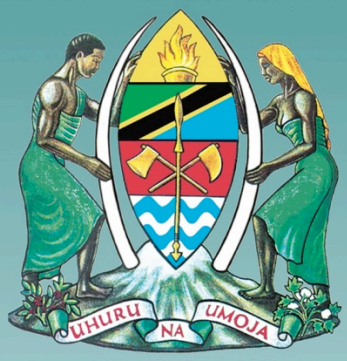


MANUAL FOR STRENGTHENING AND INTEGRATING ADMINISTRATIVE DATA INTO NATIONAL OFFICIAL STATISTICS OF TANZANIA

June, 2024

TABLE OF CONTENTS

FOREWORD

The manual for strengthening and integrating administrative data into National Official Statistics of Tanzania is the first guideline for administrative data to be developed in an attempt to bridge the huge data gap in reporting Tanzania's performance in various national, regional, and global development programs in Tanzania such as Tanzania Development Vision 2025 and the planned Vision 2050, East Africa Agenda 2050, The Africa we Want (Agenda 2063), the SDG 2030, and others.

The manual was developed by National Bureau of Statistics (NBS) and Office of the Chief Government Statistician (OCGS) in collaboration with stakeholders whose sectors are better placed in producing and supplying the needed data resulting from their administrative records and/or routine data systems. These stakeholders were from both Tanzania Mainland and Zanzibar; hence this manual is intended to be applied in the United Republic of Tanzania.

This Manual has been prepared within the implementation process of the Tanzania Statistical Master Plan II whereas the major goal amongst others is to improve the quality of administrative data produced by public and private agencies within the NSS. Currently, reporting Tanzania's progress towards the various development programs is mainly relying on data from surveys and/or censuses that contribute less than fifty percent of the requirements. On the other hand, various sectors have been generating data as part of discharging their day-to-day functions mandated under their sectors. These data for quite some times have not been used as part of official statistics mainly due to inherent quality challenges.

The main purpose of the Manual is to provide guidelines and methodologies for harmonizing and strengthening administrative data produced by institutions within NSS with the ultimate goal of enhancing the its quality that will pave way for the data to integrated into official statistics that are used in reporting the country's performance. The manual will ensure consistency, accuracy, and reliability in production of administrative data.

NBS and OCGS welcome all users and producers of administrative data to adopt this manual to guide the understanding of collection, compilation, analysis, publication and dissemination of quality administrative data that will lead into improved products and services rendered by the sectors.

Dr. Albina Chuwa
Statistician General
National Bureau of Statistics

Salum Kassim Ali
Chief Government Statistician
Office of the Chief Government Statistician
Zanzibar

ACKNOWLEDGEMENT

The National Bureau of Statistics (NBS) and Office of the Chief Government Statistician (OCGS) are sincerely acknowledging the Government of United Republic of Tanzania for approving and facilitating the implementation of the second Strategy for Development of Statistics in Tanzania that is known as the Tanzania Statistical Master Plan Phase Two 2022/23 – 2026/27 (TSMP II). We extend our sincere gratitude to the representatives from various institutions from both Tanzania Mainland and Zanzibar for their participation and cooperation from the stage of assessment of status of production of administrative data and identifying the challenges that can impact data quality, reliability and usability through development of this first-ever manual in Tanzania.

Our appreciations are extended to NBS and OCGS technical staff from different departments and units for their dedication and tirelessly playing key technical role in developing this manual. A number of technical staff from both coordinating institutions (NBS and OCGS) played pivotal role in this assignment. They include the following: Mr. Emilian Karugendo, Ms. Mariam Kitembe, Ms. Hellen Mtove, Ms. Elide Mwanri, Ms. Tumaini Kalindile, Ms. Jocelyn Rwehumbiza, Mr. Joseph Milanzi, Mr. John Mwangi, Ms. Johnia Kakiziba, Ms. Phausta Ntigiti, Ms. Jillahoma Wemmao, Ms. Rainer Kiama, Ms. Nacky Sinda, Ms. Veronica Mwangoka, Ms. Beatrice Rwegoshora, Mr. Saruni Njipay and Ms. Mary August from NBS. Also, deserving recognition are Mr. Bakari Makame, Ms. Hamisa Suleiman, Ms. Moza Omary and Mr. Suleima Suleiman from OCGS and Mr. Ali Shamte from Zanzibar Health Service Fund. Their dedication and commitment is what has resulted into this guiding document that it is our belief that it be implemented by players in the intended members of the NSS and that we shall be able to receive feedback within less extended duration for improvement of this manual, if such a need arise.

We also acknowledge the invaluable support of Dr. Lucia Barbone and Dr. Danielle Burke from Office for National Statistics (ONS) UK for providing technical assistance and mentorship during the process of developing of the manual.

We extend our gratitude to TSMP II Coordinator, Department of Research and Standards at NBS and Department of Standards, Statistical Coordination and

Researches at the OCGS for the coordination that made this manual a success. The manual for strengthening and integrating administrative data in the Tanzania National official statistics is expected to be reviewed in short intervals as it finds its usefulness among the unlimited number of stakeholders who house administrative data under their custody and would wish to see its contribution in bridging data gap the country is experiencing.

Dr. Albina Chuwa
Statistician General
National Bureau of Statistics

Salum Kassim Ali
Chief Government Statistician
Office of the Chief Government Statistician
Zanzibar

ACRONYMS

API	Application Programming Interface
AU	African Union
DPs	Development Partners
EAC	East African Community
FYDP	Five-Year Development Plan
GEWE	Gender Equality and Women Empowerment
GSBPM	Generic Statistics Business Process Model
KILM	Key Indicator for Labour Market
MDAs	Ministries, Departments and Agencies
MoU	Memorandum of Understanding
NBS	National Bureau of Statistics
NSAs	Non-State Actors
NSDS	National Strategy for the Development of Statistics
NSO	National Statistics Office
NSS	National Statistical System
OCGS	Office of the Chief Government Statistician
R&D	Research and Development
SDGs	Sustainable Development Goals
SADC	Southern Africa Development Cooperation
TDV	Tanzania Development Vision
TSMP II	Tanzania Statistical Master Plan Phase Two 2022/23 – 2026/27
UN	United Nations
URT	United Republic of Tanzania
ZADEP	Zanzibar Development Plan
ZDV	Zanzibar Development Vision

GLOSSARY OF TERMS

Data: A quantitative representation of facts, concepts, or instructions in a formal manner, suitable for communication, interpretation, or processing by humans or by automatic means.

Statistics: Any quantity (total, average etc) computed from values in a sample which is considered for a statistically purposes.

Statistical indicator: representation of statistical data for a specified time, place or any other relevant characteristic, corrected for at least one dimension (usually size) so as to allow for meaningful comparisons.

Statistical information: Any organised data obtained from census, surveys or administrative data.

Official statistics: Statistics produced, validated, compiled and disseminated by or under the authority of the NSOs.

Framework: Provides a general structure, guidelines offer specific guidance on how to implement that structure.

Para data: Data about the processes by which statistics are compiled. Paradata can refer to micro-data or aggregate data, and can be used to manage process quality.

Statistics production system: A systematic and scientific process which starts from planning going through designing, collection, processing and analysis to acquire required statistical information

CHAPTER ONE: INTRODUCTION

1.0 Background

The National Bureau of Statistics (NBS); The Chief Government Statistician (OCGS), as per the Statistical Act Cap 351 and The Office of Chief Government Statistician Act No.9 of 2007 are two Government agencies responsible for coordination, supervision, production and dissemination of official statistics and for the custodianship of official statistics in the country. These two agencies coordinate and supervise the National Statistical Systems (NSS) in the country. The coordination role of these statistics offices is to ensure that official statistics produced by all institutions within Tanzania are in line with the acceptable quality. This means that, statistics which are produced have to meet the set and agreed national and international rules, standards and methodologies.

It is fact that there are abundant administrative data that exists in various institutions. However, these data have not been used for compilation and production of statistics due to several reasons including insufficient understanding of the use of such data or being subjected to under coverage. Administrative data can be integrated or pooled from multiple institutions that produce similar type of data by using record linking or statistical matching or using modelling approaches. Though quality of administrative data may be good enough for administrative purposes but not sufficient for statistical purposes. Transforming administrative data into statistical data requires improving the quality of such data and dealing with conceptual differences.

In this regard, the two National Statistics Offices (NSOs), through the Tanzania Statistical Master Plan Phase Two (TSMP II) 2022/23 – 2026/27, aimed at improving production and use of administrative data through developing a manual of integrating administrative data in the production of official statistics. This manual explains guidance on concepts, definitions and standards on production of statistics; legal and policy framework; and improving quality of data and their accessibility and use. This Manual was developed by adopting other national and international guidelines. Cross reference from the Generic Statistics Business Process Model (GSBPM) and other statistics production frameworks has been extensively made to align the content of this manual with the steps and procedure which already exist.

1.1 Rationale

The complexity of statistical development needs more sources, it can be from censuses, surveys or through generating data from administrative records. Administrative data plays an important role in completing the statistical information and must be one of the sources for official statistics. Findings from recent assessments on data availability and gaps show that administrative data contribute to more than 60 percent (TSMP II) of the data required to monitor and report the implementation of national, regional and global development agendas. Yet, the administrative data system underperforms in terms of provision of required statistics and addressing data gaps compared to survey data system. Administrative data is rightful and very strategic information due to the continuity, completeness, and representation of small area statistics. The statistics derived from the administrative data is intensively used by various government agencies. Other benefits of administrative data include cost savings, reduced respondent burden, and improvements in the efficiency of macro-level estimators and small-area statistics.

The effective use of administrative data as well as the data from surveys and censuses is essential for informed decision-making, policy formulation, and program evaluation across various sectors in the country.

The utilization and harmonization of administrative data pose significant challenges. In most cases administrative data does not use statistical concepts and standards, but more on technical and operational approach due to the daily practice in delivering the public services. Timeliness, coverage and completeness of administrative data are unpredictable. To address these issues, NBS and OCGS decided to develop a comprehensive manual for integrating administrative data into national official statistics of Tanzania to strengthen coordination and production of reliable administrative data in the NSS. This will be achieved by strengthening modalities for coordination, production, management of administrative data systems, data assessment and sharing of routine data.

1.2 Objectives

The aim of producing this manual is to enable all institutions to improve the process of collecting, managing and using administrative data to produce reliable statistics in the country. This aims at providing guidelines and methodologies for harmonizing administrative data for filling data gaps in the monitoring of mid-term and long-term development plans (SDGs, FYDP III, ZADEP etc) and official statistical indicators across NSS to ensure consistency, accuracy, and reliability.

The manual for integrating administrative data will help to achieve the following specific objectives:

i. Establishing common data standards and formats to ensure uniformity in data collection, storage, and reporting;

ii. Addressing challenges related to data linkage, matching, and duplication;

iii. Implementing quality control measures in administrative data;

iv. Improved data quality and reliability through standardized data collection and processing procedures; and

v. Enhanced efficiency and effectiveness of data utilization across various stakeholders.

1.3 Scope of the Manual

The manual is designed for use across NSS. It is fit for use by all players engaged in the production of administrative data within the country. The goal is to improve the quality of administrative data by addressing the integration and harmonization; and aligning with the statistical production process to bridge administrative data gaps. This will result in enhanced overall administrative data quality.

1.4 Uses and Users of the Manual

The Manual on Strengthening and Integrating Administrative Data is intended for use by all players of the NSS. The Manual is designed to provide guidance to administrative data producers and compilers on best practices to ensure they produce statistics of acceptable standards and quality through administrative data system. However, uses of the manual may differ from one type of player to another. For instance, producers of statistics may find this manual to be of more use to them

compared to users of statistics. On the other hand, users of statistics may find this manual to be more useful when it comes to know how statistics have been produced with the intention of assessing their quality and fitness for purpose.

1.5 Structure of the Manual

This manual is organised into five chapters. Following this introduction chapter, the second chapter scans the situation on administrative data system of Tanzania. Chapter three and chapter four are the core chapters of this manual. Chapter three explains step-by-step procedure to be followed when producing statistics accordingly while considering the GSBPM. The fourth chapter introduces the aspect of quality management and how quality can be assured as far as production of administrative data is concerned. The last chapter explains the coordination of the administrative data and its overall management in the NSS.

CHAPTER TWO: ADMINISTRATIVE DATA SYSTEM IN TANZANIA

2.0 Introduction

The administrative data system in Tanzania encompasses various aspects of data collection, compilation, management, classification, and statistical analysis across social, demographic, and economic domains. In the following sections, practices of statistics production by NSS using administrative data and their management have been extensively explained.

2.1 National Statistical System of Tanzania

2.1.1 Composition and Structure

The National Statistical System (NSS) comprises the NSOs and all other stakeholders of statistics in the country. The NSS is composed of producers, compilers and users of statistics. This encompasses different groups and types of stakeholders including the NSOs (NBS and OCGS), Government Ministries, Departments and Agencies (MDAs), Local Government Authorities (LGAs), Research and Development (R&D) institutions, Development Partners (DPs) and other Non-State Actors (NSAs). Each one among these groups, has specific roles in the NSS.

To better coordinate and manage the NSS, the Government, through responsible agencies, set up arrangements on how these different players of the NSS are related and interlinked. The Government, through NSOs, regularly designs National Strategies for the Development of Statistic (NSDS) which is currently known as TSMP that serve as a tool to facilitate mechanism of coordinating and managing the NSS. Within this tool, different players of the NSS are brought together through structures like Statistical Working Groups (SWGs) and joint committees. The cornerstone for effective operations of these structures is their coordination and management.

2.1.2 NSS Legal and Policy Framework

The legal framework for statistics consists of the laws and legally-binding regulations that govern the statistical agencies, relationship between members of the national statistical system and respondents for the provision of data, as well as all aspects of data compilation, dissemination and data use. Legal framework gives the statistical agencies a clear mandate to collect data and to compile and disseminate statistics, and makes relevant information, such as supplementary data sources available for compilation.

Well-developed national legislation regulating the production of official statistics is necessary to ensure the implementation of the principles described in UN Fundamental Principles of Official Statistics. The overall target is to ensure the production of official statistics of a high level of quality, meeting users' needs and being trusted by users and other stakeholders. For the National Statistical Offices in Tanzania, The Statistics Act Cap. 351 and The Office of Chief Government Statistician Act No. 9 of 2007 for Zanzibar give mandate to NBS and OCGS to coordinate the National Statistical System. Other institutions have their legal and policy frameworks applied in their institutions according to their core functions but are not specific for production of official statistics. Both Statistics Act considers the official statistics produced by NSS if they met criteria and standard set by the Statistician General/Chief Government Statistician and they are not allowed to collect data through census or survey except with approval of Statistician General/Chief Government Statistician. There is a need for the legal and institutional framework of other statistical agencies to guarantee the fundamental values and principles that govern the development, production and dissemination of official statistics and be respected at all political levels and by all stakeholders in NSS. Other institutions are advised to propose legislation with a long-term strategy because opportunities to pass legislation are limited and effectively impacting legal frameworks requires substantial effort.

A successful national statistical system has flexibility to respond to changing conditions and circumstances without needing to change its primary legislation frequently. A flexible legal environment minimizes legislation changes and the associated risk of political interference with the legislation when it is opened for revision. Once the legal and policy frameworks are in place to permit the use of

statistics, including administrative data, it is necessary to consider the organizational arrangements to facilitate data flows in the form of written agreement of Memorandum of Understanding (MoU).

2.2 Official Statistics

In Tanzania, as in any other country, statistics are produced by various entities and come in various formats. However, not all statistics which are produced deserve to be taken and published as country statistics. For statistics to be published as country statistics, they need to be official as stipulated under both Statistics Act and regulations.

The statistics produced by any entity, including government institution and other agencies, shall qualify to be official statistics if they meet criteria and standards set approved by Statistician General of NBS or the Chief Government Statistician of OCGS. These two offices are the only one which have been legally mandated to stamp statistics as being official. However, this mandate does not deprive any organization an opportunity of producing statistics for its own use. In order for the statistics to be endorsed as being official, there are set procedure that producer has to follow to have such endorsement. These processes and procedures are stipulated in various manuals and guidelines which are developed by the NSOs in Tanzania.

2.3 Data needs and supply within NSS

National Statistics Offices in Tanzania are producing official statistics through census and surveys. These statistics are used for planning, monitoring and evaluation; and reporting of programs as well as for policy formulation and decision making. The statistics produced by NSOs are not enough to satisfy the needs and demand of data users. Administrative data is essentials to bridge the data gaps and be able to track progress and demonstrate results on national, regional and international programs over short, medium and long-term periods. Administrative data is rich in information although not necessarily designed for statistical purposes, it is however useful in providing comparable statistics used to assess countries compliance and performance to national, regional and international programs. The following are the development plans, framework and programmes committed by Tanzania which has high demand of statistics.

2.3.1 Global programs, frameworks and agendas

The United Republic of Tanzania (URT) committed to implement the 2030 Agenda for Sustainable Development Goals (SDGs) since January 2016 that are mainstreamed in the National Development Frameworks such as National Five-Year Development Plans (FYDP II & III), Zanzibar Development Plan (ZADEP), Tanzania Development Vision 2025, Zanzibar Development Vision 2050 (ZDV50), as well as Sectoral Strategic Plans. About 246 out of 248 SDGs indicators recently were adopted and localized to suit the country's context of which 61 percent of the data were obtained from administrative data sources in 2016 compared to 39 percent obtained from surveys and census. In 2023, administrative data contributed 49 percent against 51 percent from surveys and census data. Therefore, there is a need to strengthen administrative data to fill the highly significant data gaps, of which its data are obtained from various sectors within NSS.

2.3.2 Regional Frameworks and Agendas

Tanzania abides itself to unite with other countries at regional and continental level to set up and implement various development frameworks that addresses development issues and challenges faced by these countries. At highest level there is an African Union (AU) in which Tanzania is a member. Below it, there is Southern African Development Community (SADC) in which our country is also a member and at the lowest, there is an East African Community (EAC) where also we are among the member state. In each of these unions, there development ambitions and aspiration that member states wish to achieve at some time. Each member country allocates resources to implement programs earmarked to achieve those ambitions and aspirations. Likely, monitoring of the progress towards achievement of the set goals need to be informed using statistics that many countries including Tanzania produce from surveys or administrative data systems.

2.3.3 National Frameworks and Plans

Planning for development in Tanzania Mainland is guided by FYDP in medium term. The FYDP in term are designed to reflect the aspirations of the TDV. On the other hand, Zanzibar relies on ZADEP as medium-term development plan. Similarly, ZADEP reflects aspiration of the long term ZDV50. Designing of all these plans and vision requires a lot of information to study and understand the trend and hence to

set up realistic targets to be achieved at some time later. In addition, monitoring of the progress and reporting the achievement also need to be well informed using statistics collected from surveys as well as administrative data system.

2.3.4 Sectoral Frameworks and Plans

Policies are designed or formulated at institutional level but its implementation has a spill over effect in the sense that it involves other stakeholders. Also, set up of the Government structure involve sectors which in some cases include more than one institution working together. In most cases, national programs are implemented at sectoral level whereby more than one institution are involved. Under such circumstance, having sectoral plans is common and therefore statistics are required for monitoring implementation of those programs as well as reporting their performance and achievements.

2.3.5 Institutional Plans

Designing and development of institutional strategic plans, business plans as well as budgeting and annual action plans need to be well informed using data and statistics produced by the respective institution. In addition, policy formulation and decision making at all level must rely on facts in order for them to be genuine and realistic. Therefore, statistics from administrative data system from any institution are key drivers of all planning and decision-making process within that particular institution let aside the importance of such statistics to other institutions, sectors and the country at large.

2.4 Administrative Data Management

Tanzania's administrative data management system is integrated in institutional and sectoral structures. The NBS and OCGS play a central role in coordinating and overseeing the administrative data system through development of guidelines and set up standards on how statistics should be produced.

2.5 Classifications of Types of Administrative Data

The NBS and OCGS adopted the global classification of types of statistics. On the other hand, administrative data in Tanzania are classified into various categories based on the sectors or domains they represent. This includes data related to social services, such as health and education, as well as economic activities, agriculture, and other areas.

2.6 Social and Demographic Statistics

Social and demographic statistics cover a wide range of topics related to population characteristics, and social services such as health, education, good governance, crime, justice and other aspects of social welfare. Demographic data includes population size, growth rate, age distribution, fertility rates, mortality rates, migration patterns, and urbanization trends. Social statistics encompass indicators such as literacy rates, school enrolment rates, access to healthcare services, housing conditions, poverty levels, and social protection coverage.

Most of the produced statistics under the social and demographic domain are household based, whereby household is the primary source for data. Among the key statistics collected at household level are statistics on poverty measure. Currently there is no alternative mechanism for collecting data on poverty beside using household-based surveys. However, in administrative data systems, statistics within that domains are collected based on individual level records compiled by their respective authoritative institutions.

2.7 Economic Statistics

Majority of the key macro-economic statistics indicators are compiled and produced by NBS and OCGS using data from different sources including the administrative data systems. Economic statistics in Tanzania focus on measuring various aspects of economic activities including production, consumption, investment, trade, and employment. Key economic indicators include Gross Domestic Product (GDP), inflation rate, unemployment rate, sectoral contributions to GDP (such as agriculture, industry, and services), balance of payments, and foreign direct investment. Economic surveys and administrative data sources provide valuable information for analysing economic trends, monitoring sectoral performance, and formulating

economic policies. Information collected from the administrative data systems serve as potential inputs to compilation of these indicators.

2.8 Cross-Cutting Issues

Cross-cutting issues refer to themes or concerns that transcend specific sectors and have implications across multiple domains. Efforts to address cross-cutting issues often involve integrating data from different sources, conducting specialized surveys or studies, and implementing policies and programs that promote inclusive and sustainable development.

In the context of Tanzania NSS, among others, cross-cutting issues include the following themes:

i. Gender;
ii. Governance;
iii. Environment; and
iv. Climate change.

2.9 Challenges in Administrative Data collection

Efficient management and utilization of administrative data across these domains are essential for informing policy making, monitoring progress towards development goals, and addressing socio-economic challenges in Tanzania. Continuous efforts to improve data quality, enhance data governance frameworks, and strengthen statistical capacity are vital for maximizing the value of administrative data for evidence based decision making.

Beside efforts made Tanzania on improving the NSS, production of statistics through administrative data systems still faces various challenges that can impact data quality, reliability, and usability. Among the challenges in administrative data collection include:

i. Inconsistent data entry practices;
ii. Lack of standardized data collection procedures;
iii. Limited scope associated with under coverage;
iv. Inconsistent statistics on the same indicator if reported from different sources;
v. Insufficient training and capacity among staff;
vi. Technological limitations;

vii. Limited sharing of data among key actors due to lack standard data exchanges;
viii. Limited availability and use of metadata;
ix. Data privacy and security concerns.

Addressing these challenges requires a concerted effort from organizations to invest in quality improvement measures, adopt best practices in data collection and management, and prioritize data integrity, accuracy, and security.

CHAPTER THREE: GUIDELINES FOR PRODUCTION OF ADMINISTRATIVE STATISTICS

3.0 Frameworks for Statistical Production

Frameworks that guide statistical production include legal framework, policies, manuals, guidelines and plans. Statistical agencies and other producers should strive to shape data collected through their statistical production frameworks in a way that satisfies the requirements of national and international standards as well as meeting user needs. Some of the key elements that are recommended and common in most of the existing statistical production frameworks are:

i. ***Data protection:*** Statistical agencies should have data protection legal framework that regulate the collection, processing, and storage of personal data. These frameworks establish principles for data confidentiality, such as limiting access to authorized personnel, obtaining consent for data collection, and implementing security measures to prevent unauthorized disclosure;

ii. ***Confidentiality:*** Statistical agencies and third-party entities that handle administrative data may be required to enter into confidentiality agreements guided by existing policies to protect the confidentiality of the data they access or process. These agreements typically outline the obligations of the parties to maintain confidentiality and the consequences of unauthorized disclosure.

iii. ***Access controls:*** Statistical agencies should implement access controls to limit access to administrative data only to authorized personnel with a legitimate need to know. This includes user authentication mechanisms, role-based access controls, and audit trails to monitor access and detect unauthorized activities.

iv. ***Anonymization:*** To mitigate the risk of re-identification and unauthorized disclosure, NBS and OCGS encourage or require the anonymization of administrative data be in undertaken before sharing or processing. Anonymization involves removing or encrypting personally identifiable information.

v. ***Encryption:*** Statistical agencies are required to encrypt administrative data during storage, transmission, and processing to protect it from unauthorized

access or interception. Encryption techniques such as encryption-at-rest and encryption-in-transit help safeguard data confidentiality by rendering it unreadable to unauthorized parties.

vi. ***Data minimization:*** Statistical agencies may promote the principle of data minimization, which involves collecting and retaining only the minimum amount of data necessary for a specific purpose. By reducing the amount of sensitive information held, data minimization helps to limit the potential impact of a data breach or unauthorized disclosure.

vii. ***Data breach notification requirements:*** Policy and legal frameworks often include provisions requiring government agencies to notify affected individuals and relevant authorities in the event of a data breach involving administrative data. Prompt notification allows affected individuals to act appropriately to protect themselves from potential harm.

viii. ***Penalties for non-compliance:*** Legal frameworks typically establish penalties for non-compliance with confidentiality requirements, such as fines, sanctions, or legal liability for damages resulting from unauthorized disclosure or misuse of administrative data. These penalties serve as a deterrent against negligent or intentional breaches of confidentiality.

ix. ***Oversight and accountability mechanisms:*** Policy and legal frameworks may establish oversight mechanisms, such as data protection authorities or regulatory agencies, responsible for monitoring compliance with confidentiality requirements and enforcing penalties for violations. These mechanisms help to ensure accountability and promote trust in the handling of administrative data.

x. ***Data Classification:*** Implement a data classification scheme to categorize administrative data based on its sensitivity and confidentiality level. This helps identify which data requires the highest level of protection and ensures that appropriate safeguards are applied accordingly***.***

xi. ***Audit and Monitoring:*** Implement regular audits and monitoring mechanisms to ensure compliance with confidentiality policies and detect any unauthorized access or misuse of administrative data. This includes conducting periodic

security assessments, reviewing access logs, and investigating any suspicious activities or breaches.

Any agency which intends to collect data for the purpose of producing official statistics should observe and adhere to these elements. These elements are normally documented in statistical legislation, agreements or MoU, strategic documents, manuals and guidelines.

3.1 Generic Statistics Business Process Model

The official statistics are produced based on the international processes that are described in Generic Statistics Business Process Model (GSBPM) to ensure the quality of the statistics produced (Annex 3).

The GSBPM provides a standard framework and harmonised terminology to help NSOs and other organizations which produce statistics to modernise their statistical production processes, as well as to share methods and components. It can also be used for integrating data and metadata standards, as a template for process documentation, for harmonising statistical computing infrastructures, and to provide a framework for process quality assessment and improvement. The model is designed to be applicable regardless of the data source, so it can be used for the description and quality assessment of processes based on surveys, censuses, administrative data including registers, and other non-statistical or mixed sources.

3.2 Processes for production of administrative statistics

The GSBPM is a reference model for any statistical program. It is intended that the GSBPM may be used by organisations to different degrees. An organisation may choose to either implement the GSBPM directly or use it as the basis for developing customised version of the model. It may be used in some cases only as a model to which organisations refer when communicating internally or with other organisations to clarify discussion. The various scenarios for the use of the GSBPM are all valid.

In Tanzania, it has been used as a basis for developing a customized version of the guidelines for production of administrative statistics (Figure 3.1). Through this customization, the following phases have been designed to be adapted during the process.

i. Specify needs

ii. Design and build statistical program

iii. Data collection

iv. Data processing

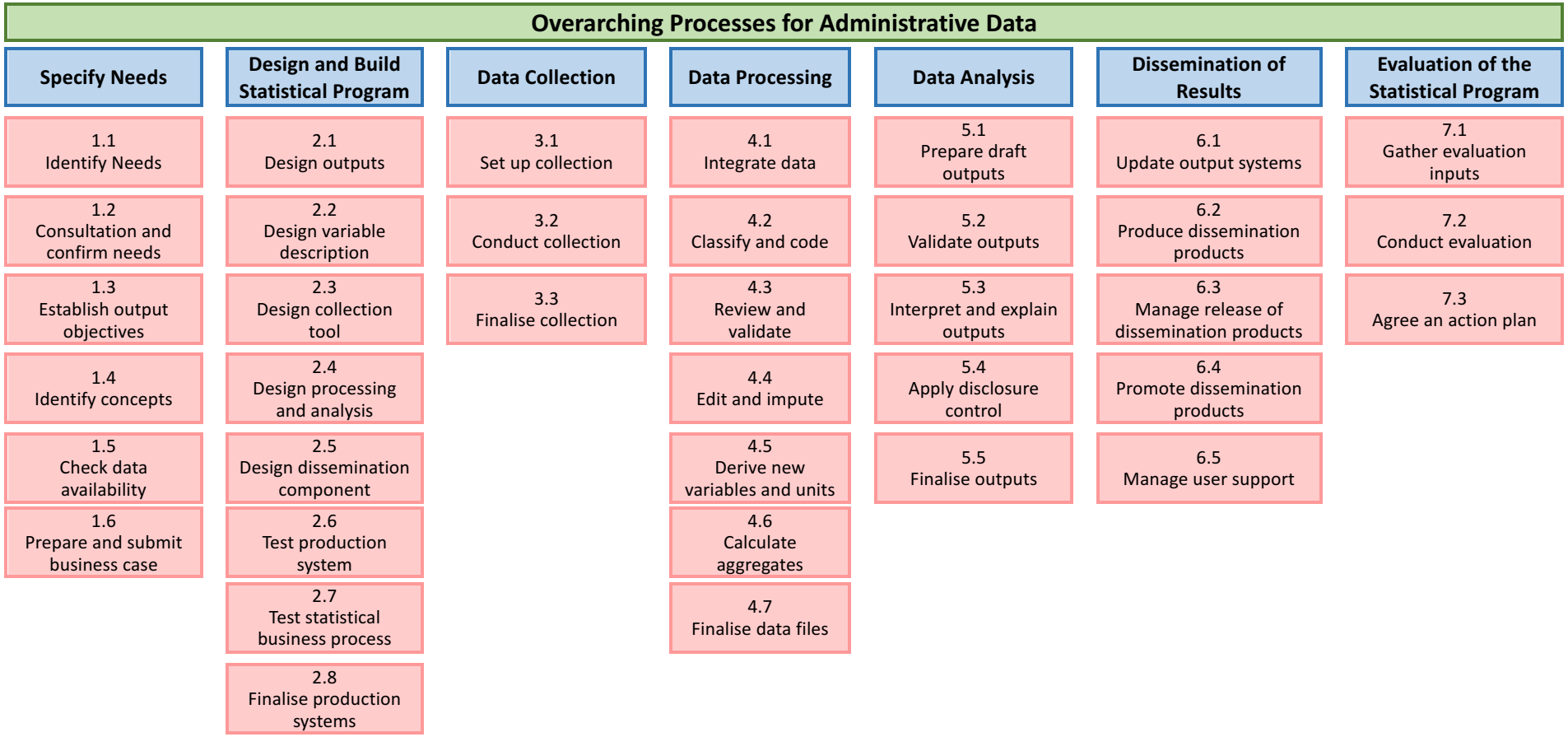
v. Data analysis

vi. Dissemination of results

vii. Evaluation of the statistical program

Thorough description and processes in each of these phases will be provided in the following sections.

Figure 3.1: The structure of the Tanzania National Statistics Business Process Model



3.2.1 Specify needs

Whenever there are new statistics that need to be produced, it is critical that the exact needs of such statistics be identified and established. Also, identifying needs is crucial when some ongoing routine statistics production program need to be reviewed. In some cases, needs might come from users which will trigger changes to the ongoing statistics production process. In all these and other circumstances that may arise, the ultimate goal of any statistics production process is to ensure it meets user needs.

During the design of any statistics production, consideration should be made at all levels of reporting. Beside institutional requirements and needs which may call for the need of new statistics production, there are sectoral, national, regional and global needs of such statistics. In order to identify such needs, users and stakeholders' involvement is very critical when identifying needs. The following guidelines elaborate mechanisms through which producer of statistics can ensure inclusive of needs from users.

3.2.1.1 Identify Needs

Make an initial investigation and identification of what statistics are needed and for what purpose. This initial investigation may involve internal review of institutional requirement and review of the requirements from national and other frameworks and agendas. This process will form an initial draft of requirements.

3.2.1.2 Consultation and confirm needs

Undertake consultation to various intended users of the statistics to be produced in order to confirm the initial draft that was proposed. Users may involve internal and external stakeholders.

3.2.1.3 Establish output objectives

Once needs have been identified and confirmed, the producer needs to set clear output objectives of producing the intended statistics. All considerations (legal and technical) should be taken on board to avoid any constraint during production process;

3.2.1.4 Identify concepts

Identify and clarify the concept to be measured. This is a point where producer has to link and align users' needs with existing statistical standards and concepts.

3.2.1.5 Check data availability

Check availability of data from any source that will make it possible to address the identified needs. The aim here is to minimize the cost of collection and production of new statistics. If partial data is available from existing source, then the new production should be made just to compliment what is missing from the available data. It is important to conduct an exhaustive assessment to all possible sources to identify any existing data prior to decide on the designing of new data collection. If some data are available from other organization, it is better to set up legal arrangements of accessing those data through data sharing agreements or any mechanism to ensure that such data can be accessed and utilized for statistics production.

3.2.1.6 Prepare and submit business case

Once needs for new data collection has been identified and existing data from other sources being reviewed and marked, the following step is to develop and submit a business case. This should be in the form of either complete proposal or a concept note or just Terms of References. In either form, it should be comprehensive and identify clearly the needs. The business case to be developed may include, but not limited to, the background, objectives, benefits, costs, deliverables, time frame, budget, required technical and human resources, risk assessment and impact on stakeholders for each option.

3.2.2 Design and build statistical program

This phase involves the development, designing, building and testing the statistics production solution required to define the identified statistical needs outputs, concepts, methodologies, collection instruments and operational processes as proposed by the business case. For statistical outputs produced on a regular basis, this phase usually occurs for the first iteration and whenever improvement actions are identified. This phase makes substantial use of international and national

standards and guidelines in order to reduce the length and cost of the design and build process, and enhance the comparability and usability of outputs.

The following guidelines explain how this phase could be successfully implemented:

3.2.2.1 Design outputs

The outputs are designed based on the objectives of the data collection and user needs. Statistics are normally produced through indicators, therefore, all indicators intended to be produced need to be clearly identified with their associated metadata. Metadata will enable producer of statistics to identify variables which are required to be involved in the questionnaire of form which will be used for data collection.

3.2.2.2 Design variable description

The next process is to design all variables that need to be collected according to indicators which need to be addressed. Consideration should also be made to variables which will be computed using one or more variables which have been already created. For example, if the interest is to collect age, the design should consider design a variable which will satisfy the need of producing age groups as well as single year age. Under such circumstance, the designed variable should be single year from which age groups can be derived as a computed variable. Some variables may not be included as core variables but rather as classification or disaggregation variables. These may include, for example, variable on sex, geographical location, etc.

3.2.2.3 Design collection tool

Once all variables have been identified and well designed, a collection tool (questionnaire or form) is designed. Prior to design the questionnaire, a response unit need to be clearly identified. If more than one response unit is targeted (for example, an establishment and an individual) then separate questionnaires should be designed, one for each response unit. Try to avoid mixing questions for two different units of response in one questionnaire. The questionnaire or form which will be designed at this process have to meet at least the following characteristics:

i. Includes only necessary questions and avoid redundant questions;

ii. Have questions which are not ambiguous and easy to be understood by respondent;

iii. Questions should be properly sequenced;

iv. Have pre-coded questions as much as possible; and

v. Have questions whose responses are exhaustive.

The designed questionnaire or form may then be computerized so that it can be administered using Computer Assisted Personal Interview (CAPI). A comprehensive manual that explains details for each question in the form or questionnaire should also be developed to assist administration of that form during data collection, processing and analysis of data that will be collected.

3.2.2.4 Design processing and analysis

The following process is to design and build statistical processing and analysis methodology. This process includes:

i. Designing rules for coding;

ii. Designing rules for editing and imputation;

iii. Designing tabulation plan;

iv. Designing and building dummy tables; and

v. Designing report structure.

3.2.2.5 Design dissemination component

Design and build dissemination products may involve designing of key findings report, brochures and leaflets, dashboard, video and audio clips. Presentation of statistics that has to be released during dissemination stage may be of different forms including tables, graphs and charts or maps. These should be clearly identified and if there are additional needs required for their preparation, they have to be identified and designed at this stage.

3.2.2.6 Test production system

Once all design and building of statistics production system has been completed, the next stage is to test the system. Testing the system involves testing the data collection instruments (questionnaires or forms) to see if they work and return response as they are expected. Any discrepancy from the expected results should be addressed and rectified at this stage. It is important to pre-test the data collection instruments in environment at which they are going to be administered.

3.2.2.7 Test statistical business process

The next process is to pilot a complete statistical business process. This is beyond just testing the instrument but rather includes all processes involved in the production system including logistics and administrative arrangement, budgets and even the look of the final outputs.

3.2.2.8 Finalise production systems

Rectify any observed discrepancy of the pilot from the design. Changes should be made whenever necessary and everything then have to be finalised at this process. To complete this process, the following activities should be accomplished:

i. Documentation of all processes;

ii. Production of user manuals; and

iii. Training users of various processes including data collection, editing, quality check and administration of each process.

3.2.3 Data collection

Data collection of administrative records involves collecting or gathering all necessary information (e.g. data, metadata and paradata) using different collection modes (e.g. acquisition, collection, extraction, transfer), and loads them into the appropriate environment for further processing. Whilst it can include validation of data set formats, it does not include any transformations of the data themselves, as these are all done in the "Process" phase. For statistical outputs produced regularly, this phase occurs in each iteration.

The data collection phase of administrative data is implemented through the following three processes which are generally sequential, but can also occur in parallel, and can be iterative. These sub-processes are preparation, run collection and finalise collection.

3.2.3.1 Set up collection

Set up collection to ensures that the people, processes and technology (e.g. CAPI, web-based applications, GPS system) are ready to collect data and metadata, in all modes as designed. It takes place over a period of time, as it includes the strategy, planning and training activities in preparation for the specific instance of the statistical business process. This process includes the following sub-processes:

i. Preparing a collection strategy;

ii. Training staff who will fill the form on the administrative data collection instruments and the system in general;

iii. Ensuring collection resources are available (e.g. laptops, collection apps, APIs);

iv. Configuring collection systems to request and receive the data;

v. Ensuring the security of data to be collected;

vi. Preparing collection instruments (e.g. printing data collection forms, pre-filling them with existing data, loading data collection forms and data onto responsible staff computers, APIs, web scraping tools);

vii. Translating of materials into Kiswahili if necessary; and

viii. Ensures that the necessary processes, systems and confidentiality procedures are in place, to receive or extract the necessary information from the source. This includes:

- Evaluating requests to acquire the data and logging the request in a centralised inventory;
- Initiating contacts with organisations providing the data, and sending an introductory package with details on the process of acquiring the data;

- Checking detailed information about files and metadata with the data provider and receiving a test file to assess if data are fit for use; and
- Arranging secure channels for the transmission of the data.

3.2.3.2 Conduct collection

The next following process is actual collection of data using the collection instruments to collect or gather the information which may include raw microdata or aggregates produced at the source, as well as any associated metadata. It can include the initial contact with providers and any subsequent follow-up or reminder actions. It may include manual data entry at the point of contact depending on the source and collection mode. It records when and how providers were contacted, and whether they have responded. Depending on the geographical frame and the technology used, geo-coding may need to be done at the same time as collection of the data by using inputs from GPS systems, putting a mark on a map, etc. This process also includes the management of the providers involved in the current collection, ensuring that the relationship between the statistical organisation and data providers remains positive, and recording and responding to comments, queries and complaints.

Proper communication with reporting units and minimization of the number of non-responses contributes significantly to a higher quality of collected data. For administrative data, geographical or other non-statistical data, the provider is either contacted to send the information or sends it as scheduled. This process may be time consuming and might require follow-ups to ensure that data are provided according to the agreements. In case where data are published under an Open Data license and exist in machine-readable form, they may be freely accessed and used. This process also includes supervision and monitoring of data collection and making any necessary changes to improve data quality.

3.2.3.3 Finalise collection

Finalise collection by loading the collected data and metadata into a suitable electronic environment for further processing. This may include manual or automatic data capture, for example, using clerical staff or optical character recognition tools to extract information from data collection forms, or converting the formats of files or encoding the variables received from other organisations. In cases where there is a physical collection instrument such as data collection form, which is not needed for further processing, this process manages the archiving of that material. When the collection instruments use software such as an API or an app, this process also includes the versioning and archiving.

3.2.4 Data processing

Data processing phase describes the processing of input data and their preparation for analysis. It is made up of sub-processes that integrate, classify, check, clean, and transform input data, so that they can be analysed and disseminated as statistical outputs. For statistical outputs produced regularly, this phase occurs in each iteration. The following are processes which guide and can be applied to data from administrative sources:

3.2.4.1 Integrate data

Integration of data can be from one or more sources. It is where the results of sub process in the collect phase are combined. The input data can be from a mixture of external or internal data sources, and a variety of collection modes, including extracts of administrative data resulting in linked data. Data integration include matching data from multiple sources and prioritising, when two or more sources contain data for the same variable, with potentially different values. Following integration, depending on data protection requirements, data may be anonymised, that is stripped of identifiers such as name and address to help protect confidentiality.

3.2.4.2 Classify and code

Classification, coding and imputation of data where the coding should be done during the designing of data collection form or after collection using an automated process or an interactive manual process. For example, automatic

(or clerical) coding routines may assign numeric codes to text responses according to a pre-determined statistical classification to facilitate data capture and processing.

3.2.4.3 Review and validate

Reviewing and validating by examining data to identify potential problems, errors and discrepancies such as outliers, item non-response and miscoding.

3.2.4.4 Edit and impute

Editing and imputation where data are considered incorrect, missing, unreliable or outdated, new values may be inserted or outdated data may be removed in this process. The specific steps include determination of whether to add or change data; selection of the method to be used; adding/change data values; writing the new data values back to the data set, and flagging them as changed; and production of metadata on the editing and imputation process. When working with administrative data, it is often useful to suggest to data provider the inclusion of automated editing rules within the collection system, as this can lead to a dramatic increase in the quality of the administrative data.

3.2.4.5 Derive new variables and units

Deriving new variables and units that are not explicitly provided in the collection, but are needed to deliver the required outputs. New variables are derived by applying arithmetic formulae to one or more of the variables that are already present in the dataset, or applying different model assumptions. This activity may need to be iterative, as some derived variables may themselves be based on other derived variables. It is therefore important to ensure that variables are derived in the correct order. New units may be derived by aggregating or splitting data or by various estimation methods.

3.2.4.6 Calculate aggregates

Calculating aggregates to create aggregate data and population totals from microdata or lower-level aggregates. It includes summing data for records sharing certain characteristics (e.g. aggregation of data by demographic or geographic classifications), determining measures of average and dispersion. It is at this process where indicators are computed.

3.2.4.7 Finalise data files

Finalising data files to brings together results from other processes in this phase in a data file (usually macro-data), which is used as the input to the "analyse" phase. Sometimes this may be an intermediate rather than a final file, particularly for business processes where there are strong time pressures, and a requirement to produce both preliminary and final estimates.

3.2.5 Data analysis

In this phase, statistics are produced, examined in detail and made ready for dissemination. This phase includes the processes and activities that enable statistical analysts to understand the statistics produced. The data analysis phase is broken down into the following processes:

3.2.5.1 Prepare draft outputs

Preparation of draft outputs where the data collected are transformed into statistical outputs. It includes the production of additional measurements such as indices, trends or seasonally adjusted series, as well as the recording of quality characteristics.

3.2.5.2 Validate outputs

Validating outputs to check the quality of the outputs produced, in accordance with a general quality framework and with expectations. Validation activities can include:

i. Checking that the population coverage and response rates are as required;
ii. Comparing the statistics with previous cycles (if applicable);
iii. Confronting the statistics against other relevant data (both internal and external);
iv. Investigating inconsistencies in the statistics;
v. Performing macro editing;
vi. Validating the statistics against expectations and domain intelligence

3.2.5.3 Interpret and explain outputs

Scrutinizing and explaining where the in-depth understanding of the outputs is gained. Use that understanding to scrutinize and explain the statistics produced for this cycle by assessing how well the statistics reflect the initial expectations, viewing the statistics from all perspectives using different tools and media, and carrying out in-depth statistical analyses.

3.2.5.4 Apply disclosure control

Applying disclosure control in order to ensures that the data (and metadata) to be disseminated do not breach the appropriate rules on confidentiality. This may include checks for primary and secondary disclosure, as well as the application of data suppression or perturbation techniques.

3.2.5.5 Finalise outputs

Finalizing outputs in order to ensure that the statistics and associated information are fit for purpose and reach the required quality level, and are thus ready for use. It includes:

i. Completing consistency checks;

ii. Determining the level of release, and applying caveats;

iii. Collating supporting information, including interpretation, briefings, measures of uncertainty and any other necessary metadata;

iv. Producing the supporting internal documents;

v. Pre-release discussion with appropriate internal subject matter experts; and

vi. Approving the statistical content for release.

3.2.6 Dissemination of results

Dissemination of statistical products is the last step after the collection and analysis activities in order to ensure that the produced statistics are used for planning and decision making processes at different administrative levels from national, regional, district and down to lower administrative levels such as wards/shehia and villages.

Statistical products should be disseminated to stakeholders in various forms of publications such as hard copies, soft copies and by posting on the websites and dashboards. In order for the product to be user-friendly, charts, tables and relevant attachments should normally be included.

During dissemination of statistical products, producers should ensure that the products meet users' needs as follows:

3.2.6.1 Update output systems

Make dissemination and data preservation plan early in the statistical production that includes archiving, publishing and distribution. This is to verify and ensure that the released statistical products after all the processing steps are consistent with the source data. In the case of the derived variables, it means that one should be able to reproduce the same results from the source data;

3.2.6.2 Produce dissemination products

Produce dissemination products through: preparing the product components (explanatory text, tables, charts, quality statements etc.); assembling the components into products; and editing the products and checking that they meet publication standards. This could include printed publications, press releases, seminars, awareness program and web sites;

3.2.6.3 Manage release of dissemination products

Manage release of dissemination products for ensuring that all elements for the release are in place including: Timeliness and punctuality; Accuracy and reliability; Transparency; Accessibility and clarity; Coherence and comparability; and Statistical Confidentiality and security;

3.2.6.4 Promote dissemination products

Promote dissemination products for communicating statistical information to users. This concerns the active promotion of the statistical products produced in a specific statistical business process, to help them reach the widest possible audience.

3.2.6.5 Manage user support

Manage user support to ensure that users' queries and requests for services such as micro-data access are recorded, and that responses are provided within agreed deadlines. These queries and requests should be regularly reviewed to provide an input to the overarching quality management process, as they can indicate new or changing user needs.

3.2.7 Evaluation of the statistical program

This phase manages the evaluation of a specific instance of a statistical business process. It logically takes place at the end of the instance of the process, but relies on inputs gathered throughout the different phases. It includes evaluating the success of a specific instance of the statistical business process, drawing on a range of quantitative and qualitative inputs, and identifying and prioritising potential improvements. For statistical outputs produced regularly, evaluation should, at least in theory occur for each iteration, determining whether future iterations should take place. However, in some cases, particularly for regular and well established statistical business processes, evaluation may not be formally carried out for each iteration. In such cases, this phase can be seen as providing the decision as to whether the next iteration should start from the Specify Needs phase, or from some later phase (often the Collect phase). This phase is made up of three processes, which are generally sequential, from left to right, but which can overlap to some extent in practice. These processes are:

3.2.7.1 Gather evaluation inputs

The first process is gathering evaluation inputs where evaluation material can be produced in any other phase or sub-process. It may take many forms, including feedback from users, process metadata, paradata, system metrics, and staff suggestions. Reports of progress against an action plan agreed during a previous iteration may also form an input to evaluations of subsequent iterations. This sub-process gathers all of these inputs, and makes them available for the person or team producing the evaluation.

3.2.7.2 Conduct evaluation

The second process is to conduct the actual evaluation during which analysis of the evaluation inputs is done and synthesises into an evaluation report. The resulting report should note any quality issues specific to this iteration of the statistical business process, and should make recommendations for changes if appropriate. These recommendations can cover changes to any phase or process for future iterations of the process, or can suggest that the process is not repeated.

3.2.7.3 Agree an action plan

The third process is to develop and agree an action plan which will brings together the necessary decision-making power to form and agree an action plan based on the evaluation report. It should also include consideration of a mechanism for monitoring the impact of those actions, which may, in turn, provide an input to evaluations of future iterations of the process.

3.3 Fundamental Principles of Official Statistics

The fundamental principles of official statistics adopted by United Nations Statistical Commission at the global level for the purpose of ensuring the statistics produced provide the reliable information of high quality statistics to support evidence-based decision-making. For administrative data to be recognized as official, the NSS have to comply with the ten adopted principles of official statistics as follows:

i. **Relevance, impartiality and equal access**: Official statistics provide an indispensable element in the information system of a democratic society, serving the government, the economy and the public with data about the economic, demographic, social and environmental situation. To this end, official statistics that meet the test of practical utility are to be compiled and made available on an impartial basis by official statistical agencies to honour citizens' entitlement to public information.

ii. **Professional standards and ethics**: To retain trust in official statistics, the statistical agencies need to decide according to strictly professional considerations, including scientific principles and professional ethics, on the

methods and procedures for the collection, processing, storage and presentation of statistical data.

iii. **Accountability and transparency**: To facilitate a correct interpretation of the data, the statistical agencies are to present information according to scientific standards on the sources, methods and procedures of the statistics.

iv. **Prevention of misuse**: The statistical agencies are entitled to comment on erroneous interpretation and misuse of statistics.

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

vi. **Confidentiality:** Individual data collected by statistical agencies for statistical compilation, whether they refer to natural or legal persons, are to be strictly confidential and used exclusively for statistical purposes.

vii. **Legislation**: The laws, regulations and measures under which the statistical systems operate are to be made public.

viii. **National coordination:** Coordination among statistical agencies within countries is essential to achieve consistency and efficiency in the statistical system.

ix. **Use of international standards**: The use by statistical agencies in each country of international concepts, classifications and methods promotes the consistency and efficiency of statistical systems at all official levels.

x. **International cooperation:** Bilateral and multilateral cooperation in statistics contributes to the improvement of systems of official statistics in all countries.

3.4 Integration of Gender Dimensions in Administrative Data

To ensure the integration of gender dimension into administrative data, the following are to be taken into account:

i. Specify gender data needs;
ii. Build the relationships with users of gender-specific data and create awareness of data needs and the processes of data production;

iii. Equip with adequate knowledge of gender-sensitive SDG indicators and highlighting those relevant to their domain;
iv. Review data collection forms of administrative data to include gender issues relevant to sector;
v. Build strong coordination and cooperation with NBS/OCGS to ensure harmonization of methodologies and data production and management processes, and timely production of indicators;
vi. Adopt better data management processes;
vii. Create awareness among users regarding the type of gender-sensitive data that existing administrative data systems regularly produced;
viii. In collaboration with NBS/OCGS conduct periodic data quality assessments of existing administrative data systems;
ix. Work with the SDG monitoring agency in the country to communicate monitoring, reporting and evaluation mechanisms related to the SDG framework to improve compliance and sustainability in gender data production and dissemination;
x. Publish gender-sensitive administrative data reports in collaboration with NBS/OCGS; and
xi. Establish systems to increase the accessibility of gender-sensitive administrative data and related products to users through regular statistical releases, publications and data sharing.

CHAPTER FOUR: QUALITY MANAGEMENT IN THE PRODUCTION OF ADMINISTRATIVE DATA

4.0 Introduction

Quality management is defined in the GSBPM as the process that includes quality assessment and control mechanisms. It recognizes the importance of evaluation and feedback throughout the statistical business process. It ensures that an organization's products or services are consistent, meet users' expectations, and comply with the standards.

In assuring transparency statistical agencies' policies, management practices, terms and conditions under which statistics are developed, produced and disseminated should be documented and available to users, respondents, owners of data source and the public.

4.1 Guiding Principles

i. Identify quality indicators in order to measure compliance with the respective quality principles and requirements as defined in the Tanzania Data Quality Assurance Framework.

ii. Set levels of requirements for the quality indicators in the form of quality targets which can serve as a tool for monitoring quality developments over time.

4.2 Importance of Quality in Administrative data

High-quality administrative data is essential for several reasons:

i. It forms the basis for informed realistic decision-making.

ii. It facilitates accurate monitoring and evaluation of programs and activities.

iii. It facilitates compliance with regulatory requirements.

iv. It supports transparency and accountability.

4.3 Strategies for improving Quality of Administrative Data collection

Improving the quality of administrative data collection involves implementing strategies aimed at enhancing accuracy, completeness, timeliness, consistency, relevance, and accessibility of the data. Here are several strategies to achieve this:

4.3.1 Clear Documentation of the Processes

Documenting data collection processes helps ensure consistency and transparency. Some key elements to include in the documentation are purpose and objectives, data collection methodology, data variables and definitions, roles and responsibilities and training materials. Organizations should develop clear guidelines and procedures for data collection, including definitions of variables, data entry protocols, and quality control measures.

4.3.2 Building Data Collection instrument

The collection instrument is generated or built based on the design specifications created during the design stage. A collection may use one or more means to receive the data. Collection instruments may also be data extraction routines used to gather data from existing statistical or administrative data sets. Administrative data collection instruments are specifically designed to gather information relevant to administrative purposes within organizations or institutions, this includes forms, applications, checklists and templates.

Building data collection instruments involves designing tools or forms that facilitate the collection of relevant information. The choice of instrument depends on factors such as the nature of the data, the target population, and the intended objectives.

This also includes preparing and testing the contents and functioning of that instrument (e.g. testing the questions in a questionnaire). It is recommended to consider the direct connection of collection instruments to the statistical metadata system, so that metadata can be more easily captured in the collection stage. Connection of metadata and data at the point of capture can save work in later stage.

4.3.3 Data Collection Standardization

Data collection standardization refers to the process of establishing and implementing uniform procedures, methods, and protocols for collecting data in a consistent and systematic manner. It involves defining and adhering to a set of

standards and guidelines to ensure that data is collected accurately, reliably, and efficiently.

Data collection standardization is essential for ensuring the integrity, comparability, and usability of collected data for analysis, interpretation, and decision making purposes. It helps minimize errors, enhance data quality, and increase confidence in the findings derived from the collected data.

Data collection standardization involves the following:

i. **Scope**: This refers to standardizing the specific techniques or approaches used to collect data.

ii. **Focus**: It emphasizes consistency in the way data is gathered, measured, or recorded.

iii. **Goal**: The primary goal is to ensure uniformity and comparability in the data collected, making it easier to analyze and interpret results across different contexts or time periods.

An organization should establish uniform data collection tools, such as forms and templates, and provide training to staff on their use.

4.3.4 Training and capacity building

Investing in staff training and capacity building is crucial for improving data collection quality. Training programs should cover topics such as data collection techniques, data quality standards, and the use of data collection tools.
The following steps can help organizations develop their workforce's proficiency in this area:

i. Conduct a needs assessment to identify the specific training needs and capacity gaps of personnel involved in data collection. This assessment should consider factors such as existing skills, knowledge levels, and areas requiring improvement.
ii. Engage subject matter experts, experienced data collectors, or external trainers with relevant expertise to facilitate training sessions, share best practices, and provide real-world examples.
iii. Provide ongoing support and mentoring to personnel throughout the training process and beyond. Encourage open communication, feedback, and

collaboration to address questions, troubleshoot issues, and reinforce learning objectives.

iv. Develop comprehensive training materials, manuals, guides, and resources to supplement training sessions and serve as reference materials for personnel. These materials should be accessible, up-to-date, and easily understandable.

v. Evaluate the effectiveness of training programs through pre-and post-training assessments, surveys, and feedback mechanisms.

4.3.5 Utilization of technology

Technology can enhance data collection efficiency and accuracy. Organizations should explore the use of electronic data capture systems, data validation tools, and other technological solutions to streamline data collection processes; select appropriate tools and technologies for data collection based on the nature of the information being collected; and ensure that data collection tools are reliable, user-friendly, and capable of capturing the required data accurately.

4.4 Metadata, Concepts and Definitions

Statistical metadata is commonly defined as data that describes information about other data. Metadata ensures the quality, interpretability and usefulness of datasets. The first and most fundamental purpose of metadata is to help users of statistical data to interpret, understand, and analyse statistical data.

4.4.1 Guiding Principles

i. Create a codebook that provides question level metadata matched to variables in the dataset at data processing stage of administrative data production;

ii. Provide sufficient metadata to ensure quality and add value to administrative data. Provide information covering the underlying concepts and definitions of the data collected and statistics produced, the variables and classifications used, the methodology of data collection and processing, and indicators of the quality of the statistical information in general, sufficient information to enable the user to understand all of the attributes of the statistics, including their limitations;

iii. Make available, comprehensive and clear metadata; and documentation about the administrative source in order to understand and interpret the data. Without this, it is not possible to understand and assess the administrative source against the intended use;

iv. Make available the system that stores all the necessary information collected during data collection; and

v. The metadata management system of the data producers should be well defined, documented, archived and disseminated according to internationally accepted standards. The data producers should ensure that the following elements are assured for managing metadata.

4.5 Access and Use of Administrative Data and Statistics

Statistics Act provide access to public and other records where the Statistician General/ the Chief Government Statistician is of the opinion that, the collection of statistics relating to any matter may be obtained from any Government institution, agency, or user or producer of statistics, he/she shall grant access to any authorized officer or staff of NSOs for the purposes of getting the information required.

A legal framework for administrative data addressing confidentiality issues is crucial for ensuring the protection of sensitive information collected by government agencies and other entities. Confidentiality concerns often arise when dealing with personal or sensitive data, such as health records, financial information, or other Personally Identifiable Information (PII).

The general policies on data sharing within government bodies, which will influence the right of access to administrative data for statistical purposes can be developed. These policies are developed to ensure that data producers and users follow the accepted standards for the long-term preservation and dissemination of data to the wider public. In particular, standards should be supported that aim to facilitate data harmonization and exchange across different stages of the administrative data production and across institutions including common data structure definitions and code lists, and the integration of data flows and processes within NSS.

4.5.1 Guiding Principles

i. Make dissemination and data preservation plan early in the production of administrative data that includes archiving, publishing and distribution. Verify and ensure that the released data after all the processing steps are consistent with the source data. In the case of the derived variables, it means that one should be able to reproduce the same results from the source data.

ii. Preserve sustainable copies of all key data and documentation files produced during the data collection process, as well as those files made available for secondary analyses. Consider;

- To define the long-term preservation standards and protocols used.
- To maintain older versions of important data and documentation files so that users can follow the changes made from one version to the next.
- Archiving collections in one archive which would keep master copies of files in several locations but minimize the possibility of conflicting versions of data and documentation files.

iii. Conduct a disclosure analysis to protect respondent confidentiality. The key goal of disclosure risk analysis is to ensure that the data maintain the greatest potential usefulness while simultaneously, offering the strongest possible protection to the confidentiality of the individual respondents.

iv. Think about the production of both public and restricted use of data files. Considering the following:

- Make data files fully available to the users by establishing clear rules under which users can obtain the data.
- Establish clear policies for how users may access the restricted data files by creating a set of application materials and restricted-use data agreement that specify how users can obtain and use such data.
- In order to provide optimal utility for the users, produce a variety of products for varied constituencies;
- Produce set-up files and ready to use portable files if applicable in software packages to address the needs of those who seek to do intensive statistical analyses.
- Consider disseminating data on websites.

4.6 Agreement and sharing of administrative data

The assessment carried out by NSOs in 2024 showed that many institutions use the policy as a guiding document for sharing data and not MoU. Given the legal and policy frameworks required to permit the use of administrative data, written agreements (MoU) are often necessary to detail and facilitate the transfer of knowledge and data (UN, 2011). MoU will be applicable only if the law or policy of respective institutions do not offer a window for sharing of its data.

Data sharing among agencies, refers to those methods whereby agencies can obtain access to one another's data on a timely basis. The collection process is duplicate if different agencies collect similar data on the same source. Therefore, when the institutions gather data together and share such information, it will help to reduce this duplication and hence unnecessary costs. Although data sharing has many benefits, it raises issues regarding to 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.

Beyond the act of sharing itself, data sharing entails a commitment to maintaining the integrity and reliability of the shared data throughout its lifecycle. This means not only making data accessible to all stakeholders but also ensuring that it retains its quality, coherence, and usefulness for the processing and analysis by data users. A crucial part of this process involves data producers carefully documenting and labelling sets of data, including providing detailed descriptions and clear definitions so that others can easily find, discover, and understand the shared data.

In addition, data sharing implies making data accessible to the relevant individuals, domains, or organizations using robust access controls and permissions. This ensures that only authorized personnel can access specific data sets, thus adhering to regulatory compliance demands and mitigating risks associated with breaches and data misuse.

4.6.1 Internal vs. External data sharing

In the landscape of modern business operations, NSS must distinguish between internal and external data sharing, with their different approaches for organizations to disseminate information. **Internal data sharing** is all about the exchange of

information within the confines of an institution while **External data sharing** extends beyond the institution's boundaries to include partners, clients, suppliers, and regulatory bodies. Given its nature, external data sharing is subject to stricter regulatory compliance and security measures, necessitating robust protocols to protect sensitive information and maintain trust between the organization and its external stakeholders.

4.6.2 Key features to consider during data sharing

i) Compliance requirements

Institution should maintain data privacy and compliance with regulatory requirements. As agency, sharing of data with external partners or stakeholders, they must navigate complex privacy laws and regulations governing the collection, storage, and sharing of personal or sensitive information.

ii) Ensuring data security

With cyberattacks becoming more sophisticated, ensuring data security both during transmission and when stored, has taken centre stage. Institution must ensure safeguarding sensitive information from unauthorized access, data breaches, and potential leaks. Implementing robust encryption, firewalls, and intrusion detection systems is paramount, but these measures also come with costs and require constant updating in the face of evolving threats.

iii) Maintaining data privacy

Privacy concerns, the organization should be adhering to data protection regulations while ensuring they don't infringe upon user privacy.

iv) Managing data volume

The volume of data generated today poses storage, transmission, and processing challenges. Deciding what data is essential for sharing, ensuring timely data transfer, and making sense of vast datasets require advanced storage solutions, efficient data pipelines, and powerful analytics tools.

v) Ensuring data accuracy and integrity

Sharing inaccurate or corrupted data can have dire consequences, especially when it informs critical business decisions. Establishing rigorous data validation and

cleansing processes is crucial, but it also demands resources and introduces potential bottlenecks in data flow.

vi) Addressing trust concerns

Sharing data, especially with external partners, requires a foundation of trust. Concerns about data misuse, intellectual property theft, or gaining a competitive edge can inhibit data-sharing initiatives. Building transparent data government and clear terms of engagement are vital to fostering this trust.

4.6.3 Agreement between institutions

It is important to have formal agreements in place, which can take the form of MoUs between the relevant institutions. These agreements can help ensure that the collaboration is more institutionalized and less dependent on individuals, and clarify the roles and responsibilities within a concrete time frame.

4.6.4 Characteristics of a successful MoU

i. The institution must have appropriate policies and procedures in place to support data sharing activities. Secondly, the working staff of the institution should be supportive to the project's goals, results oriented, experienced with the data, and able to work cooperatively with people in their agency and the partner agency.
ii. There must be a mutual interest in order to reach a successful conclusion in a way that data sharing arrangements must benefit among agencies.
iii. The goals should be stated narrowly, down to the level of which particular fields in a database will be shared between agencies; and being flexible as it can change according to situation.
iv. Cross-agency data sharing projects can require significant involvement by agency leadership.

CHAPTER FIVE: COORDINATION OF ADMINISTRATIVE DATA PRODUCTION

5.0 Levels of Coordination

In performing the coordination among other functions, the NBS and OCGS shall set standard for data collection, analysis and publication of statistics to ensure uniformity in quality, adequacy of coverage and reliability of statistical information; and promote cooperation and rationalization among users and producers of statistics so as to avoid duplication of efforts and ensure optimal utilization of scarce resources.

Coordination of production of administrative data is essential to strengthen the production of quality of data, dissemination and use. The mandate of coordinating all producers of administrative data and other official statistics was given to the NBS and OCGS but within the NSS a coordination unit can be established and given the following responsibilities

i. Coordinate and implement statistical programs and represent the organization in national and international statistical meetings and bilateral and multilateral consultations whenever necessary;
ii. Working systematically towards achieving national and international agreements about common concepts, classifications, standards and methods;
iii. Coordinating technical cooperation and capacity building activities with national and international partners to avoid duplication of effort and to encourage complementarities and synergies;
iv. Establishing internal coordination mechanisms, which facilitate the discussion of responsibilities, methodologies, concepts and common standards;
v. Ensuring official statistics cover all important subject;
vi. Enhancing the image of official statistics through branding and common release; and
vii. Ensuring that the information needs are incorporated efficiently into the national program to reduce the need for additional statistical surveys to be carried out.

In addition, the Administrative Data Steering Committee consisting of the heads of administrative data producing agencies and data users that meets regularly can be formulated to provide guidance on the production of administrative data in the respective NSS. The flexibility of such a steering committee is essential, and therefore detail functional and governance modalities should not be ruled in a legal text. Apart from Steering committee, the Technical Working Group who will be the champions in ensuring production of administrative data should be established.

5.1 Financing Administrative Statistics Production

The budgetary provisions may affect the process for budget allocation to the NSS producing administrative data. The unit responsible for production of statistics within NSS should be directly involved in the planning and budgetary process. The production of administrative data should be considered in the organization strategic plan and budget allocation within the NSS in every financial year.

However, implementation of administrative data system may be supported through extra budget which comes in the form of development. This budget, unlike the usual recurrent budget, aims at investment activities such as human resources capacity building, procurement of equipment, designing and setting up information systems and compliment the undertaking of periodic large national surveys.

5.2 Monitoring the use of the Manual

As part of their coordination role, NSOs should monitor on annual basis the use of the manual to ensure that quality of the data collected adheres to the guidelines provided. This will help produce high quality administrative data, leading to the production of official statistical indicators. To achieve this, NSOs should:

i. Prepare a quality assessment checklist to verify that the statistical production processes follow the provided guidance;
ii. Provide awareness training to NSS on the proper use of this manual
iii. Utilize the TSMP II Sector Working Groups meetings through the planning, monitoring, and evaluation unit to follow up on NSS statistical operations and ensure adherence to the guidance.

Additionally, the coordination or planning unit within the NSS can conduct monitoring to ensure the manual is used correctly.

ANNEXES

Annex 1

Key requirements for data sharing 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.

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:

i. Provision of metadata: It is important that data flows be accompanied by the relevant metadata, which may include definition, computation, units, dates, and descriptions for any codes etc.

ii. 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.

iii. Period of agreement. Agreements will normally be for a fixed period, but should include provisions for renewal or extension if necessary.

iv. 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.

v. 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.

Annex 2

Sample of MoU

No.	DATA SHARING AGREEMENT	
1.	**Organizations**	This Data Sharing Agreement is entered between **The National Bureau of Statistics (NBS)/ Office of the Chief Government Statistician (OCGS)** and ……
2.	**Purpose of the Agreement**	This data sharing agreement between The National Bureau of Statistics/ Office of the Chief Government Statistician and ……is necessary to ensure the availability of data necessary to create……..indicators in ……..
3.	**Project Description**	The ……is a joint effort between the National Bureau of Statistics (NBS) and Office of the Chief Government Statistician (OCGS)……
4.	**Duration of Agreement**	This Agreement commences on**……..,** and will end on**……,** unless terminated/extended by mutual agreement of both parties.
5.	**Data Custodian responsibility and data stewardship**	The ……will remain the primary owner of the data. The National Bureau of Statistics will store this data in a consolidated database that will house data from other ministries, agencies, and departments (MDAs) whose systems are also integrated into the ……
6.	**Permissible data use, linking and sharing under this agreement**	The National Bureau of Statistics/ Office of the Chief Government Statistician has the mandate to communicate official statistical information to the public. Therefore, NBS/OCGS will only share public data.
7.	**Timing and frequency of updates.**	Depending on the frequency of collection and the process from preparation to dissemination; The National Bureau of Statistics/ Office of the Chief Government Statistician and ……will decide accordingly.
8.	**Constraints on Use of Data**	Data supplied by the**……..**is the property of the……..and shall not be shared with third parties without the written permission of the…… **The National Bureau of Statistics (NBS)/ Office of the Chief Government Statistician (OCGS)** shall not sell or use, internally or externally, for any purpose not directly related to the scope of work defined in this agreement without the written permission of the ……..
9.	**Indemnification**	In the case of legal claims against any of the parties to the agreement, normal legal rules and principles will apply. If

No.	DATA SHARING AGREEMENT	
		one party becomes aware of a claim against the other, they should inform the other party in a timely manner.
10.	**Termination and Modification**	This agreement may be modified if one or both parties request a modification in 90 days prior to modification or termination.

By the signatures of their duly authorized representative below, **The National Bureau of Statistics (NBS)** and the **……**intending to be legally bound, agree to all the provisions of this Data Sharing Agreement.

For and on behalf of
The National Bureau of Statistics/ Office of the Chief Government Statistician

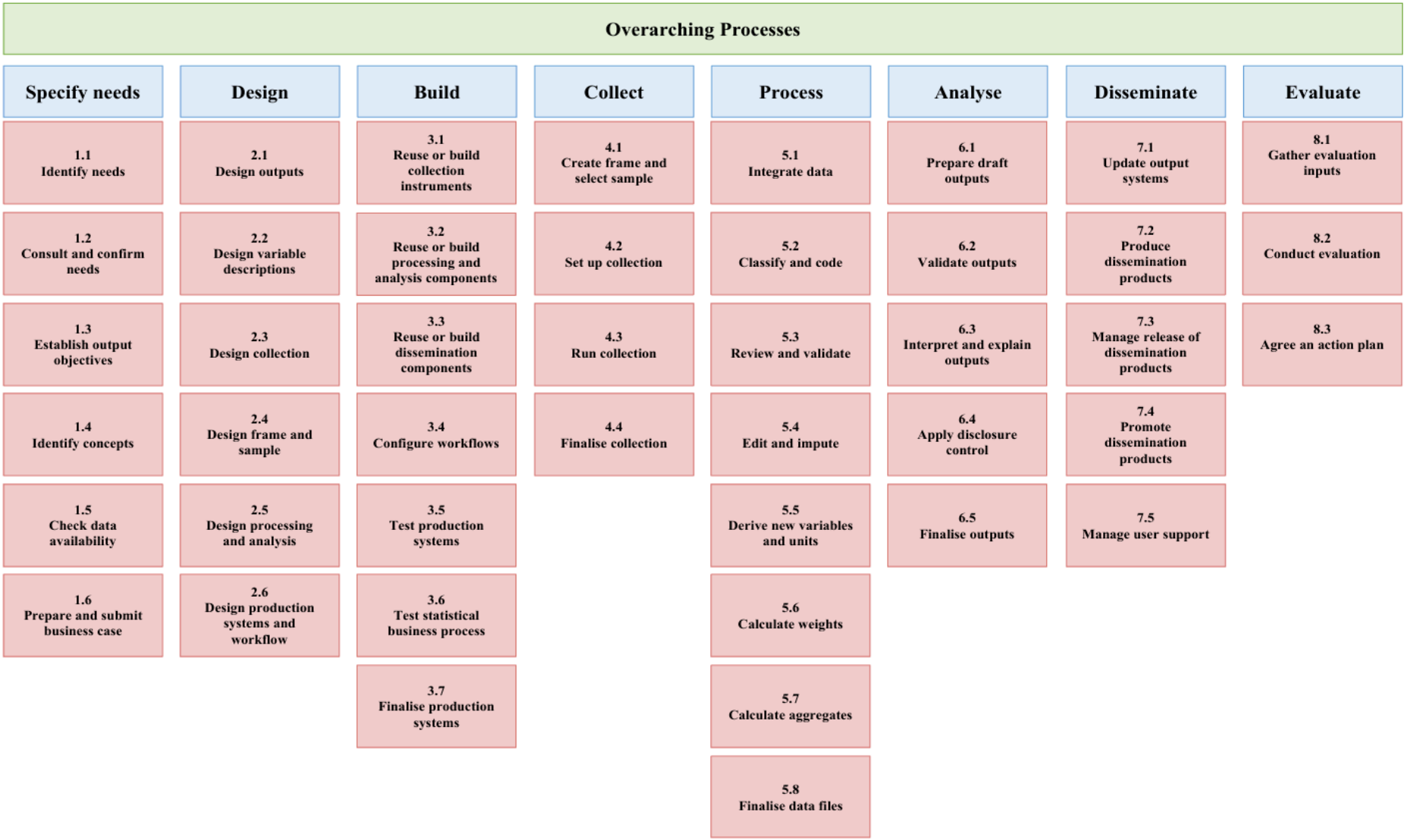
Name: _____
Signature:_____
Position:_____
Date: _____
Place: _____

For and on behalf of
……

Name:_____
Signature:_____
Position:_____
Date: _____
Place: _____

Annex 3

The structure of the Generic Statistics Business Process Model



REFERENCES

1. UN WOMEN and UNSD: Toolkit for Quality Assessment of Administrative Data for Official Statistics, version 01. By Silja Emmel, December 2022
2. Tanzania Statistical Master Plan Phase Two (TSMP II) 2022/23-2026/27, Strategic Document
3. 2022 UN handbook on Management and Organization of National Statistical Systems
4. Statistics Act Cap.351 and its associated Statistics Regulations of 2017
5. The Statistics Act No. 9 of 2007 and its associated Statistics Regulations of 2018 (Zanzibar)
6. United Nations, (2019), National Quality Assurance Frameworks Manual for Official Statistics
7. United Nations - Department of Economic and Social Affairs Statistics