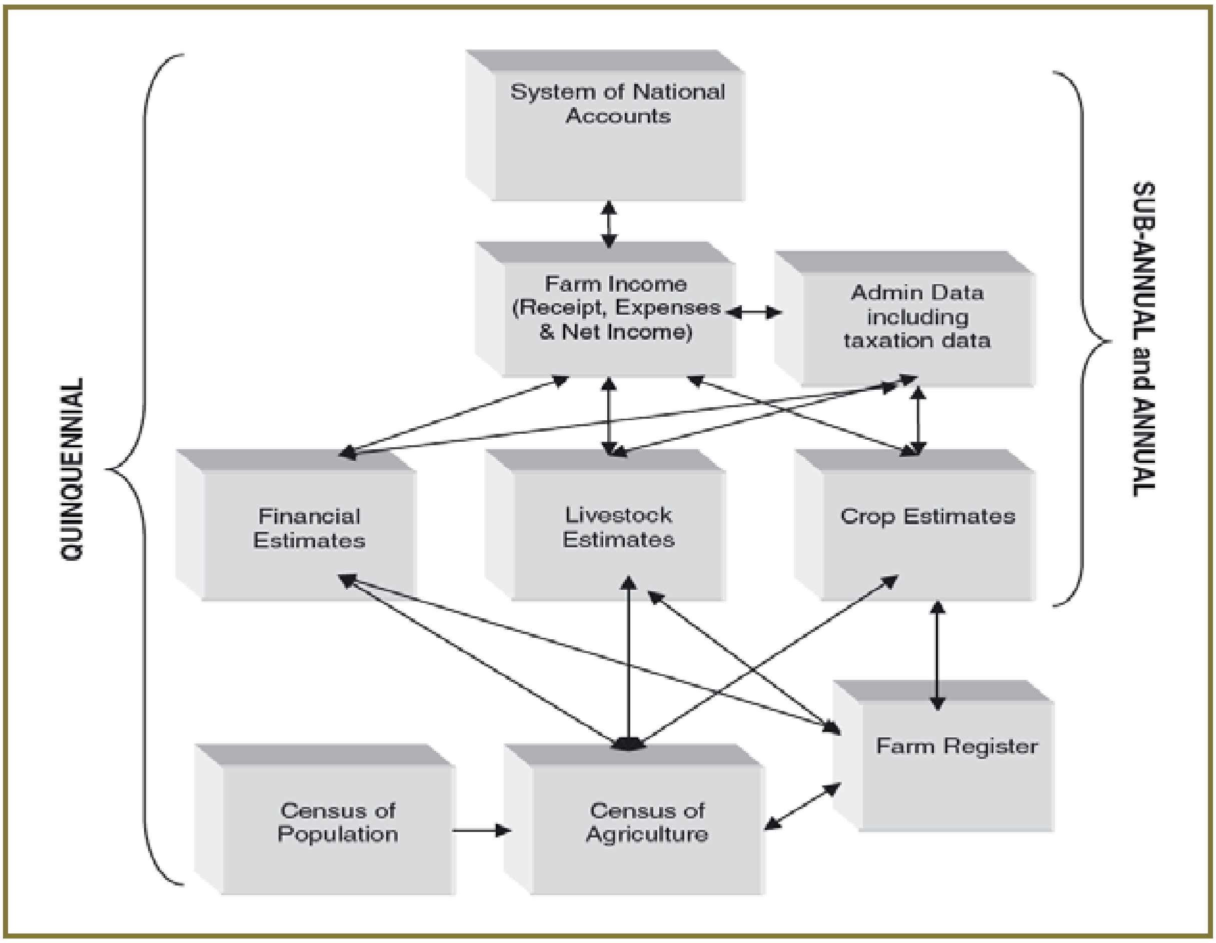
6.1 INTEGRATION OF THE ADSAS INTO THE SPARS

It is crucial for each country to establish a clear strategy regarding the production of agricultural statistics that should take into account the improvement and use of administrative data. Accordingly, the Global Strategy recommends that countries elaborate SPARS taking into account administrative data, and develop a standard methodology to design these plans (GSARS, 2014). SPARS provide a basis for establishing policy strengths and priorities, and the respective data needs, critical gaps, deficiencies, duplications and inconsistencies. They should cover the entire agricultural and rural sector, including all data collection, analysis, dissemination and use from censuses, surveys and administrative systems (GSARS, 2014). The SPARS design methodology includes the assessment of statistical outputs from existing agricultural data sources, including administrative reporting systems.

The methodology recommends that the SPARS be realistic and pragmatic with regard to resources. This implies, for example, prioritization, sequencing and cost-effectiveness, considering alternative ways of compiling data, such as administrative sources and sample surveys. The proper integration of administrative data in the agricultural statistical system will enable important cost savings, as described in GSARS (2015b).

An example of integration of the ADSAS into the national agricultural statistics system is provided in figure 3 below.

FIGURE 4. VISUAL DEPICTION OF STATISTICS CANADA'S AGRICULTURAL STATISTICS FRAMEWORK.



Source: Dion, Chartrand, and Murray (2010).

6.2 LONG-TERM PERSPECTIVE ON THE INTEGRATION OF ADMINISTRATIVE DATA INTO STATISTICAL SYSTEMS

The Global Strategy to Improve Agricultural and Rural Statistics (World Bank, FAO and UN, 2011) discusses the integration of agriculture into the national statistical system to improve agricultural statistics. A long-term perspective on this integration may be a register-based agricultural statistics system. Wallgren and Wallgren (2017) describe how statistics regarding agriculture and the rural population are produced in a country where statistics production is based on registers, and discuss the modernization of the national statistical system from a traditional census-based system into a register-based system. Here, a long-term strategy which closely follows the guidelines proposed by Wallgren and Wallgren (2017) is proposed to construct an integrated system of statistical registers utilizing administrative data sources. It is sought to provide a plan to create an integrated statistical farm register that can be regularly updated with multiple administrative data sources.

6.2.1 Modernization of the integrated register-based system

i. 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 connecting 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.

ii. Improving the administrative system

Administrative data should be of the same quality as survey data. Key variables such as identity numbers, registration of births and deaths, and migration are also important in linking data across sources.

iii. 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.

iv. Centralization, cooperation and legislation

Another key condition is the construction of a centralized statistical system. Wallgren and Wallgren (2007) 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 national statistical institutes and other organizations is necessary for administrative systems to operate effectively.

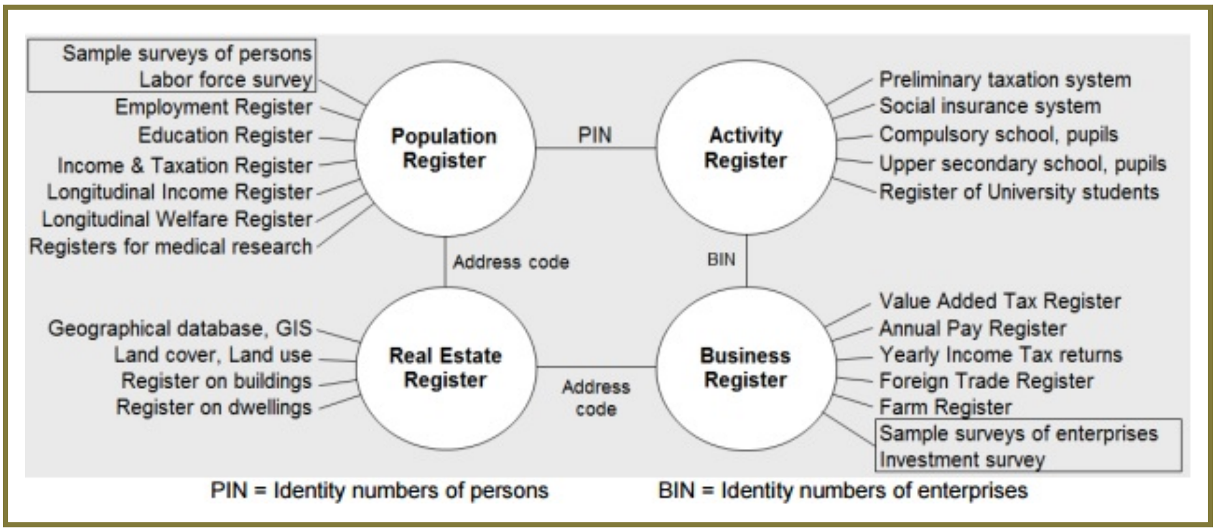
v. Quality assessment

As discussed in chapter 5, administrative data are frequently affected by 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).

6.2.2 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 may be built. Figure 4 illustrates a general statistical production system, with base registers, other statistical registers and sample surveys. The term "integrated system" is used to emphasize that the populations and variables are consistent and the estimates are coherent.

FIGURE 5. A REGISTER-BASED PRODUCTION SYSTEM, THE REGISTER SYSTEM, AND SAMPLE SURVEYS.


Sample surveys of persons
Labor force survey
Employment Register
Education Register
Income & Taxation Register
Longitudinal Income Register
Longitudinal Welfare Register
Registers for medical research
Population Register
PIN
Activity Register
Preliminary taxation system
Social insurance system
Compulsory school, pupils
Upper secondary school, pupils
Register of University students
Address code
BIN
Geographical database, GIS
Land cover, Land use
Register on buildings
Register on dwellings
Real Estate Register
Address code
Business Register
Value Added Tax Register
Annual Pay Register
Yearly Income Tax returns
Foreign Trade Register
Farm Register
Sample surveys of enterprises
Investment survey
PIN = Identity numbers of persons
BIN = Identity numbers of enterprises

Source: Wallgren and Wallgren (2007).

The creation of new statistical systems depends on the construction of a registration system. Details of the system procedures proposed by Wallgren and Wallgren (2007) 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 cadastre, 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 (2007) 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 that certain administrative data, such as expert judgments and eye estimates, are not handled in an integrated register-based system. However, if financial resources and institutional 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 for those developing countries without land registers.

SUMMARY

This final chapter discusses the importance of the integration of administrative data in agricultural statistical systems. It is recommended to take into account the improvement and use of these data in strategic plans for the production of agricultural statistics. It is highlighted that the SPARS developed and recommended by the Global Strategy should take into account the improvement and use of administrative data, as discussed in detail in the previous chapters of this publication. Finally, a strategy and methods to develop a register-based agricultural statistical system are discussed as a long-term perspective.

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. Last accessed 12 April 2017.

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

Bakker, F.M.B. 2012. Estimating the validity of administrative variables. *Statistica Neerlandica*, 66: 8–17.

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. Last accessed 12 April 2017.

Battese, G.E., Harter, R.M. & Fuller, W.A. 1988. An error-components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83: 28–36.

Beckler, D.G. 2013. *Administrative Data Used by the National Agricultural Statistics Service*. Proceedings of the *Sixth International Conference on Agricultural Statistics*, 23–25 October 2013. Rio de Janeiro, Brazil.

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: Trento, Italy.

Berg, S., Hall, J., Färnstrand, J. & Norlander, M. 2007. *Administrative data in the Swedish SBS*. Statistics Sweden Publication: Stockholm.

Berka, C., Humer, S., Moser, M., Lenk, M., Rechta, H. & Schwerer, E. 2012. Combination of evidence from multiple administrative data sources: quality assessment of the Austrian register-based Census 2011. *Statistica Neerlandica*, 66: 18–33.

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.

Bins, B.O. & Dale, P.F. 1995. *Cadastral surveys and records of rights in land*. FAO Land Tenure Studies. FAO Publication: Rome.

Borg, A. & Seriyar, M. 2010. The Record Linkage Package: Detecting Errors in Data. *The R Journal*, 2: 61–67.

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. Last accessed 12 April 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. Last accessed 12 April 2017.

Da Silva, D.A., Romeo, S.M.O., Soares, T.S. & Xavier, V.L. 2011. Study of Record Linkage Software for the 2010 Brazilian Post Enumeration Survey. Proceedings of the 58th World Statistical Congress, Dublin, pp. 1056–1063.

Day, C. 1995. Record Linkage 1: Evaluation of Commercially Available Record Linkage Software for Use in NASS. United States Department of Agriculture, National Agricultural Statistics Service, Research Division. USDA Publication: Washington, D.C.

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., Bee, Espa, R. & Piersimoni, F. (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*. Publication of the 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.

Fellegi, I.P. & Sunter, A.B. 1969. A theory for record linkage. *Journal of the American Statistical Association*, 64(328): 1183–1210.

Fuller, W.A. 2009. *Sampling Statistics*. John Wiley and Sons: New Jersey, USA.

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.

Geuzinge, L., van R.ooijen, J. & Bakker, B.F.M. 2000. *Integrating administrative registers and household surveys*. Statistics Netherlands Publication: Voorburg/Heerlen.

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

Goel, B. 2002. Agricultural Surveys in India – Some Thoughts. *Journal of the Indian Society of Agricultural Statistics*, 55: 32–46.

Global Strategy to improve Rural and Agricultural Statistics (GSARS). 2014. *SPARS: Strategic Plan for Agriculture and Rural Statistics*. GSARS Technical Report: Rome.

_____ 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.

_____ 2017. *Improving the methodology for using administrative data in an agricultural statistics system. Final Report*. Technical Report n.24. Global Strategy 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. Last accessed on 12 April 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*. Available at: http://cenex-isad.istat.it/. Last accessed on 12 April 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.

Harris, J.M. & Clark, C.Z.F. 2013. *Strengthening Methodological Architecture with Multiple Frames and Data Sources*. Proceedings of 59th ISI World Statistics Congress, 25–30 August 2013, Hong Kong SAR.

Herzog T.N., Scheuren F.J. & Winkler, W.E. 2007. *Data Quality and Record Linkage Techniques*. Springer Science+Business Media: New York, USA.

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. Last accessed on 12 April 2017.

__________2008b. *Report of WP2: Recommendations on the Use of Methodologies for the Integration of Surveys and Administrative Data*. Available at: http://cenex-isad.istat.it/. Last accessed on 12 April 2017.

__________2008c. *Report of WP3: Software Tools for Integration Methodologies*. Available at: http://cenex-isad.istat.it/. Last accessed on 12 April 2017.

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

Jensen, P.V. & Larsen, K. 2016. *Register based statistics – Agriculture*. Component C Infrastructure for Agricultural Statistics. Activity C.1.a Data sources for agricultural statistics and farms frame. Available at: http://www.dst.dk/ext/6046544347/0/israel2016/Annex-C1a-10-Register-based-statistics-Agriculture--pdf. Last accessed on 7 October 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 ofAdministrative Data Research and Development*. Statistics Sweden Publication: Stockholm.

Larsen, M.D. & Rubin, D.B. 2001. Iterative automated record linkage using mixture models. *Journal of the American Statistical Association*, 96: 32–41.

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. Last accessed on 12 April 2017. National Statistical Office Publication: Lilongwe.

Maligalig, D.S. 2017. *Administrative Reporting Systems for Agriculture in Asia*. Paper presented at the 61st ISI World Statistics Congress, 16–21 July 2017, Marrakech, Morocco.

Namibia Statistics Agency. 2014. *Press Release Update on the Census of Agriculture*. Available at: http://cms.my.na/assets/documents/p19dmo8lrf15ok8t71u3t1j7k7q1.pdf.

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. Last accessed on 12 April 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.

Omernik, J.M. 1987. Ecoregions of the conterminous United States. Map (scale 1:7,500,000). *Annals of the Association of American Geographers*, 77: 118–125.

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.

Roberts, R.A.J. 2005. *Insurance crops in developing countries*. FAO Agricultural Services Bulletin. Available at: ftp://ftp.fao.org/docrep/fao/008/y5996e/y5996e00.pdf.

Sariyar, M. & Borg, A. (2010). The Record Linkage Package: Detecting Errors in Data. *The R Journal*, 2: 61–67

Sarndal, C.E., Swensson, B. & Wretman, J. 2005. *Model Assisted Survey Sampling*. Springer-Verlag: New York, USA.

Sarndal, C.E. & Lundstrom, S. 2005. *Estimation in Surveys with Nonresponse*. John Wiley & Sons: Chichester, England, UK.

Seljak, R. 2007. Use of the tax data for the purposes of the short-term statistics. Seminar on Registers in Statistics – methodology and quality, Helsinki.

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. Last accessed on 12 April 2017.

United Republic of Tanzania. 2011. *Monitoring and Evaluation Framework; Revised Final Draft, March 2011*. 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*. Available at: www.resakss.org/2014conference/docs/Tanzania_JSR_Assessment.pdf. Last accessed on 12 April 2017.

Thomsen, I. & Holmøy, A.M.K. 1998. Combining data from surveys and administrative record systems. The Norwegian experience. *International Statistical Review*, 66, 201–221.

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

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. 12 April 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. Last accessed on 12 April 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 Agriculture (USDA). 2011. *Price Program: History, Methodology, Analysis, Estimates, and Dissemination*. USDA Publication: Washington, D.C.

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.

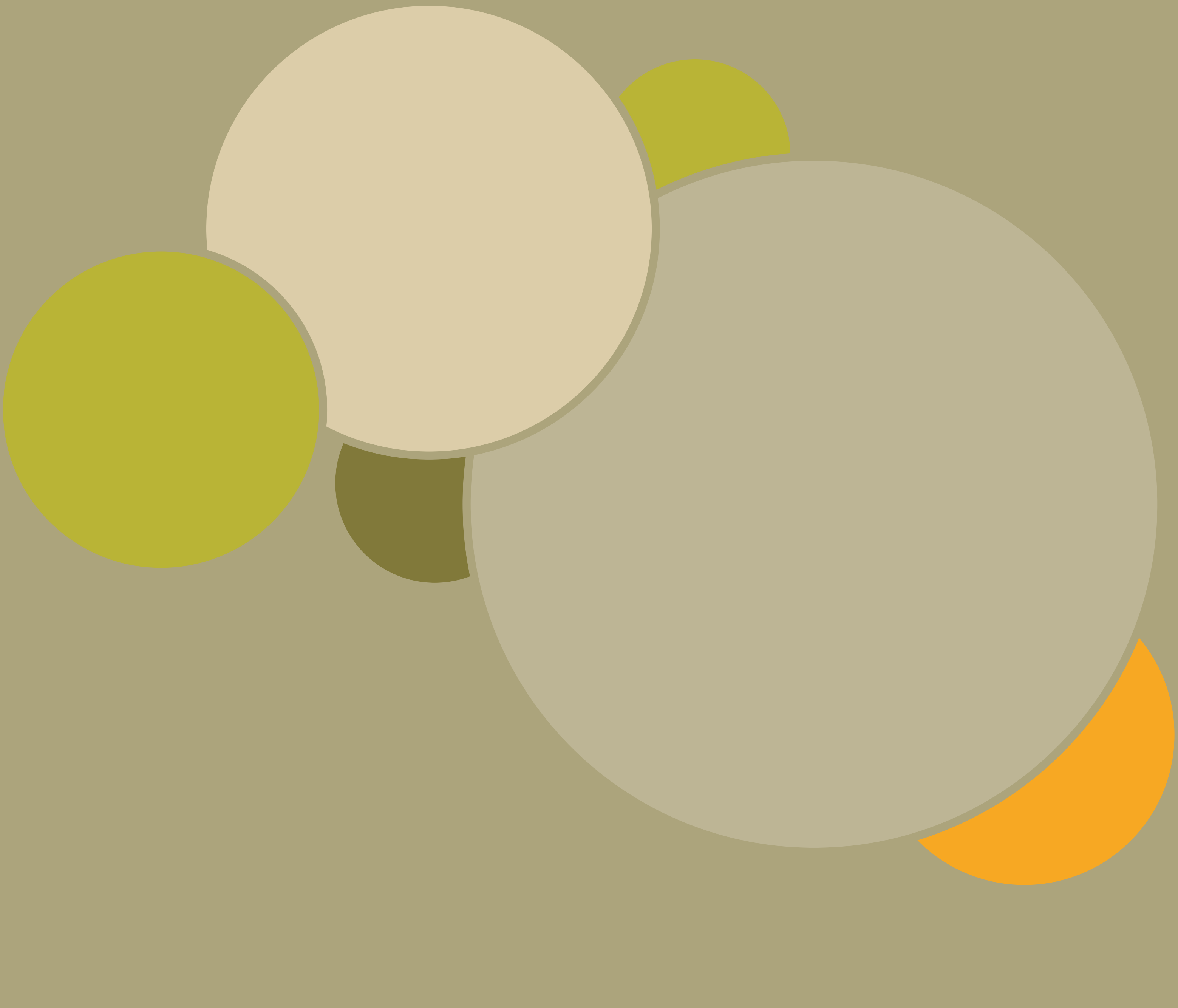
__________ 2011. *To understand the possibilities of administrative data, you must change your statistical paradigm*. In Proceedings of the American Statistical Association Section on Survey Research Methods, pp. 357–365. American Statistical Association Publication: Alexandria, VA, USA.

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*. Report 56719-GL. World Bank Publication: Washington, D.C.

Yung, W., Rancourt, E. & Hidiroglou, M. 2007. Administrative Data in Statistics Canada's business Surveys: The Present and the Future. Seminar on Registers in Statistics – methodology and quality, 21–23 May 2007, Helsinki.

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



Annex

ANNEX 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.*

    a. *The administrative data is generated at village level and consolidated at subdistrict level to facilitate planning at that level.*
    b. *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	Subdistrict 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/ smartphones	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 airtime/data	Tablets and computers	500	12	Months	14.93	89 552.24
6	Procure programming services	System design and support	10	2		74.63	1 492.54
	Subtotal						**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 subdistrict 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	Communication 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 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**

ANNEX 2
ELEMENTS, ATTRIBUTES AND INDICATORS OF QUALITY OF ADMINISTRATIVE REGISTERS.

Quality elements	Quality attributes	Quality indicators
Administrative data source	1. Relevance	1.1. Utility
		1.2. Intended use
		1.3. Demand for information
		1.4. Satisfaction of primary users
	2. Information security and limitations on the use of the information	2.1. Legal framework
		2.2. Personal data protection
		2.3. Limitations due to confidentiality regulations
		2.4. Confidentiality agreements
		2.5. Secure data transfer
		2.6. Confidentiality, integrity and availability of information
		2.7. Data protection
		2.8. Data backup policies
	3. Data delivery commitment	3.1. Costs associated with the delivery
		3.2. Delivery agreements
		3.3. Frequency of deliveries
		3.4. Dates of last five deliveries
		3.5. Punctuality
		3.6. Risks because of lack of data
		3.7. Alternative method to replace the lack of information
		3.8. Means of data delivery
		3.9. File format
		3.10. Data selection
	4. Control and continuous improvement	4.1. Data collection
		4.2. Consistency control
		4.3. Change control
		4.4. Continuous improvement
	5. Data treatment	5.1. Control of objective units
		5.2. Control of variable content
		5.3. Control of outliers
		5.4. Changes
		5.5. Reasons for not changing
		5.6. Changes according to procedure
		5.7. Use of Database Management System
		5.8. Database documentation
		5.9. Database integrity

Quality elements	Quality attributes	Quality indicators
Metadata	1. Metadata documentation	1.1. Metadata documentation
	2. Completeness and clarity	2.1. Definition of population units
		2.2. Description of variables
		2.3. Communication of changes in definitions/concepts
	3. Use of unique keys	3.1. Identification keys
		3.2. Comparability of identification keys
		3.3. Unique combinations of variables
	4. Comparability	4.1. Comparability of the objective unit definition
		4.2. Comparability of variable definitions
Data	1. Technical controls	1.1. Readable data
		1.2. Redefinition of concepts and metadata in case of more than one data source
		1.3. Correspondence between data and metadata
		1.4. Record linkage method
		1.5. Verification of effectiveness of the record linkage method
	2. Coverage	2.1. Overcoverage
		2.2. Classification errors
	3. Record linkage	3.1. Rate of record linkage
	4. Completeness	4.1. Rate of unit non response
		4.2. Rate of item non response
	5. Measurement	5.1. External control (audit)
	6. Identification keys	6.1. Rate of records with unique key
	7. Data processing	7.1. Data editing
		7.2. Imputation
	8. Data accuracy	8.1 Data accuracy
	9. Coding	9.1. Use of standard coding
		9.2. Verification of coding
		9.3. Rate of coding errors
		9.4. Rate of records without code
	10. Data freshness	10.1. More than 90 percent of the objective units created during year t have been registered before the end of the year t+1
	11. Multiple records	11.1. Rate of multiple records of the same unit
	12. Other controls	12.1. Rate of units with valid values into identification keys
		12.2. Tables of the statistical operation have been validated through automatic procedures

Quality elements	Quality attributes	Quality indicators
Statistical product	1. Comparability	1.1. Length of comparable time series
		1.2. Comparability of microdata over time
	2. Relevance	2.1. Identification of users
		2.2. Information about users
		2.3. Rate of final user satisfaction
		2.4. Utility (intended uses)
	3. Coherence	3.1. Coherence of statistics with different periodicity
		3.2. Coherence of statistics with the same socioeconomic scope
	4. Availability and clarity	4.1. Accessibility by Internet
		4.2. Rate of completeness of metadata
	5. Accuracy	5.1. Coefficient of variation
		5.2. Rate of unit nonresponse
		5.3. Rate of item nonresponse
		5.4. Rate of imputation
		5.5. Rate of editing
		5.6. Rate of overcoverage
		5.7. Rate of classification errors
	6. Timeliness and punctuality	6.1. Punctuality of the statistical product dissemination
		6.2. Length of time between its availability and the event or phenomenon it describes
		6.3. Duration between reference time point of administrative data and date of availability to the statistical office

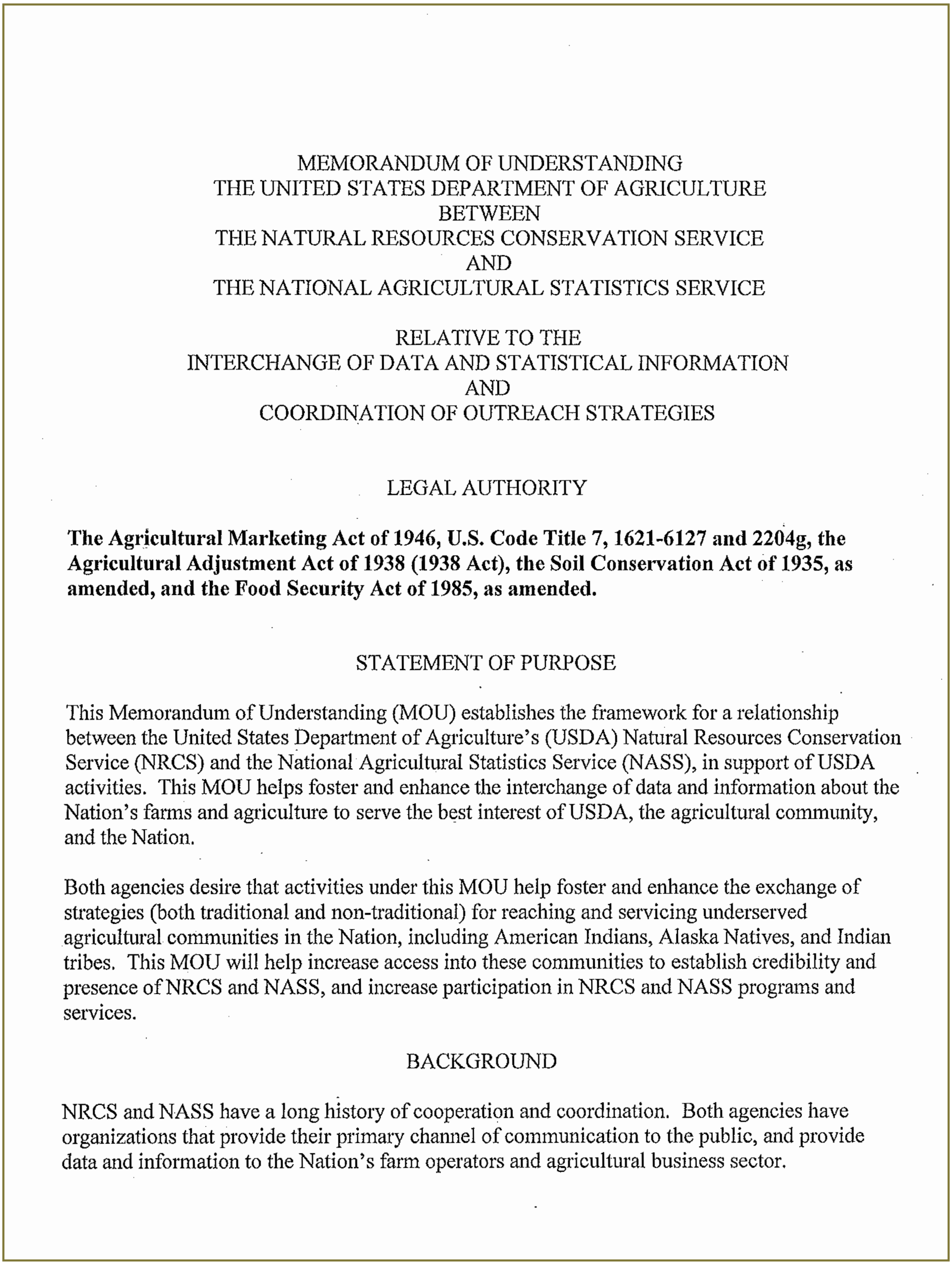
ANNEX 3
SUMMARY OF AGRICULTURAL ADMINISTRATIVE DATA GAPS AND WAYS OF IMPROVEMENT.

Section/Topic	Gap/weakness	Possible solution(s)
Institutional framework	Poor co-ordination between the NSO and the various agencies involved in administrative agricultural data collection and management	**India** provides a **good example on how to co-ordinate** the various federal and state institutions
	Weak data relevance due to lack of communication between data producers and users	Forums for communication with stakeholders such as the National Agricultural Statistics Technical Committee Strengthen Producer–User committees and hold regular dissemination workshops.
	Inadequate physical and statistical infrastructure	Technical support to improve institutional infrastructure Learn from India, which has a detailed cadastral survey maps, frequently updated land records and the institution of a **permanent village reporting agency**
	Lack of, or poor, training and supervision. Quality indicators reflect poor training.	Provide proper guidelines, for example, hierarchical training systems and the **Supervision and backstopping Best Practices** inform the United Republic of Tanzania ARDS. Also, refer to the special training and strict quality control procedures in India
Resources	Many MDAs in the agricultural sector do not have a specific statistics section or unit	Create statistics units
	Inadequate number of qualified staff and low staff retention mainly due to low salaries and poor working conditions	Assess human resource needs and incorporate finances required for the recruitment and training of staff into the national budget and budget of the Ministry, parastatal agencies and local governments
	Poor incentive structures among employees leading to poor behaviour, such as shirking, among employees of institutions.	Set up an efficient compensation and incentive plan to motivate employees based on their actual performance
	Extension staff and/or chiefs who often collect administrative agricultural data also have many other functions	Make needs explicit in budget and pool resources (reduce overlap, streamline activities to make data collection clear in job description) to cut costs
Financial resources; Sources of funding and sustainability strategies	Inadequate and unsustainable financial resources partly due to provision of very low funding for agriculture in the national budgets	Greater budgetary allocations in the national budgets and mainstreaming the costs of administrative agricultural data
	Much of the funding of activities tends to be from donor agencies having a short lifespan. Once donor funding is terminated, the systems usually collapse.	Same as above, plus **Advocacy best practices** – Lobbying government in **Asia Pacific**, has done very well in gaining access to the national budget for routine data collection systems.
	Resource wastage, given the duplication of data collection activities by various agencies	Better co-ordination between MDAs, **Pooling resources** (human and financial) and harmonizing administrative data activities will cut down on costs.
	Lack of information on cost effectiveness of Agricultural Routine Data Systems	Determine costs of data collection, analysis, dissemination, database management.

Section/Topic	Gap/weakness	Possible solution(s)
Data collection and management	Poor data collection tools – questionnaires and manuals	Best practices from India and the United Republic of Tanzania of clear questionnaires and manuals with clear guidelines on who/ how/when to fill the various forms
	Subjective reporting of crop area, production, and realized and forecasted yields	Objective area measurement and production estimation Develop land registration systems such as that in India, and programs that encourage farmers to maintain records
	Estimating production for food and minor crops	Look for new potential data sources and special studies and/or surveys
	Gaps in the methods of making estimates under mixed and continuous cropping	Expect lessons from the Global Strategy study on *"Methodologies for Estimation of Crop Area and Crop Yield under Mixed and Continuous Cropping"* conducted by the Indian Agricultural Statistics Research Institute (IASRI)
Statistical software and ICT	Application of statistical standards and methods and use of modern technology	Introduce and enhance use of modern technologies such as GPS equipment, mobile phone, PDAs, scanners, remote sensing, etc., adapted for agricultural administrative data processes. Train and equip staff to be able to use these technologies for data capture, processing and analysis, compilation, storage and dissemination.
Availability of core data items	Often incomplete, out-of-date, inconsistent and unreliable routine data on livestock and use of non-uniform formats across local administrations	Improve skills in data-handling and processing, insufficient resources, etc. Harmonize formats in data collection and analysis. There are some good practices in the United Republic of Tanzania, Uganda and Mozambique.
	Data on other core data items, such as forestry, fisheries, agricultural inputs are also incomplete.	Look for new potential sources, e.g. farmers' associations and inputs producers.
	Crops: minor crops and continuous cropping systems pose challenges	Special studies/surveys for minor crops and results from IASRI study on continuous cropping
	Trade data/exports and imports: Customs office provides official trade statistics, but informal trade is undocumented	Carry out surveys on informal cross-border trade surveys.
	Land cover: definitions of classifications inadequate	Improved coordination to define relevant categories
	Data accessibility: difficult to access	Use CountrySTAT
Data quality	Divergence between different measures of same item – lack of consistency, coherence, and comparability	Improve specificity of definitions, borrow ideas for training and supervision from India, monitor quality of administrative data, as in India and Mexico. The Improvement of Crop Statistics (ICS) scheme involves supervising data collectors to verify the accuracy of the basic data. The Timely Reporting Scheme (TRS) is an effort to improve the timeliness of the data.

Section/Topic	Gap/weakness	Possible solution(s)
Uses of administrative data, especially in developed countries	Many uses not widely applied in developing countries:	Adapt the various uses to developing countries and apply them
	Integrating data from multiple sources. Inconsistent data published without any clarification. Non-use of record linkages	Methodology for record linkage and evaluation measurement error is necessary to maintain high-quality databases that integrate multiple administrative files. Combining census and administrative data to create efficient frames, statistical models and manual review processes and probabilistic record linkage. Software is also available. There are also examples from India, Mozambique and Uganda.
	Small-area estimation	Use administrative records as covariates in constructing model based small area estimates and forecasts. Draw lessons from the example of Ethiopia.
	Improving efficiency of survey-based estimators, e.g. provide population control totals	Use calibration (Carfagna and Carfagna, 2010)
	Improving frame construction and sampling design	Construct area frames

ANNEX 4
SAMPLE MEMORANDUM OF UNDERSTANDING BETWEEN THE NATIONAL AGRICULTURAL STATISTICS SERVICE AND THE NATIONAL RESOURCES CONSERVATION SERVICE IN THE USA



MEMORANDUM OF UNDERSTANDING
THE UNITED STATES DEPARTMENT OF AGRICULTURE
BETWEEN
THE NATURAL RESOURCES CONSERVATION SERVICE
AND
THE NATIONAL AGRICULTURAL STATISTICS SERVICE

RELATIVE TO THE
INTERCHANGE OF DATA AND STATISTICAL INFORMATION
AND
COORDINATION OF OUTREACH STRATEGIES

LEGAL AUTHORITY

The Agricultural Marketing Act of 1946, U.S. Code Title 7, 1621-6127 and 2204g, the Agricultural Adjustment Act of 1938 (1938 Act), the Soil Conservation Act of 1935, as amended, and the Food Security Act of 1985, as amended.

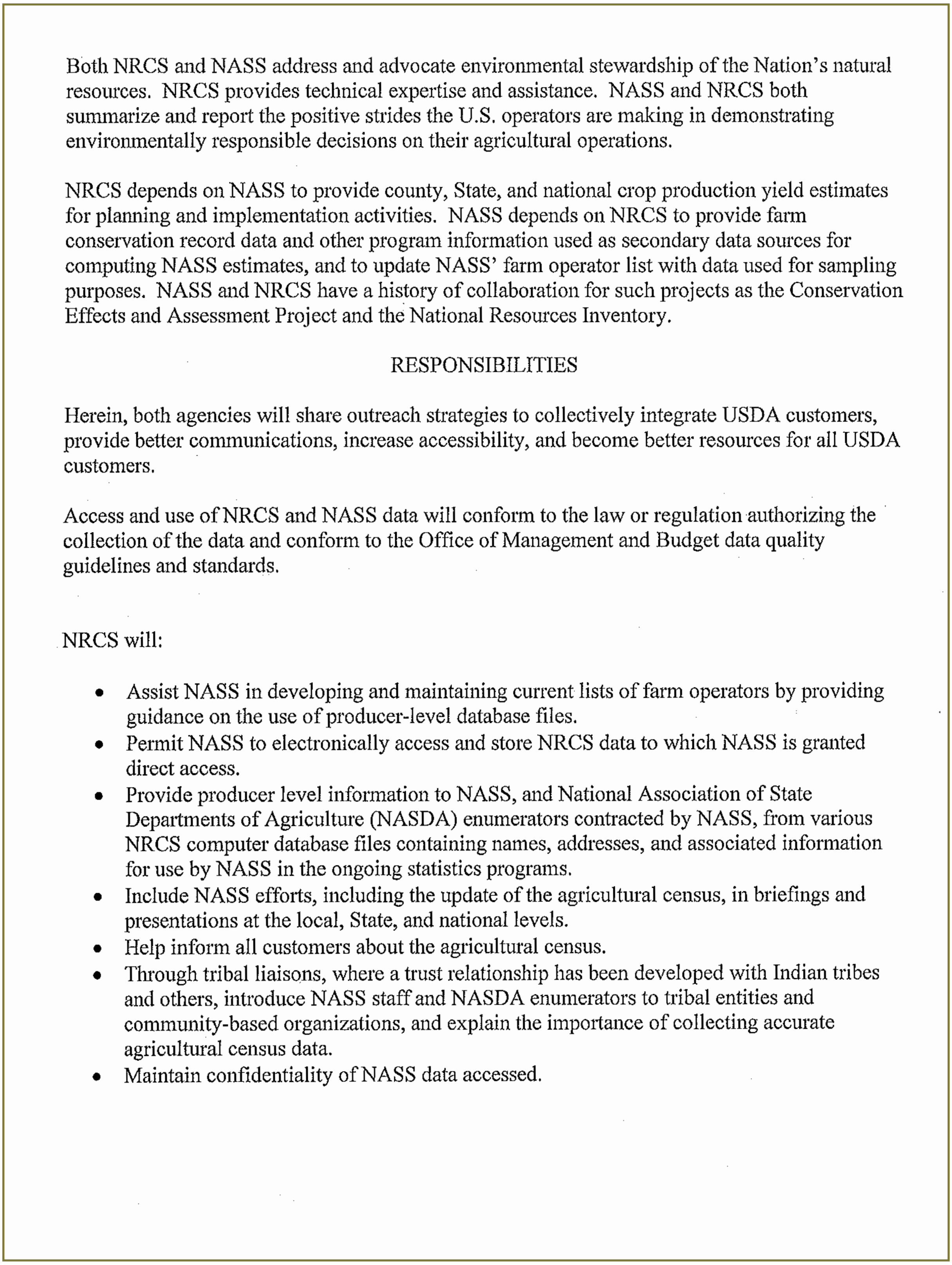
STATEMENT OF PURPOSE

This Memorandum of Understanding (MOU) establishes the framework for a relationship between the United States Department of Agriculture's (USDA) Natural Resources Conservation Service (NRCS) and the National Agricultural Statistics Service (NASS), in support of USDA activities. This MOU helps foster and enhance the interchange of data and information about the Nation's farms and agriculture to serve the best interest of USDA, the agricultural community, and the Nation.

Both agencies desire that activities under this MOU help foster and enhance the exchange of strategies (both traditional and non-traditional) for reaching and servicing underserved agricultural communities in the Nation, including American Indians, Alaska Natives, and Indian tribes. This MOU will help increase access into these communities to establish credibility and presence of NRCS and NASS, and increase participation in NRCS and NASS programs and services.

BACKGROUND

NRCS and NASS have a long history of cooperation and coordination. Both agencies have organizations that provide their primary channel of communication to the public, and provide data and information to the Nation's farm operators and agricultural business sector.



Both NRCS and NASS address and advocate environmental stewardship of the Nation's natural resources. NRCS provides technical expertise and assistance. NASS and NRCS both summarize and report the positive strides the U.S. operators are making in demonstrating environmentally responsible decisions on their agricultural operations.

NRCS depends on NASS to provide county, State, and national crop production yield estimates for planning and implementation activities. NASS depends on NRCS to provide farm conservation record data and other program information used as secondary data sources for computing NASS estimates, and to update NASS' farm operator list with data used for sampling purposes. NASS and NRCS have a history of collaboration for such projects as the Conservation Effects and Assessment Project and the National Resources Inventory.

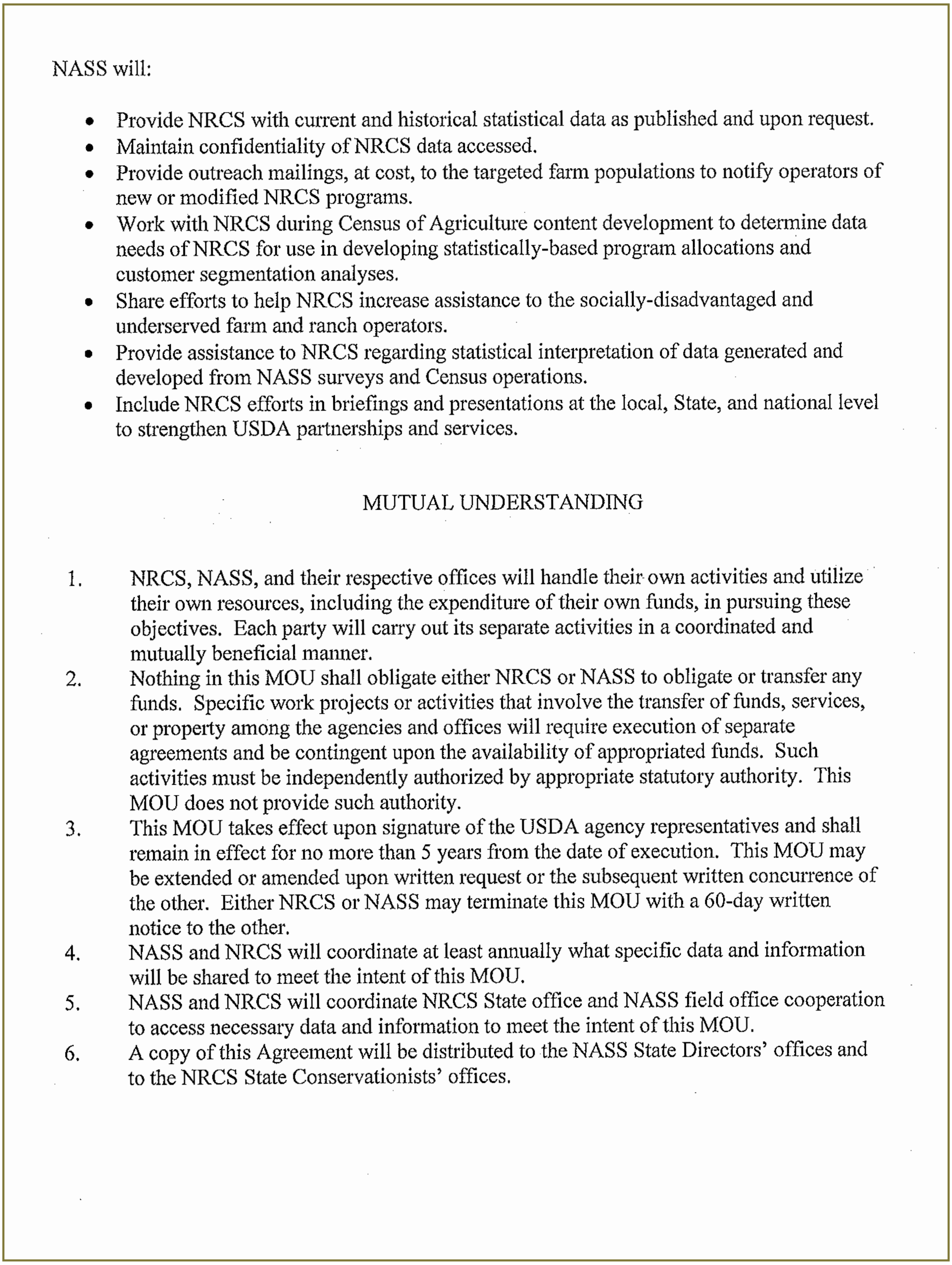
RESPONSIBILITIES

Herein, both agencies will share outreach strategies to collectively integrate USDA customers, provide better communications, increase accessibility, and become better resources for all USDA customers.

Access and use of NRCS and NASS data will conform to the law or regulation authorizing the collection of the data and conform to the Office of Management and Budget data quality guidelines and standards.

NRCS will:

- Assist NASS in developing and maintaining current lists of farm operators by providing guidance on the use of producer-level database files.
- Permit NASS to electronically access and store NRCS data to which NASS is granted direct access.
- Provide producer level information to NASS, and National Association of State Departments of Agriculture (NASDA) enumerators contracted by NASS, from various NRCS computer database files containing names, addresses, and associated information for use by NASS in the ongoing statistics programs.
- Include NASS efforts, including the update of the agricultural census, in briefings and presentations at the local, State, and national levels.
- Help inform all customers about the agricultural census.
- Through tribal liaisons, where a trust relationship has been developed with Indian tribes and others, introduce NASS staff and NASDA enumerators to tribal entities and community-based organizations, and explain the importance of collecting accurate agricultural census data.
- Maintain confidentiality of NASS data accessed.

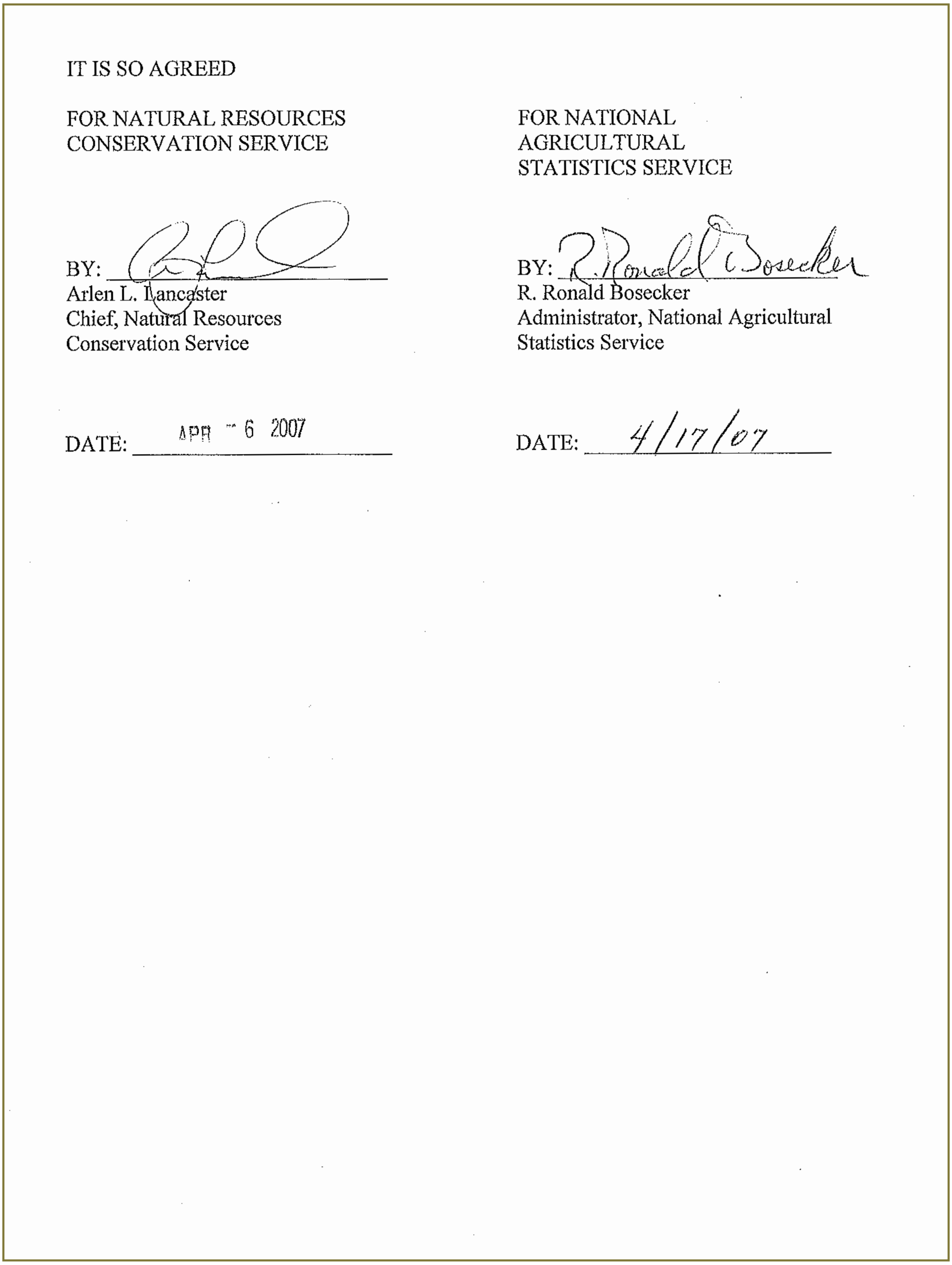


NASS will:

- Provide NRCS with current and historical statistical data as published and upon request.
- Maintain confidentiality of NRCS data accessed.
- Provide outreach mailings, at cost, to the targeted farm populations to notify operators of new or modified NRCS programs.
- Work with NRCS during Census of Agriculture content development to determine data needs of NRCS for use in developing statistically-based program allocations and customer segmentation analyses.
- Share efforts to help NRCS increase assistance to the socially-disadvantaged and underserved farm and ranch operators.
- Provide assistance to NRCS regarding statistical interpretation of data generated and developed from NASS surveys and Census operations.
- Include NRCS efforts in briefings and presentations at the local, State, and national level to strengthen USDA partnerships and services.

MUTUAL UNDERSTANDING

1. NRCS, NASS, and their respective offices will handle their own activities and utilize their own resources, including the expenditure of their own funds, in pursuing these objectives. Each party will carry out its separate activities in a coordinated and mutually beneficial manner.
2. Nothing in this MOU shall obligate either NRCS or NASS to obligate or transfer any funds. Specific work projects or activities that involve the transfer of funds, services, or property among the agencies and offices will require execution of separate agreements and be contingent upon the availability of appropriated funds. Such activities must be independently authorized by appropriate statutory authority. This MOU does not provide such authority.
3. This MOU takes effect upon signature of the USDA agency representatives and shall remain in effect for no more than 5 years from the date of execution. This MOU may be extended or amended upon written request or the subsequent written concurrence of the other. Either NRCS or NASS may terminate this MOU with a 60-day written notice to the other.
4. NASS and NRCS will coordinate at least annually what specific data and information will be shared to meet the intent of this MOU.
5. NASS and NRCS will coordinate NRCS State office and NASS field office cooperation to access necessary data and information to meet the intent of this MOU.
6. A copy of this Agreement will be distributed to the NASS State Directors' offices and to the NRCS State Conservationists' offices.



IT IS SO AGREED

FOR NATURAL RESOURCES
CONSERVATION SERVICE

BY: ________________________
Arlen L. Lancaster
Chief, Natural Resources
Conservation Service

DATE: APR - 6 2007

FOR NATIONAL
AGRICULTURAL
STATISTICS SERVICE

BY: R. Ronald Bosecker
R. Ronald Bosecker
Administrator, National Agricultural
Statistics Service

DATE: 4/17/07

Layout:

- Laura Monopoli

Cover photos:

- © FAO/Giuseppe Bizzarri
- © FAO/Jon Spaull
- © FAO/Florita Botts

www.gsars.org