
INSTITUTE OF COMMONWEALTH STUDIES
ICwS
SCHOOL OF ADVANCED STUDY UNIVERSITY OF LONDON
A Matter of Trust
Records as the foundation for building integrity and accountability into data and statistics to support the UN Sustainable Development Goals
Concepts, issues and potential strategies
Arts & Humanities Research Council
sas.ac.uk

A Matter of Trust

Records as the foundation for building integrity and accountability into data and statistics to support the UN Sustainable Development Goals

Concepts, issues and potential strategies

Instiutute of Commonwealth Studies
School of Advanced Study
University of London
Senate House
Malet Street
London WC1E 7HU

http://commonwealth.sas.ac.uk

First published 2018

ISBN 978-1-912250-17-2

Cover image: spainter_vfx/shutterstock.com

This report emerges from a UK Arts and Humanities Networking Grant, 'Digital Records as Evidence to Underpin Global Development Goals' (AH/P006205/1) based at the Institute of Commonwealth Studies (ICwS). The Institute would like to thank the AHRC for this support and acknowledge in particular the work of John McDonald (former Director of the Information Management Standards and Practices Division of the National Archives of Canada) and Anne Thurston (Director, International Records Management Trust) who played a central role in drafting the report. We would like to thank all the participants in the project workshops for their contributions, which this report seeks to capture, and in particular Justus Wamukoya, our lead international collaborator from the School of Information Sciences at Moi University, Kenya.

The situation analysis

In September 2016, 193 United Nations member countries agreed the Sustainable Development Goals (SDGs) to help guide national and global development policies in the period to 2030. These cover areas such as climate change, economic inequality, sustainable development, peace and justice. Each of the 17 goals is to be assessed based on agreed targets and indicators.[1] These targets are global in nature and universally applicable, taking into account divergent realities, capacities and levels of development and respecting national policies and priorities.

Member countries, supported by organisations such as the UN Statistics Division, are responsible for collecting and processing the data and generating the statistics needed to measure the indicators. National statistical agencies are expected to serve as data hubs and focal points for producing government statistics. However, problems with collecting reliable data at the national level often mean that countries have to rely on data collected by international agencies, which can lack the granularity needed to respond to local needs.

Each goal presents a considerable challenge in terms of collecting and analysing relevant data and producing the statistics needed to measure progress. Most governments in lower resourced countries haven't yet developed the controls needed to produce high quality, reliable data.[2] Data collection processes tend to be fragmented and poorly administered because those responsible are too few in number and lack the expertise to introduce the necessary policies, standards and procedures and accountability structures, making it difficult to rely on the data to make decisions and set direction.

The implications are hugely significant. The inability to measure progress because of inadequate, inaccurate, incomplete, inaccessible or flawed statistics can lead to misguided decisions and is likely to undermine the achievement of the SDGs. Failure to 'get the statistics right' can result in wrong decisions being made, wrong strategies being adopted and wrong laws, policies and standards being established. It also can lead to a significant waste of resources.

Fortunately, the quality of the statistics used to measure the achievement of the goals has been recognised as an issue by a variety of global organisations, including the Sustainable Development Solutions Network and the Global Partnership for Sustainable Development Data.[3] These organisations recognise that the quality of the statistics depends upon the quality of the data used to produce the statistics. These, in turn, depend upon the quality of the processes supporting the collection, manipulation and analysis of the data and the production of the statistics.[4] There is also growing recognition that the quality of these processes depends on the availability, completeness and integrity of the records documenting the processes.

Without records to provide evidence of how the data was gathered and analysed or how the statistics were produced and disseminated, it is not possible to confirm that the statistics used to measure the SDG indicators are complete, accurate, relevant and meaningful. The quality and integrity of data and statistics depend on the quality and integrity of the supporting processes, which, in turn, depend on the quality and integrity of the supporting records. Failure to address the records issue adequately undermines efforts to address the quality of statistics and data. For instance, the accuracy of health data depends on reliable and complete patient records. Access to justice cannot be measured if transparent and effective court processes are not in place and if case records are not well-kept and readily available.

In order to explore the records issue and its implications, a UK Arts and Humanities Research Council project was set up toward the end of 2016 at the University of London's Institute of Commonwealth Studies (AHRC Networking Grant, AH/P006205/1). Under the banner Digital Records as Evidence to Underpin Global Development Goals, workshops were held in April 2017 and May 2018 to explore the relationship between data, statistics and records as the primary types of information for measuring goals and to initiate an interdisciplinary dialogue among humanities scholars, development experts and information professionals, including statisticians, data experts and records management professionals.

The participants recognised the need to reach beyond the worlds of data and statistics to incorporate the role that records play in enabling member countries to prove the integrity not only of the data and statistics but also of the processes used to collect and analyse them. The worlds of statistics, data and records can be quite different, but viewing them as part of a whole helps to ensure the quality and integrity of the information for all who need it. The participants also recognised that the quality, completeness and integrity of records can be difficult to achieve without effective policies, procedures, standards, systems and qualified records management expertise. Among the specific points raised in the discussions were the following:

- Census data in many countries tends to be limited, out of date, inaccurate, irretrievable or simply has not survived; undocumented changes in data structures and data entry errors make it difficult to compare data through time.
- Governments are not able to make informed decisions where the existing data Is weak or inconsistent or where the data they need does not exist.
- There is often a lack of clarity about where data has come from. How was it compiled? Why was it compiled? What was

1 There are 169 targets, setting out quantitative and qualitative objectives for the next 15 years. The global indicator framework developed by the Inter Agency and Expert Group on SDG Indicators (IAEA-SDGs) was agreed by the UN Statistical Commission in June 2017. It is supported by the SDG database dissemination platform, maintained by the UN Statistics Division, which provides a metadata repository containing the latest information available about the indicators.

2 Jerven (2013)

3 With the approval of the global indicator framework, the Inter Agency and Expert Group on Sustainable Development Goal Indicators has formed three working groups to address specific areas relevant to SDG indicator implementation: statistical data and metadata exchange, geo-spatial information and interlinkages. In November 2017, the Inter Agency and Expert Group produced a consultation draft on Guidelines and Best Practices on Data Flows and Global Data Reporting for Sustainable Development Goals, which highlights many of the challenges that UN member countries face in producing the high-quality statistics required to measure the SDG indicators.

4 There is an illustration of the interrelationships among data, statistics and records in the context of a 'process' supported by a framework of policies, standards, technologies, people and governance in Appendix A.

the sample set? How was it amalgamated? What methods were used to analyse and interrogate it? What algorithms were used to enable its interpretation? Who published the data and when? When multiple data sets are combined, who owns the amalgamated data set? Has the process been carried out transparently? How effectively has the data been anonymised?

- Mapping together data structures from multiple data sets, many of them with broken data linkages, to arrive at reliable statistical findings, can be a challenging and complex process.
- Many members of the data community are self-trained and may not follow rigorous data science methodologies.
- Partial metadata, opaque provenance and undocumented chains of custody raise doubts about the identity and integrity of the data, particularly during aggregation, as does the lack of information about the systems for managing the data.
- IT systems create records but often lack the full functionality needed to keep them reliable and authentic for as long as they are needed. Too often, they have been developed without the supporting framework of policies and systems needed to protect, preserve and make digital evidence available through time.
- Digital records are fragile. Digital media deteriorate, software changes and hardware becomes obsolete. Digital records may be stored on personal drives, un-networked computers, unmanaged network drives or mobile devices, which can make them unavailable as a national resource and unlikely to survive. They can be altered, deleted, fragmented or corrupted through malicious interference or inadequate management; their meaning may be lost when metadata is not captured, is imprecise or becomes separated from the records when technology changes. They can be difficult to retrieve after a few years, months or even days.
- Data drawn from poorly kept records, particularly where the processes documenting how they were created are poorly documented, can contribute to skewed findings and misguided policy recommendations and misplaced funding.
- In some countries, where digitisation has been introduced, it has been discovered that the records are in such disarray that records management teams have to be brought in to organise the records before they can be digitised.
- If records cannot be accessed, if they are lost or corrupted through time, or if they are so poorly kept that the cost of locating them is prohibitive, access to information laws will have little impact. The true cost is the relationship between the citizen and the state.
- In many cases, no one knows who is responsible for the selection and long-term preservation of government data, and no one is paying attention.
- Digital information (data, statistics and records) is being lost regularly on a very large scale, particularly in lower resource countries where the structures needed to protect and preserve it have not been developed.

Following the workshops, it was decided that it would be important to establish a shared understanding of the terminology and concepts among the disciplines involved in creating and managing digital information and that it would be valuable to illustrate the situation through a fictional situation report based on a fictional government in a lower resourced country. This would make it possible to present key issues without identifying individual countries. It also would allow countries to assess the extent to which the fictional situation matched their own reality, with a close match suggesting greater relevance. Overall it was hoped that the illustration would increase awareness and understanding of issues and their consequences, leading to concrete action.

A fictional situation analysis, in the form of a consultant's report, was developed drawing on experience in lower resourced countries in Africa, a region where issues such as poverty, climate change and the availability of clean water have life threatening consequences and where a significant number of countries are participating in the SDGs. Its specific objectives were to clarify some of the basic concepts around statistics, data and records, to explore key questions around their generation, organisation and reliability, and to contribute to the development of strategies for quality control now being developed by organisations such as the Sustainable Development Solutions Network and the Global Partnership for Sustainable Development Data.

The hypothetical consultant's report that follows describes a situation analysis conducted for the government of a fictional country called Amania. According to the report, the idea of the situation analysis grew out of the government's realisation that if it were to meet the targets set for the SDGs, it needed to assess its capacity to generate and process complete, and accurate, data, statistics and records and to be able to prove their integrity.

The situation analysis begins by explaining the background to the study, its objectives, how it was undertaken and the definitions of key terms used. The results are reported at two levels: in individual ministries and at the National Bureau of Statistics, which coordinates the preparation of statistics provided to the Ministry of Government Affairs for transmission to the UN Statistics Department. Specific issues are identified at both levels, but the report focuses primarily on broader issues related to the framework of policies, standards, systems and people needed to manage data, statistics and records. Managing records is emphasised because of the widespread lack of awareness of the importance of the role records play in supporting the quality and integrity of the data and statistics.

The situation analysis is an awareness-setting document, not a guide. However, it concludes with practical information on potential strategies for strengthening quality controls and a set of maturity levels to guide the systematic implementation of the strategies. The closer the analysis comes to reflecting reality, the more relevant the strategies will be for strengthening the quality and integrity of the data, statistics and records used to measure the SDG indicators.

It is important to recognise that the issues addressed in the analysis have consequences for all countries working towards achieving the SDGs. The quality and integrity of data, statistics and records used to measure the goals are global issues requiring a global response. Hopefully, this illustration will contribute to awareness and communication among the disciplines that need to work together to develop that global response by ensuring that the data, statistics and records used to measure the achievement of the SDGs are accurate, relevant and trustworthy.

The quality of data, statistics and records used to measure progress towards achieving the United Nations Sustainable Development Goals in Amania

A situation analysis

JLM Information Management Consulting Inc.

A report produced for
the Ministry of Government Affairs
Government of Amania
October 2018

Contents

Introduction

Background

Amania is one of 193 countries that has signed on to the Sustainable Development Goals (SDGs) initiative led by the United Nations. Within the government, the Ministry of Government Affairs is responsible for coordinating the government's commitments under the initiative and for submitting the statistics to the UN Statistics Department. These statistics, generated by the Amania National Bureau of Statistics and based on its own in-house surveys, as well as on surveys and other data sources supported by individual ministries, are being used to measure the SDG indicators.

The government's efforts to implement the Millennium Development Goals (MDGs) revealed significant weaknesses in the data and statistics used to measure their achievement. In a number of cases, it was found that the statistics produced to measure the indicators were flawed, often because the data used to generate them was also flawed. In some cases it was possible to identify where and why the data was flawed, but in too many others it was impossible because there was a lack of supporting documentation. The records that should have documented the processes tended to be fragmented or missing altogether. It was embarrassing to the government when it was discovered that certain MDGs had not been achieved or when it was clear that the government's statistics could not be trusted. This caused several development partner organisations and private sector investors to question whether to trust the statistics when deciding on the level of donor support to provide.

In order to avoid the same issues emerging in the SDG initiative, the Ministry of Government Affairs commissioned this situation analysis to assess the quality, completeness and integrity of the data and statistics used to measure the SDG indicators. The government recognised that while issues associated with the quality and integrity of data and statistics were becoming understood, the role of records was poorly defined. This helped shape the scope of the study, the analysis of issues and the development of suggested strategies. The following key issues were identified:

- The quality and integrity of **statistics** are based on the quality and integrity of the data input.
- The quality and integrity of **data** input to the statistics relies on the quality and integrity of the processes for collecting, processing, analysing and maintaining the data, as well as on the processes for producing and reporting the statistics.
- The quality and integrity of the **processes** can be demonstrated by complete, authentic and accurate records of sufficient quality and integrity to provide evidence of decisions and actions supporting the processes.

The Ministry of Government Affairs retained an information management expert to undertake a situation analysis. In addition to improving the quality and integrity of the data and statistics used to measure the SDGs, the government expects that the analysis will also improve the data, statistics and records that support operational and administrative programmes as well as transparency, openness and accountability for citizens.

Organisation of the report

The report describes the results of the situation analysis. It defines the methodology for the study and the terms used, and it analyses the quality and integrity of the processes followed in collecting and manipulating data and producing statistics used to measure the SDG indicators, with an emphasis on the quality and integrity of the records that document the processes. The implications for achieving the SDGs are highlighted at the levels of the ministries and of the National Statistics Bureau.

The report then goes on to suggest strategies for addressing the issues that have been identified by establishing a framework of policies, standards, systems and people supported by an effective management structure. The last section describes a series of maturity levels to help the government establish milestones for implementing the strategies, and it recommends immediate first steps.

Methodology

The consultant's activities focused on:

- Conducting research to identify current relevant initiatives underway at the international level.
- Conducting interviews and reviewing documentation to identify and describe the work processes and management frameworks for collecting, analysing and presenting data and statistics needed to measure the SDG indicators.[5]
- Conducting interviews and reviewing documentation to identify and describe the characteristics of the records (correspondence, documents, completed forms, data files, logs and so on) needed to collect data and produce statistics as well as the supporting management frameworks.
- Analysing and assessing the level of authenticity, completeness, accuracy and integrity of data, statistics and records documenting the supporting processes.

5 Information for this step was derived from *SDG Indicators and Metadata Repository*, United Nations, 2017, https://unstats.un.org/sdgs/metadata/.

- Identifying areas where the authenticity, completeness, accuracy and integrity of the data, statistics and records are at risk and why.
- Preparing an overview of the implications on risks to the government's ability to deliver Its SDG commitments.
- Proposing strategies to address the issues and reduce or eliminate risks.
- Defining a road map describing the way forward based on a set of progressively more sophisticated maturity levels.
- Providing a set of immediate next steps that can be taken to strengthen the management of data, statistics and records.

Definitions

When government officials were interviewed, including those responsible for data, statistics, records management, IT, audit and programme management, it was clear that their understanding of basic concepts and processes differed. For some, the concept of data embraced all recorded information, regardless of physical form, from information recorded on paper or electronically, to the highly structured information recorded digitally on computer-readable media. For these individuals, records in digital form, including email and reports, contained data that could be manipulated and exploited just as readily as the data recorded in highly structured computer-based data files. Records were just another form of data.

Others made a clear distinction between records and data. According to them, data is highly structured codified information recorded in computer-readable form for processing and manipulation by computers. They saw records, whether in paper or electronic form, as information recorded with the primary purpose of documenting actions and decisions and serving as evidence to meet various accountability requirements. Records, for these individuals, were static, never-changing documents, rather than data that could be manipulated. Some were even more focused in their views, believing that records are the paper files that they use, while data is what is held in databases that IT people use.

Given this range of views, it was decided to use definitions that reflected a balance, respected the scope and objectives of the study and were, as far as possible, based on authoritative sources. The following definitions resulted from considerable discussion among representatives from the various disciplines.

- **Data** refers to a set of values of qualitative or quantitative variables (recorded in multiple physical forms) generated, manipulated and analysed to support the production of statistics.
- **Statistics** are the result of manipulating and analysing data. They are a type of data. For the purposes of the study, they are the instruments used to measure the SDG indicators.[6]
- **Records** refers to recorded information produced or received in the initiation, conduct and/ or completion of activities and that documents those activities. They are a special form of recorded information or data. When well-managed, they comprise content, context and structure sufficient to provide evidence of the activities.[7] Records are not simply correspondence or other documents generated to oversee management of the activity. They include all forms of recorded information, including data and statistics, that can serve to document the activity.
- **Metadata** refers to data that provide context for data and statistics used to measure the SDG indicators and the supporting processes. It is also an important attribute of the records that document an SDG activity, such as the conduct of a survey, the analysis of data or the production of statistics. Metadata describes the relationships among the records, providing a documentary trail of the activity, and placing the records in the context of their creation, management and use. In short, metadata makes it possible for the information in data, statistics and records to be understood, verified and used in context.
- **Process** refers to a collection of related, structured steps or tasks needed to achieve a specific service, product or goal.[8] For the purposes of this study, it refers to the structured steps or tasks involved in collecting, processing and manipulating data to produce the statistics that are used to measure SDG indicators. This includes, for instance, the steps involved in planning and approving the survey, designing and testing the survey methodology, conducting the survey, collecting the data, processing and analysing the data, producing and reporting on the statistics, and performing an evaluation of the entire exercise. Data, statistics and records are generated continuously throughout the process.
- **Records management** is the management function responsible for efficient and systematic control of the creation, receipt, maintenance, use and disposition of records.[9] It enables ongoing capture and continued accessibility of high-quality, authentic, reliable, accurate, complete, relevant and timely records. This includes data files which, as part of a given documentary trail, must share these characteristics if they are to be trusted.

6 According to the Inter Agency and Expert Group on Sustainable Development Goal Indicators, 'official data' refers to a set of values of qualitative or quantitative variables, which are produced and/or disseminated by an official source such as the National Statistical Office or another governmental agency or department including non-traditional types of data. 'Official statistics' means a part of official data, which is produced in compliance with Fundamental Principles of Official Statistics. See: *Guidelines and Best Practices on Data Flows and Global Data Reporting for Sustainable Development Goals*, 9 November 2017, page 4 (https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-06/20171108_Draft%20Guidelines%20and%20Best%20Practices%20for%20Global%20SDG%20Data%20Reporting.pdf.)

7 Derived from definitions provided by the International Council on Archives, https://www.ica.org/en.

8 Derived from Wikipedia, https://en.wikipedia.org/wiki/Business_process.

9 For a complete definition, see ISO Standard 15489, *Records Management*, which states that: 'records management is the field of management responsible for the efficient and systematic control of the creation, receipt, maintenance, use and disposition of records, including processes for capturing and maintaining evidence of and information about business activities and transactions in the form of records'.

- **Records management framework** refers to the policies, standards and practices, systems and technologies, and governance structures for managing records. Just as policy frameworks govern the management of personnel, finances and security, a records management framework must be based on a government-wide policy. The records management framework is designed such that records in all of their different forms can play multiple roles.
  - **Records** provide evidence of how a ministry or a person conducted business including decisions, actions, non-decisions and inactions.[10] For instance, **records could be used to prove that a statistical survey was deliberately manipulated to show a favourable outcome**.
  - Records enable organisations to hold themselves accountable in relation to laws and practices. For instance, records documenting the process for extracting statistical data from a government database could be requested under an access to information law; and **the same records could also support an audit of a government programme responsible for extracting the data.**
  - Records support individual rights and freedoms. For instance, **records documenting processes for producing land settlement statistics could be used to locate original survey forms completed by individuals seeking to substantiate their claims**.
  - **Records are the source of qualitative and quantitative data that can be used for multiple purposes beyond those that led to their creation.** For instance, the master data file produced as a result of the annual health survey could be merged with census data and data from other sources to perform analyses not possible using the master data file alone.

10 Derived from State of Queensland – Department of Public Works, Glossary of Archival and Recordkeeping Terms (Queensland: Queensland State Archives, 2010) as described in International Council on Archives Multilingual Terminology, https://www.ica.org/en/online-resource-centre/multilingual-archival-terminology.

Analysis

Introduction

This section describes issues that present high risks for the quality of data, statistics and records, based on an analysis of:

- **processes** that support the collection and analysis of data used to produce statistics for measuring the SDG indicators and for disseminating the statistics themselves;
- **data** and **statistics** generated by the processes; and
- **records** documenting the processes, the data and the statistics.

The government of Amania and the SDGs

Ministries are required to submit final versions of data files and statistics to the National Bureau of Statistics as the basis for measuring specific SDG indicators.

The Bureau, which serves as a coordinating hub, incorporates the data and statistics in reports that it submits to the Ministry of Government Affairs. It also collects, verifies, analyses and produces its own data and statistics. In some cases, it merges data from a ministry with data from other ministries as well as its own data to produce statistics covering multiple SDG indicators. It may also undertake additional data processing to ensure that statistics reported to the Ministry of Government Affairs are presented in a consistent format. In turn, the Ministry reviews the reports, confirms their acceptance and produces a summary report that it submits to the UN Statistics Division according to a predefined schedule.

Data collection and analysis at the ministry level

Twelve of the 23 government ministries are responsible for collecting data used to measure the SDG indicators. A ministry may be responsible for one or several SDGs, or it may only support one of several indicators associated with a given SDG. The ministries may use one or more of the following methods to collect data and produce statistics.

Survey data

Survey data is collected through longitudinal (repeated observations of the same variables over short or long periods of time) or one-time surveys using questionnaires or interviews. Some are large, such as the census; others are small, such as household surveys. In some cases, ministries have designed entirely new surveys to meet the requirements of a specific SDG.

Surveys are generally designed and administered by research divisions within the ministries. Typically, data is collected on survey forms by mail or distributed by contract survey staff to a sample of a target population. In a few cases, data is submitted online. The data is then transferred to coding sheets and input digitally to a raw data file. Several process files may be produced as the data analysis moves through various stages. Where the data contains personal information, anonymised versions may be created.

The resulting statistics are formatted into tables and embedded in various reports for distribution to a wide range of audiences in paper form or, in a few cases, in digital form via a ministry website. In some cases, especially for longitudinal survey results, the data file might be input on a database of data files from previous surveys. A customised report, together with a copy of the master data file produced as a result of the analysis, is forwarded to the National Bureau of Statistics, which produces a standardised report containing the statistics and submits it to the Ministry of Government Affairs to report on progress towards meeting specific SDGs.

Records documenting decisions and actions relating to planning, designing and conducting surveys may exist in various forms. These include emails, paper-based correspondence and reports about a given survey; process and master files created as a result of the survey; survey documentation, such as completed coding sheets, survey design reports and descriptions of the methodology; and records describing the business context for planning, designing and carrying out the survey. Together, these records, when well-managed, provide evidence that can substantiate the integrity and trustworthiness of the data and statistics used to measure relevant SDG indicators. Examples of these include the proportion of the population living in households with access to basic services (1.4.1) with safely managed drinking water (6.1.1) or in slums, informal settlements or inadequate housing (11.1.1).

Registration and administrative data

This type of data results from administrative activities, such as personnel and finance or operational registration activities, such as licensing. Personnel and finance data tends to be generated in relation to ministry-wide standards and procedures or to workflows associated with hiring and retaining staff, processing expenditure and preparing budgets. Work processes established for registration activities, such as licensing, vary depending on the process, and most are well defined. For instance, in the case of a typical licensing process, licensing applications are received by the responsible ministry and reviewed for completeness and suitability; applicants are notified of whether or not they are accepted, the information is processed and included in a database, and licenses are issued to applicants.

Records documenting these processes may be in multiple forms. For instance, records documenting a licensing process might include emails, paper-based correspondence and reports about a given application for a license; completed application forms; copies of notifications; completed data conversion forms or logs documenting entry of the data into a database; review and analysis documentation (to verify qualification for a license); copies of approval and notification documents; and reports documenting review and renewal actions. Together these records provide evidence that verifies the integrity and trustworthiness of the data and statistics used to measure relevant SDG indicators. Examples of these would include death rates due to traffic injuries (3.6.1), the proportion of women in managerial positions (5.5.2) and annual growth rate of real GDP per capita (8.1.1).

Scientific data

This includes instrument readings measuring natural or physical phenomena, such as weather (for example temperature, rainfall), geology (for example soil composition, erosion) and hydrology (for example water levels, pollutants). Data generated from instruments is stored within the instrument or transmitted to receivers that store the data separately. In the case of weather data, for instance, readings are made on a regular basis from instruments located around the country. These are transmitted to a satellite, which transmits the data to ground stations supported by computers that automatically convert the readings and merge them in a master database holding not only the readings but also the processed data that underpins weather reports. There is very little human intervention.

To take another example, water acidity measurements are taken by staff and volunteers for the Natural Resources Ministry using instruments that take water quality readings, including acidity levels, on an annual basis in selected areas of the country. These are recorded on coding sheets, submitted to the Ministry, converted to digital form, analysed and used to produce a wide range of statistics including the average marine acidity statistics used to measure SDG indicator 14.3.1. Some of the statistics are also combined with land use data to measure the impact of agricultural land use on the levels of water pollution.

Again, records documenting these processes may exist in multiple forms. From the water quality example, these might include emails, paper-based correspondence and reports about a given water quality activity, completed water quality measurement logs, completed data input forms, data verification logs, extract files (data files created from the master database) and report files describing statistics resulting from analysis of the data. Documentation on planning, designing and operating the water quality measurement process and the programme administering the process and the database all form part of the documentary trail. Examples of relevant SDGs supported by scientific data might include the proportion of agricultural area in productive and sustainable use (2.4.1), average marine acidity (14.3.1) or forest area as a proportion of total land area (15.1.1).

Data and records issues at the ministry level[11]

The findings that follow are based on interviews with selected staff and on-site observations in the 12 ministries responsible for measuring the SDGs. Quotes from some of those interviewed for the study are included to illustrate the practical issues involved in measuring the SDG indicators reliably using official data and statistics. In general, ministry staff tend not to recognise the need to build a documentary trail to support the processes of collecting and processing data and producing statistics. They are often unaware aware of the kinds of records that need to be in place, how the records can be related to one another, where and how they should be organised and stored and how ongoing accessibility should be managed.

In most of the ministries responsible for conducting longitudinal or one-time surveys, data files are not well described; documentation on data structures, coding and formats is fragmented and data verification and quality control procedures are weak or, often, non-existent. Little care has been taken to ensure that a documentary trail is in place to provide evidence of how surveys are designed and conducted, how data is collected and processed and how statistics are produced.

> ***Manager in a research division:***
> *The ministry wants us to document our surveys but I don't know what this means. The minister was worried about a sensitive data file that had errors and we couldn't explain where the errors came from. It's not because of us in the research division. We've tried to follow some data management standards and survey guides we found online. Now the ministry says we need to document things like why the surveys were done and how they were managed. That information is mostly with other people in emails and memos that I don't see. Action officers in other divisions have that information on their desktops.*

11 Information for this section was inspired by Jerven (2013)

Large operational databases in the participating ministries tend to be well managed, but data extracted from the databases to measure SDG indicators is often poorly documented. Records documenting data extraction tend to be fragmented or non-existent, and procedures for managing the data after it is extracted and used are generally poorly defined. In some cases, the lack of metadata makes it hard to understand the relationship between the extracted data and the source data in the database. Without records documenting changes made to the structure of the extracted data or to the definition of key fields, it is often difficult to know to what extent statistics are inaccurate or misleading.

> ***Manager, IT division:***
> *We had a senior managers' meeting and someone said we should store all our data with an outside service bureau that has better storage conditions than we do. Other managers agreed and said storing in the 'cloud' was the answer to storing the government's data. But I think it is too risky. We don't know if it's secure. I think we must keep our data inhouse. Anyway, we would still have data quality issues. Storing outside is not the answer.*

In a number of ministries, staff managing large databases are being asked to generate statistics to support measuring the SDGs. This is a new task, and many staff do not have the expertise needed to document the processes generating the statistics or to ensure quality control.

> ***Database manager:***
> *I was asked by senior management to get statistics from our immigration database. I was told to send them to the Bureau of Statistics because they needed them for the Sustainable Development Goal indicators. I can write a program to extract the data in a report, but I was told I needed to produce the statistics according to industry standards. I don't know what that means. What are the right industry standards? I've talked to other IT managers, but no one seems to know anything about industry standards. We need training.*

Given the need for rigorous standards for collecting and analysing scientific data, the quality of the documentary trail is somewhat better than for administrative and survey data. However, data reliability is undermined by the failure to keep records of changes in the instruments used to make scientific measurements, by changes in sampling methods or by failure to update metadata schema.

> ***Staff member, Environmental Monitoring Division, Environment Ministry:***
> *We monitor marine acidity at stations along the coast and take manual samples. But we don't have trained staff to take samples and equipment has been stolen from some sites and at others it's broken down. How are we supposed to generate good statistics? The ministry still wants us to use the data we have for the SDG indicator 14.3.1 – the one about marine acidity. I've told the minister that our data is not good enough to do the analysis, but he wants us to try anyway.*

Several of the surveys used to measure the SDG indicators lack sufficient documentation about the metadata schema supporting the surveys. There are inadequate definitions of key terms, which has led to confusion when interpreting some of the statistics generated from the surveys.

> ***Official in the Social Development Ministry:***
> *The Labour Ministry uses a different definition of 'employment status' from us. I think we should include more people like part-time street traders and part-time farmers, even children. How can we report statistics for employment if we are using different definitions? I've searched in our files to find out why we use our definition, but I can't find any records. Maybe there aren't any. I've asked the Labour Ministry where their definition comes from, and sent reminders, but no reply.*

Problems in finding, retrieving and understanding data held in older data files for trend analysis purposes are hindering the government's ability to regularly measure SDG indicators.

> ***Official, Ministry of Agriculture:***
> *The chairman of one of our farmers' associations asked us for data on crop production. He wanted the information for his members so that they could look at trends. He also asked the Bureau of Statistics, but we can't find any data files earlier than four years ago. We don't have any record of where the data files are stored and all the staff involved have left.*

Metadata describing the context for many data and statistical files tends to be incomplete. This makes the analysis of trends very difficult and it also makes it difficult to respond to 'access to information' requests or court challenges.

> ***Government lawyer:***
> *NOPA, that's the National Oil Producers' Association, say the government has sent the UN incorrect statistics. This is in connection with Sustainable Development Goal 7 on energy. They've asked the government to provide the documentation on how the statistics were produced but we can't find the records. We've asked the Records Office and the action officers involved but no one can find anything. To be honest, I'm not even sure the methodology was properly documented.*

Many organisational units across the government are involved in developing statistics that support the SDG indicators. Multiple organisations may be involved in developing any given SDG indicator, from the initial planning for a survey or the extraction of data from a database, to the final submission of the statistics to the UN Statistics Division. In many ministries, it is practically impossible to bring together the complete story of measuring an indicator because each unit takes its own approach to capturing and classifying the records documenting its activities.

Official, Ministry of Labour:
We've had a big problem with statistics for the Sustainable Development Goals initiative. We receive data from two other ministries and merge them with our own data to produce the statistics. Now there is an expert looking at how we produce the statistics. We gave him copies of the records we send to the Bureau of Statistics. He says the records are not good enough – the quality of the data can't be trusted. He's right. We can't relate their data to the records we keep so the overall statistics can't be trusted.

When one organisational unit passes data to another unit, if the units take different approaches to capturing and managing records documenting the processes they follow, it can be challenging if not impossible to bring together the complete story of how the indicators are measured.

Assistant secretary, Social Development Ministry:
Our Research And Statistics Division has complained that its statistics have been altered. They sent the statistics to our Communications Division for submission to the Bureau of Statistics but somehow the statistics were changed. The Bureau of Statistics sent the statistics on to the Ministry of Government Affairs. The Ministry was supposed to send the statistics to the UN for Sustainable Development Goal indicator 10.1.1, but that hasn't happened because we don't know how or why the statistics were altered. I'm trying get to the bottom of this, but no one can find any record of why the statistics were changed.

Most ministries have assigned accountability internally for producing statistics to support the SDGs. However, often no one is accountable for classifying records that should document processes for collecting and analysing data and producing statistics or for ensuring that they are complete and accessible through time. Changes in methodology (for instance in the sample size) and in definitions of key concepts (for instance the target object being measured) tend not to be well documented. In a few ministries with long involvement in generating statistics there is documentation on survey methodologies (such as coding schemes and analytical techniques) and on conducting surveys (data verification checks, evaluations and audits) somewhere in the ministry, such as in the library. Even in these cases, however, there is seldom a link to the records, such as emails and correspondence, that document the conduct of the survey itself. As a result, the quality and completeness of the documentary trail for individual surveys and for the survey programme varies considerably.

Most ministries don't have records management programmes. The one exception is the Ministry of Health, which has a small records management unit with responsibility for managing all of the ministry's records and ensuring that they are accessible through time. Unfortunately, the unit does not yet support the Health Statistics Division, which is responsible for generating statistics measuring several SDG indicators. The staff are on their own in managing records documenting their surveys.

Records manager, Ministry of Health:
I only have three staff and none of us have professional qualifications. We have some training, but it only covers paper records. We keep asking for professional training or training in electronic records management. We see other people going for training, but it never comes to us. The ministry thinks we are only here to manage the paper files, but the Health Statistics Division works on data files, and I don't know anything about how to manage them.

The government of Amania does not have a digital preservation strategy. Most IT staff believe that digital preservation means storing data securely but don't understand the importance of managing the metadata that will make it possible to access and understand the data through time. They do not understand the need to convert data to new formats that new software can read or to generate and maintain complete and accurate records documenting these changes. Many look to the National Bureau of Statistics for direction and guidance, and some have suggested that it should become a centre of expertise or even a storage centre for data files with long term value. However, the Bureau lacks the necessary resources and expertise, as does the National Archives. Others think that the digital records will look after themselves:

Head of IT in a large ministry:
I think our archiving strategy is sound; we back everything up on tape.

Interviews in several ministries participating in the SDG indicators process revealed that sometimes the numbers are changed as a result of political pressure before statistics are provided to the National Bureau of Statistics. Although it does not seem to happen often, when it does, it usually isn't recorded. The combination of poor record keeping practices and corrupt actions on the part of government officials has undermined significantly the quality and trustworthiness of the statistics used to measure the SDGs.

Data and records issues at the National Bureau of Statistics

The National Bureau of Statistics maintains a large database that describes the demographic characteristics of the population, including sex, geographic location, income, education, employment status and income level. Much of the data is collected through the surveys managed by the Bureau, with some provided by ministries based on their own surveys. In some cases, the Bureau amalgamates data provided by several ministries to generate statistics on cross-cutting topics. In these cases, data submitted by the ministries is converted to formats and structures that can be matched with specific sets of data from the demographic database and matched with other survey data files.

Regardless of its sources, data held in the National Bureau of Statistics is used to produce statistics that are then incorporated into report files and submitted to the Ministry of Government Affairs before being transmitted to the UN Statistics Department. The reports exist in both digital and hard-copy form. Hard copy reports are held in filing cabinets managed by the administrative

assistant in the office of the director responsible for the demographic database. Digital versions of the report, together with any master data files, are held in the data library 'forever' and managed by the head of the IT area. Copies of the data from the ministries are also maintained in the library but disposed of after five years on the assumption that if the files are needed they can be accessed through the respective ministries.

Statisticians and IT staff in the Bureau of Statistics understand the importance of providing quality data to support the SDGs. However, the lack of resources and of a records-management infrastructure make it difficult to document processes for collecting and processing data and producing statistics. This, in turn, makes it hard to ensure that data and statistics are of high enough quality and integrity to be used effectively.

Records documenting the design of the demographic database and the management of the data (data collection, processing, analysis and reporting) are poorly maintained, and there are no documentation standards.

> ***Staff member, Socio-Economic Statistics Division, National Bureau of Statistics:***
> *Three ministries send us data for indicator 8.1.1. We convert the metadata to a standard format before we merge it with the census data. If we didn't convert the metadata to a common standard, it wouldn't match up. The problem is there are so many differences in the data, like spellings and names, that the statistics we produce are not very reliable. Plus, the ministries are always changing their staff and how they do things.*

The Bureau of Statistics assumes that ministries are submitting data files and statistics of appropriate quality and integrity.

> ***Staff member, Socio-Economic Statistics Division, National Bureau of Statistics:***
> *We sent some incorrect data to the Ministry of Government Affairs for indicator 2.3.1, but it was not our fault. The ministry said we must check the data before sending it, but it's the ministries' responsibility to check their own data. It's not our job. Even if it was, we don't have the documentation to verify the data – I am not sure if even the ministries have the documentation.*

Sometimes the documentary trail is broken when statistical files are transferred from ministries to the National Bureau of Statistics. Each participating ministry uses its own classification standards, which makes it difficult to get a complete picture of any given survey/ data collection and analysis activity. The lack of evidence of the quality and integrity of statistics increases the risk that they could be flawed.

> ***Senior official at Ministry of Government Affairs:***
> *Some months ago the Environment Ministry changed its definition of hazardous waste to make it wider. Then they had to make changes to their surveys and databases, and the way they produce statistics. Now the Environment Ministry has found out that the Bureau of Statistics has not been using the new definition in its reports for the UN. We can't find any records about why the Bureau is not using the new definition and none of the staff can explain it. We don't know how this affects the data used to measure SGD indicator 12.4.2 and other indicators.*

Although the professional staff responsible for the demographic database are concerned about preserving the data, they do not feel equipped to tackle this complex issue.

> ***Manager of the Household Surveys Division:***
> *I've read all of the literature and I think I know how I would go about developing a digital preservation strategy, but I don't have the resources. I have too many other fires to put out and, in any event, the data and its supporting documentation is in a mess.*

There are no formal retention and disposition schedules.

> ***Staff member, the Socio-Economic Statistics Division:***
> *I am worried about our policy for deleting data. The last director general made up rules for how long we keep data in our division. We are supposed to keep anonymised master files and summarised versions for ever. But input and process files must be deleted one year after we create the master files. I don't know why he came up with this idea. I've raised it at management meetings and asked if we can look at it again. If we don't keep the raw data, how can we substantiate our measurement of the SDG indicators, especially over time? It worries me.*

Implications of the failure to establish a framework

The implications for the government's inability to achieve the SDGs are:

- Poorly managed records make it hard to verify the quality and integrity of data generated to measure SDG indicators; this will undermine the government's efforts to report on progress to the UN and jeopardise its ability to make good use of the findings.
- Data can be flawed, but without a reliable documentary trail to reveal the flaws, they can go unrecognised. Without records as evidence, the government will find it difficult to demonstrate the data's integrity or to trace where a flaw occurred.
- Flawed data from one source could skew the statistics provided to the Ministry of Government Affairs, even when the quality of the data from all other sources can be proven by the existence of properly managed records. This can lead to flawed statistics being provided, inadvertently to the UN Statistics Division by the Ministry of Government Affairs.

- The government could waste resources taking action to implement SDG findings based on data that lacks integrity.
- The quality of data collected through time may be eroded as more and more flawed data joins the database. This could have significant consequences for the quality of the data and statistics used to measure SDG indicators in the future.
- The loss of credibility due to flawed data could bring the quality of other data into question, which could be problematic without records to prove the quality of the processes followed.
- In addition to the implications for measuring the SDG indicators and implementing the SDGs themselves, the impact of poor record keeping is likely to affect the government's ability to carry out its mandated responsibilities.
- Individual rights can be compromised if individuals who provided data as part of a data collection activity (such as a survey) cannot be accessed, or if data or records documenting decisions about the collection and use of the data cannot be found.
- National economic interests could be threatened if government policy and direction are based on flawed data and statistics or if the level of quality and integrity cannot be confirmed.

Strategies

This section focuses on strategies for developing a comprehensive and sustainable framework for managing complete, authentic and trustworthy data, statistics and records. Just as there are frameworks for managing human and financial resources, this framework should provide an integrated combination of laws and policies, standards and practices, systems, technologies and people supported by management and governance structures. A focus on symptoms without considering the broader causes, will only result in fragmented and ineffective strategies and short-term temporary solutions. Inevitably, there will be issues needing urgent and immediate attention, but overall the focus should be on establishing the management framework.

The section is organised according to the components of the framework. The key issues and relevant strategies are described below for each component.

Laws and policies

Issues

- There is no law that requires the government to set up a records management programme. The 'access to information' law provides the right of access to a wide range of government records but does not require the government to ensure that its records are authentic, accurate, complete and accessible. The Privacy Act requires that personal information be protected and retention standards be applied, but there is no public pressure for this to be enforced.
- Apart from the Ministry of Health and the Ministry of Government Affairs, none of the ministries participating in measuring the SDGs, including the National Bureau of Statistics, has a records management policy. At the Health Ministry, the policy focuses on managing paper records and does not yet address records in digital form. The policy for the Ministry of Government Affairs is limited to managing the paper records of the secretary and the executive committee.
- There are some policies in place for managing data and statistics and conducting surveys, but they do not address the role of records in providing evidence to document survey and other data collection and processing activities.

Strategies

It is important that existing laws, such as a national archives act, data protection legislation, statistics act or other relevant legislation, should support the effective management of information needed to measure the SDGs, enabling the government to achieve its operational and strategic goals and meet a wide range of accountability requirements. The government needs to:

- Ensure that the Freedom of Information (FOI) law enables citizens to have the right of access to the data, statistics and records generated to support measurement of the SDGs.
- Ensure that the Privacy Act gives citizens the right to access their personal information as recorded in the data, statistics and records generated to support measuring the SDGs.
- Develop a government-wide policy on managing records as evidence that embraces data and statistics as high-quality sources of information for decision making and for verifying the integrity of the processes involved.
- Strengthen policies for managing data and statistics to ensure that responsibility and accountability for documenting relevant processes are clearly defined and that there are provisions for managing data and statistics as part of the documentary trail of surveys and other data collection and analysis activities.
- Develop policies and guidance to protect personal information in relation to the data, statistics and records generated for measuring the SDGs.
- Ensure that in all contracts with private sector firms involving conducting surveys on behalf of the government, the contractor is obliged to document his or her activities, protect the data and statistics it generates, respect the government's ownership of the data and statistics, and transfer all data, statistics and supporting records to the government when the contract is completed.

Standards and practices

Issues

- There is no guidance on how to document processes for collecting and processing data and producing statistics. Ministries establish their own practices for creating and managing documentary trails, which often are not complete,

accurate and authentic. Variations in how these processes are designed and managed makes it difficult to establish standard approaches to documenting them.

- The National Bureau of Statistics follows standards for managing survey documentation, such as code books and survey methodology documentation, but generally these standards and practices are not in place in the ministries. Even when survey documentation standards are applied there is no way to link the documentation to the records, which are often in the form of emails and attachments that document decisions and actions about the management of the survey itself. Establishing a complete and comprehensive documentary record of the survey is impossible.
- There are no procedures in place for converting and sharing data across ministry boundaries. Achieving interoperability when there are multiple recording formats and diverse technologies is virtually impossible. For instance, the National Bureau of Statistics must convert data and statistics it receives from ministries in order to provide statistics to the Ministry of Government Affairs in a standard format. There has been little effort to document these conversion activities, so flaws in the data that emerge at this stage are difficult to trace.
- Retention standards for data and statistics are rarely in place, and even when they have been assigned, they are not consistent across the data, statistics and records associated with a given process. Final statistical data files may be kept 'forever', but records documenting the circumstances of their creation may be destroyed much earlier. Digital preservation presents a huge challenge for any organisation, but it is possible to take preliminary steps, such as researching possible strategies and assessing needs. At present, there is little evidence that this is happening.
- When several ministries are involved, and records documenting a given process exist in multiple forms, it is very difficult to establish a digital preservation plan for all the records associated with the process.

Strategies

- Develop criteria for identifying records that should be in place to document processes for collecting and processing data and producing statistics.
- Develop procedures to ensure that records documenting data and statistics activities are captured, managed and integrated with procedures for conducting surveys, analysing data, merging data and reporting statistics.
- Develop metadata standards and guidance for managing individual processes, linking records, data and statistics, and accessing data, statistics and records within and across processes and different media.
- Establish retention and destruction standards and guidance for all forms of records that document the collection and processing data and the production of statistics.
- Monitor, and draw from, international work on digital preservation strategies and implementation plans for the long-term accessibility and integrity of data, statistics and records.

Systems and technologies

Issues

- Technologies for managing data and statistics are usually specific to the unit responsible and the kinds of data being managed. For instance, the technology for managing data in a database may be different from technologies for extracting data from the database and processing it as statistics to support SDGs. Once the statistics are passed to the National Bureau of Statistics for further processing, other technologies may be used. Documenting the changes that take place from one technology environment to another is a significant challenge.
- Custom-designed databases are in place for managing survey documentation, but technologies have yet to be developed to manage the records of decisions and actions taken regarding surveys or other data collection activities. Nor are there systems for tracking how data is collected and processed and how statistics are produced. Records generated by these activities are not being identified, classified and managed.

Strategies

- Use generally accepted IT project management standards to plan, design, test, implement and maintain systems for managing the authenticity, integrity and continued accessibility of data files, statistics and records across space and through time.
- Use internationally approved standards to develop functional requirements for managing statistics, data files and records and incorporating them into the requirements for designing IT systems.
- Develop audit and evaluation tools for assessing the quality and integrity of data, statistics and records supporting the SDGs; integrate them into standards and practices for systems and into management audits and evaluations.

People

Issues

- The National Bureau of Statistics has some staff with professional expertise in managing data and statistics, but they do not have the records management expertise needed to manage records documenting the processes that generate data and statistics. Few people in the ministries have this expertise either. Records management staff in some ministries, such as the health ministry and the Ministry of Government Affairs, are generally only responsible for paper records. The National Bureau of Statistics recently introduced a training programme for ministry staff responsible for collecting and processing data and producing statistics to support the SDGs. This will help, but at present there are no training materials on managing records in relation to data and statistics.
- In some ministries there is a wide gulf between those responsible for technical aspects of the data (such as IT), and those responsible for processes that generate the data, statistics and records (such as programme managers); often each assumes that the other is looking after the requirement. This gap has serious consequences for the integrity and quality of data and statistics.

Strategies

- Define the work involved in managing data, statistics and records used to support the measurement of SDG goals.
- Define competencies associated with the work.
- Design and implement appropriate training programmes.
- Design and implement appropriate recruitment programmes.
- Enhance tools and techniques for measuring performance so that competencies for managing data, statistics and records can be assessed.
- Establish programmes for allocating staff with the required expertise among ministries to fill competency gaps.
- Establish partnerships, including with organisations outside of the government, in order to pool human and financial resources for developing the framework.
- Work with relevant university programmes to enhance existing courses or develop new ones to address the management of data, statistics and records.

Management and governance

Issues

- Although there are accountability frameworks for managing personnel and finance, none have been introduced for managing records documenting how data is collected and processed and statistics are produced. Accountability has not been assigned for ensuring that a complete and accurate documentary trail is in place. Audit units in ministries measuring the SDGs don't yet cover this issue in management and systems audits.

Strategies

- Establish accountability and assign roles and responsibilities[12] for staff at all levels to ensure the quality and integrity of data and statistics used to measure the SDGs.
- Establish an authority at a senior level of government with responsibility for ensuring that records are managed to support high-quality data and statistics across government.

Awareness

Issues

- Some senior managers are beginning to recognise the importance of preserving data files and statistics, but few understand the crucial role that records play. Records documenting processes by which data files were created and used and documenting the data files themselves (such as coding schemes and storage formats), must be preserved if the data files are to be accessed in the future.

12 It is important to note the difference between accountability and responsibility: accountability is always upward to someone; responsibility is for something (to be done).

- Through time, as the demand for historical data to analyse trends grows, this lack of awareness will have greater implications. The issue needs to be addressed now, rather than in the future, when data files generated early in the SDG initiative may already be inaccessible. The initiative, which asks governments to measure indicators over a 15-year period, is bringing the issue into sharp focus.
- Few citizens are aware of their rights over data collected about them in connection with the SDGs, and few have challenged the way the data is used. Government ministries have not yet felt the pressure to ensure the completeness and accuracy of the data, statistics and records for which they are responsible. However, there is growing citizen concern about these issues and growing awareness of the government's inability to manage the personal information it holds, especially in digital form.

Strategies

- Ensure that senior managers responsible for programmes and processes supporting the measurement of SDGs are aware of key concepts, issues and implications as well as possible strategies.
- Develop tools and techniques for enhancing awareness, for instance briefings and brochures for relevant staff at all levels.
- Incorporate these tools and techniques in training and awareness programmes, including orientation programmes for staff, management seminars and workshops.

The Ministry of Public Administration (which manages the civil service) could be an appropriate agency to take the lead in establishing a framework to address the quality and integrity of the data, statistics and records used to measure the SDGs and, at a more general level, to support the requirements of government programmes for authentic, complete, accurate and relevant data, statistics and records for decision making and accountability.

Implementing the strategies

Capacity levels to guide the way forward

A road map, in the form of capacity levels, will enable the government to move incrementally through defined stages to build the capacity needed to manage data, statistics and records in line with available resources. Five capacity levels are described below, the 5th level being an ideal state for a country that wants to ensure that data, statistics and records used to measure the SDG indicators are of a high enough quality to measure and implement the goals. For most organisations, achieving Level 5 or even Level 4 will be challenging.

The levels reflect diminishing levels of risk, with Level 1 representing the highest risk of loss and inaccuracy and Level 5 being the least risk. They also reflect increasing levels of sophistication in terms of the way data, statistics and records can be used to support implementation of the SDG goals and, more broadly, the government's operational and strategic goals. The roadmap for moving forward will support an objective and systematic approach.

Examples included in the maturity level description are drawn from the targets and indicators supporting SDG Goal 5:

- SDG goal 5: Achieve gender equality and empower all women and girls.
- SDG target 5.5: Ensure women's full and effective participation and equal opportunities for leadership at all levels of decision making in political, economic and public life.
- SDG indicator 5.5.2: Proportion of women in managerial positions.

Level 1 Poor quality data, statistics and records undermine SDG implementation

The organisation produces statistics to measure the SDGs, but they are unreliable. Professionals responsible for data, statistics and records lack the knowledge and skills needed to develop a reliable framework of policies, standards, practices and systems.

> ***Example***
> *Annual labour force data is collected through survey forms sent to companies across the country. Data is collected and analysed, and the resulting statistics serve multiple purposes, including measuring the proportion of women in management positions in support of SDG goal 5. The lack of metadata standards and the absence of records documenting how data was collected and processed and how statistics were produced make it impossible to relate statistics from year to year.*
>
> *Data files from previous years are poorly organised and documented, so records of decisions, including changes in survey design, data processing methods and data formats are fragmented and scattered in multiple locations. The implications will not be known for some time, but without a reliable evidence base or the expertise to prove the trustworthiness of the data, the annual statistics cannot be relied upon as an accurate measure of SDG indicator 5.5.2.*

Level 2 Data, statistics and records enable SDG implementation at a basic level

A framework of laws, policies, standards, procedures and people is in place to ensure that data and statistics are gathered and analysed to measure the SDGs. Managers are generally aware of their responsibility for ensuring that data files and statistics, with their supporting documentation, are stored properly. However, the framework is not applied universally, with some managers transferring poorly documented data and statistics. There are no standards for documenting surveys and other data gathering and analysis activities, nor have policies been developed for managing the records that should document processes for collecting and processing data and producing statistics. Records and data management professionals do not have the expertise needed to manage the interrelationships among data, statistics and records, especially those that need to be preserved through time.

> ***Example***
> *Annual labour force data, including data extracted from the labour force database to produce statistics for measuring SDG indicator 5.5.2, are collected and processed based on approved standards and procedures. However, emails, reports, logs and other records documenting the design and conduct of the survey, including changes in survey methodology, cannot be related to records documenting processes for extracting and analysing data and producing statistics. Records management professionals in the Labour Force Statistics Division that is responsible for measuring SDG 5.5.2 do not have the expertise needed to ensure that records, data and statistics are managed as a whole. The lack of a digital preservation strategy increases the likelihood that trend data needed to measure SDG goal 5 from now until 2030 will not survive.*

Level 3 The quality of data, statistics and records makes it possible to measure SDGs effectively and supports government programme activities

Data, statistics and records generated to measure SDGs are managed through a comprehensive framework of policies, standards and practices, systems and technologies, and qualified people. Records management staff work effectively with data management and other professional staff to ensure that requirements for identifying, describing, classifying, protecting and retaining data, statistics and records are integrated in the design of processes for collecting data and producing and using statistics. Managers know that they are responsible for ensuring that the data, statistics and records generated are authentic, reliable, accessible and understandable and can be retrieved when needed. Professional staff apply clear, consistent standards and practices. However, preservation is not addressed adequately; retention requirements have not been established, metadata standards for data, statistics and records have not been developed, and preservation standards, procedures and technologies are not in place.

Example

All processes for generating statistics to measure SDG goal 5 are supported by the same framework of policies, standards and practices, systems and technologies and people. For instance, data, statistics and records generated to measure SDG indicator 5.2.1 (the proportion of women in management positions) are well described, organised and managed to provide a comprehensive documentary trail of evidence. The statistics can be trusted because the comprehensive management framework itself can be trusted. Unfortunately, the lack of a digital preservation strategy means that while statistics measuring the participation of women in management positions can be compared for the past two years, the government cannot ensure the integrity of the statistics over the 15-year life of the SDG initiative.

Level 4 Well-managed data, statistics and records make it possible to measure SDG implementation effectively and consistently through time; data and statistics are of high enough quality to support government programme activities at the strategic level

Data, statistics and records generated to measure the SDG indicators can be reliably merged or combined with other data sources to support programme activities, including those supporting the organisation's strategic goals. Organisation-wide policies and standards are in place to protect records of decisions and accountability requirements, for instance 'access to information' legislation is supported by consistently applied records management policies and standards. Trends can be analysed through time, and comparisons can be made from year to year because changes to formats, coding schemes and data collection and analysis methods are well documented. Preservation standards ensure that data, statistics and records are stored properly and migrated to take account of changes in technology. The preservation programme ensures continued accessibility and authenticity of data, statistics and records through time.

Example

Gender equality is a government strategic priority. Labour force data used to produce statistics for measuring the proportion of women in management positions (SDG goal 5) is being merged with statistics from the Ministry of Industry. This is possible because of the way the data from both sources were formatted and described. The resulting database can be used to measure progress toward gender equality and contribute to the statistics needed to measure SDG indicator 5.1.2. The comprehensive framework of policies, standards and assigned accountability ensures the integrity and trustworthiness of the data, statistics and records. A preservation programme dedicated to ensuring the authenticity and completeness of the increasing volumes of data and statistics makes it possible to perform complex analyses through time.

Level 5 Processes generating data, statistics and records, and the framework for managing them are designed to make it possible to exploit data, statistics and records in new and innovative ways

Managers of SDG initiatives understand the benefits of sharing and exploiting data, statistics and records for stimulating innovative thinking on implementing the SDGs and achieving the operational goals of individual programme activities and the strategic goals of the organisation. Professional staff have the knowledge and expertise needed to design comprehensive management frameworks (sometimes covering multiple organisations and technology environments) that support the exchange and exploitation of information data, statistics and records to the greatest possible extent.

Example

Employment data from several large private enterprises have been merged with the government's labour force data and employment data to create a government–industry database. The complex inter-jurisdictional processes are well documented, data is well-managed and the statistics produced from the database can be trusted because the

management framework can be trusted. Staff have the confidence to look for new and innovative ways to exploit the data, even as its volume and complexity grows. Innovative and advanced technologies are applied, and information is published in new forms to meet the needs of a wide range of individuals and groups and to give citizens access regardless of location. A wide range of statistical products serve multiple purposes, including measuring SDG 5 and managing the government's commitments in support of the Open Government Partnership's agenda on gender equality.

Taking first steps

Rather than trying to work on everything at once, it is proposed that the government should start by identifying and defining solutions for a few processes where weak management of data, statistics and records has significant implications for achieving the SDGs.

Identify a leader and assemble a team

Given the Ministry of Government Affairs' leading role in the SDG initiative, a senior official in the ministry should oversee the initiative. This person should have a background in data management, statistics, information technology or records management, the capacity to bridge these disciplines and the ability to communicate with a variety of stakeholders, including senior management.

A steering committee should be appointed, made up of representatives from government programmes supporting the SDGs as well as programmes where the quality and integrity of data, statistics and records is particularly important. Experts in managing data, statistics, records and information technology, as well as legal experts and auditors should also be included. The committee should help select the SDG processes to be covered, identify issues and strategies and explore how to extend the results to other SDG goals.

Some government officials have argued that attempting to build a comprehensive framework is creating a 'mountain out of a mole hill' and that the focus should be on addressing immediate issues associated with specific SDG initiatives. Others have realised that systemic issues need to be addressed across the government as a whole. This tension between the need to address immediate and critical problems and the goal of developing comprehensive and sustainable solutions needs careful management.

Identify processes as examples

For each of the three process types (survey, registration/administrative and scientific), identify one or two processes that present significant challenges for measuring one or more SDG indicators and for using data, statistics and records for operational and programme delivery. These are likely to be processes where undocumented flaws or inaccuracies in data, statistics and/or records have led to embarrassment, bad decisions about the use of government resources, missed opportunities or increased risk and costs.

Describe the selected processes

The description should cover the stages of managing a given process that generates data, statistics and records.

- The stages of a survey process are likely to include planning and approving the survey; designing the survey methodology; designing the data collection tools and techniques (such as survey forms); testing the survey methodology; conducting the survey; analysing the results; reporting the findings; and reviewing how the survey was conducted.
- In the case of a registration/administration process, the stages are likely to reflect those of the systems development life cycle, including planning the system; defining functional requirements; designing the system and database; testing the design; implementing the system and the database; maintaining the system and database; and evaluating the extent to which the system and database follow the stated requirements.
- In the case of a scientific process, the stages would include planning the project; assessing data collection methods and technologies; designing the process; testing data collection and measurement tools, procedures, analytical techniques and statistical reporting methods; implementing and maintaining the process; and reviewing/evaluating the project.

It should be possible to identify the data, statistics and records created at each stage. The aim is not to describe every single stage and every piece of data, statistics and records for a given process but to identify key stages of the process and the associated data, statistics and records that are significant for measuring an SDG and providing an authentic documentary trail of the process.

Finally, the overall framework for managing both the process and the data, statistics and records should be reviewed. Policies and standards are particularly important, as is the governance structure (who is accountable to whom for what). This will provide a template for analysing the quality and integrity of the process itself and the data, statistics and records it generates.

Appendix A presents a series of graphics illustrating the process, the data, statistics and records, and the management framework relate to one another. Hopefully, it will provide a way of thinking about how the task can be undertaken.

Identify issues and implications

It should then be possible to analyse the issues, distinguish between symptoms and causes and identify solutions. For instance, a poorly documented, flawed data file input to a set of statistics is a **symptom**. The **cause** was the failure to establish metadata and documentation standards in the planning and design stages and to assign accountability for implementing them as part of the management framework. In identifying the issues, it is important to distinguish between immediate issues particular to measuring a given SDG indicator and issues related to the broader management framework for the organisation as a whole.

Finally, issues should be explained in a way that programme managers responsible for generating SDG statistics can understand. The message needs to be reinforced continually that where data, statistics and records are flawed and their accuracy cannot be established, the credibility of the manager responsible will be undermined irrevocably. By extension, society's trust that the government is capable of carrying out its obligations, including achieving the SDGs, will be eroded significantly.

Develop strategies for resolving issues

Understanding where symptoms and causes are located on the road map and how they relate to one another will help in developing strategies in relation to the overall management framework. Most of the strategies should focus on the planning and design stages of the survey, system or other data collection and analysis activity, when changes can be integrated most easily and cost effectively. For instance, enhanced metadata standards and procedures should be developed and applied at the planning stage to enable data, statistics and records to be related to one another. This approach to developing and implementing strategies can be applied to any process, from small one-time surveys to large IT systems with continuously updated databases from which data and statistics supporting SDGs are extracted.

Various international professional organisations have developed specific guidance on improving the quality and integrity of data, statistics and records, ranging from data and records management organisations, to professional associations and others focusing on managing statistics. Appendix B presents a preliminary list of standards and guides that should be helpful in developing standards-based strategies for addressing the issues that are identified.

Apply the experience to other processes and to the framework for managing data/statistics/records

The approach should result in strategies that can be applied to all processes, not just those measuring the SDG goals but any process supporting the government's programmes and services, especially where the ability to make decisions and meet accountability requirements are being placed at risk. Over the longer term, the results will be invaluable in developing a comprehensive, policy-driven standards-based framework for managing data, statistics and records, regardless of the process or business function.

Ultimately the goal is to build a comprehensive management framework to cover all government programmes and services and allow the government to demonstrate that the data, statistics and records it generates can be trusted. The outcome should be that the government is able to demonstrate a high level of credibility, both to the Amanian people and to international partners, investors, development agencies and other international organisations, including the United Nations.

Sources consulted

Publications

United Nations
Cape Town global action plan for sustainable development data:
https://unstats.un.org/sdgs/hlg/Cape-Town-Global-Action-Plan/

The Cape Town Global Action Plan for Sustainable Development, based on a consultation meeting held in November 2016 in Cape Town and adopted by the United Nations Statistical Commission in March 2017. The plan supports the need for quality and timely data, particularly by strengthening national statistical systems.

Sustainable Development Goals Report 2017: harnessing the power of data for sustainable development:
https://unstats.un.org/sdgs/report/2017/harnessing
To fully implement and monitor progress on the SDGs, decision makers need accurate, timely, sufficiently disaggregated, accessible and relevant data that is accessible and easy to use. Data availability and quality has steadily improved over the years, but statistical capacity still needs strengthening and data literacy must be enhanced at all levels of decision making. This will require coordinated efforts on the part of data producers and users from multiple data systems. It will also demand innovative ways to produce and apply data and statistics in addressing the multifaceted challenges of sustainable development.

Sustainable Development Goal indicators:
https://unstats.un.org/sdgs/

https://unstats.un.org/sdgs/indicators/database/

https://unstats.un.org/sdgs/metadata/

https://unstats.un.org/sdgs/iaeg-sdgs/

The UN Statistics Division website provides information on developing and implementing an indicator framework for the follow up and review of the SDGs. In March 2015, at its forty-sixth session, the United Nations Statistical Commission created the Inter Agency and Expert Group on SDG Indicators (IAEG-SDGs), composed of member states and including regional and international agencies as observers. The IAEG-SDGs was tasked to develop and implement the global indicator framework for the goals and targets of the 2030 Agenda. The global indicator framework, developed by the IAEG-SDGs, was agreed at the forty-eighth session of the United Nations Statistical Commission in March 2017. The IAEG-SDGs formed three working groups to address specific areas relevant to SDG indicator implementation: statistical data and metadata exchange (SDMX), geo-spatial information and interlinkages. See also:

https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-06/20171108_Draft%20Guidelines%20and%20Best%20Practices%20for%20Global%20SDG%20Data%20Reporting.pdf

https://unstats.un.org/sdgs/files/meetings/iaeg-sdgs-meeting-06/7.%20IAEG%20Draft%20Guidelines.pdf

International seminar on open data for the Sustainable Development Goals, September 2017:
https://unstats.un.org/sdgs/meetings/sdg-seminar-seoul-2017/

The seminar reviewed and recommend solutions and good practices by addressing the practical question of how to make data more open and accessible in the context of national statistical systems, from planning to a sustained and effective operation. See for instance: 'From collection to use: data and agenda 2030':

https://unstats.un.org/sdgs/files/meetings/sdg-seminar-seoul-2017/S6_P3_Jean-Louis_Sarbib.pdf

Sustainable Development Solutions Network: counting on the world: building modern data systems for sustainable development, September, 2017:
http://unsdsn.org/wp-content/uploads/2017/09/sdsn-trends-counting-on-the-world-1.pdf

There is not just one way to harness the data revolution for sustainable development, and there is not one perfect statistical system. This report brings on board a wide range of actors, each using different methodologies and approaches to producing, analysing, curating and disseminating data. It explains the kinds of data needed to achieve the SDGs and identifies the roles and responsibilities of different actors, as well as the changes needed to build systems and structures capable of responding to the growing demand for high-quality, disaggregated and geo-referenced data.

Sustainable Development Solutions Network: Trends: data and statistics for sustainable development:
http://unsdsn.org/what-we-do/thematic-networks/trends/

Trends operates independently from, but in concert with, the formal UN system and the Global Partnership for Sustainable Development Data. It focuses on including academic and technical partners as active participants in setting standards and agendas within the field of sustainable development data.

Open Government Data and Services, UN Department of Economic and Social Affairs: strengthening of capacities of developing countries to provide access for sustainable development through open government data:
http://www.unpan.org/ogdce

Global Partnership for Sustainable Development Data: http://www.data4sdgs.org
This global network of governments, the private sector and civil society organisations focuses on data availability and quality to support SDGs.

Open Data Watch: open data to support the Sustainable Development Goals:
http://opendatawatch.com/wp-content/uploads/2017/03/UNSC2017-OpenDataSupportsSDGs.pdf
http://opendatawatch.com

This international, non-profit, non-governmental organisation works at the junction of open data and official statistics, monitoring open data policies, measuring their success and impact, sharing knowledge, building partnerships and offering strategic advice and practical assistance to national governments, international organisations and other NGOs.

World Bank: how open data can drive sustainable development (discussion paper):
http://blogs.worldbank.org/ic4d/new-report-how-open-data-can-drive-sustainable-development

Development Cooperation Report 2017: data for development:
http://www.oecdilibrary.org/docserver/download/4317041e.pdf?expires=1510232633&id=id&accname=guest&checksum=A4185B1CB54456D2D5E1951CCC396D1A

This report sets out priority actions and good practices to help policy makers and providers of development assistance to bridge the global data divide, notably by strengthening statistical systems in developing countries to produce better data for better policies and better lives.

OECD: Open Government, the Sustainable Development Goals and citizen involvement: a synopsis from OECD, 2016:
https://opengovernment.org.uk/wp-content/uploads/2017/02/SDG-Summary-of-OECD-Summary-of-OECD-Report.pdf

Open Data Barometer, a survey of open data policy and practice across the world, launched by the World Wide Web Foundation – and ongoing research by the Open Data Research Network.

World Wide Web Foundation: Open Data Barometer:
https://opendatabarometer.org/

Produced by the World Wide Web Foundation as a collaborative work of the Open Data for Development (OD4D) network, and with the support of the Omidyar Network, the Open Data Barometer analyses global trends and provides comparative data on countries and regions, using an in-depth methodology that combines contextual data, technical assessments and secondary indicators. Focuses, in part, on data quality and capacity.

SDG16 data initiative:
http://www.sdg16.org/about/

This is a consortium of organisations that is attempting to pull together data sets in an open format to track SDG16 and to provide a snapshot of the current situation, and, eventually, of progress over time. It also seeks to identify potential challenges in data quality, availability and coverage that need to be addressed.

Open Government Partnership: joint declaration on Open Government for the implementation of the 2030 Agenda for Sustainable Development, 2015:
https://www.opengovpartnership.org/sites/default/files/attachments/OGP_declaration.pdf

World Resources Institute: Open Government Partnership: achieving Sustainable Development Goals through accountability and transparency:
http://www.wri.org/blog/2015/10/open-government-partnership-achieving-sustainable-development-goals-through

Open Knowledge International: The state of open government data in (June) 2017:
https://blog.okfn.org/files/2017/06/FinalreportTheStateofOpenGovernmentDatain2017.pdf

United Nations Department of Economic and Social Affairs, international knowledge-sharing workshop on Open Government data for sustainable development, The Hague, Netherlands, June, 2017:
http://workspace.unpan.org/sites/Internet/Documents/Rev-FinalReport%20-%20OGD%20International%20Knowledge-sharing%20Workshop_10-07-2017%20with%20edits%20DLB.docx.pdf

Books

Berenson, Kathy (2017) *Managing Your Research Data and Documentation* (American Psychological Association).

Jerven, Morton (2013) *Poor Numbers: How We are Misled by African Development Statistics and What to Do About It* (Ithaca, NY: Cornell University Press).

Articles and presentations

Data Science and Ensuring Data Integrity in Validation Statistics, Valgenesis (private company presentation), http://www.cbinet.com/sites/default/files/files/Soto_Ivan_pres(1).pdf

Fahari, Abdul Bari (2017) 'New research on open government data in developing countries: what we can learn from developing countries'. R&E Search for Evidence, 15 October, https://researchforevidence.fhi360.org/new-research-open-government-data-developing-countries-can-learn-case-studies

Gelbstein, Ed (2011) 'Data integrity—information security's poor relation', *ISACA*, 6, https://www.isaca.org/Journal/archives/2011/Volume-6/Pages/Data-Integrity-Information-Securitys-Poor-Relation.aspx

Kalow, Jared and O'Donnell, Megan (2017) *To Leave No One Behind, Data Disaggregation Needs to Catch Up*. Center for Global Development, 19 January, https://www.cgdev.org/blog/leave-no-one-behind-data-disaggregation-needs-catch

International Guide to the Data Management Body of Knowledge, https://technicspub.com/dmbok/

Maunu, Piritta (2017) 'More than bytes and signatures: managing computerised systems to ensure data integrity', *R&D*, 30 August, https://www.rdmag.com/article/2017/08/more-bytes-and-signatures-managing-computerized-systems-ensure-data-integrity

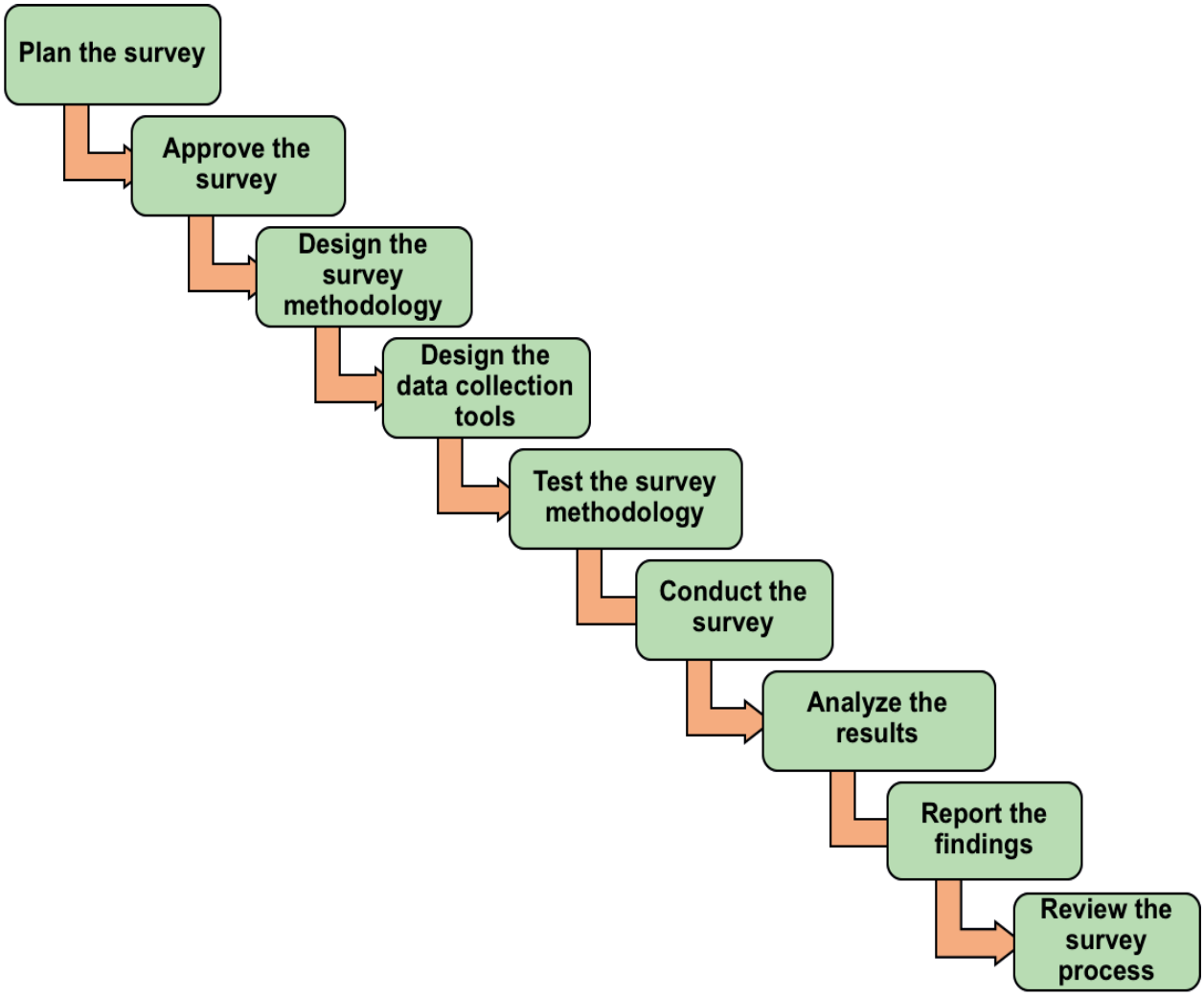
Risse, Nathalie (2017) *OECD report outlines challenges and recommendations for SDG data*. International Institute for Sustainable Development, SDG Knowledge Hub, 19 October, http://sdg.iisd.org/news/oecd-report-outlines-challenges-and-recommendations-for-sdg-data/

'Tracking the global goals: four steps to make data matter', *The Guardian*, 4 May 2016, https://www.theguardian.com/global-development-professionals-network/2016/may/04/tracking-the-goals-four-steps-make-data-matter

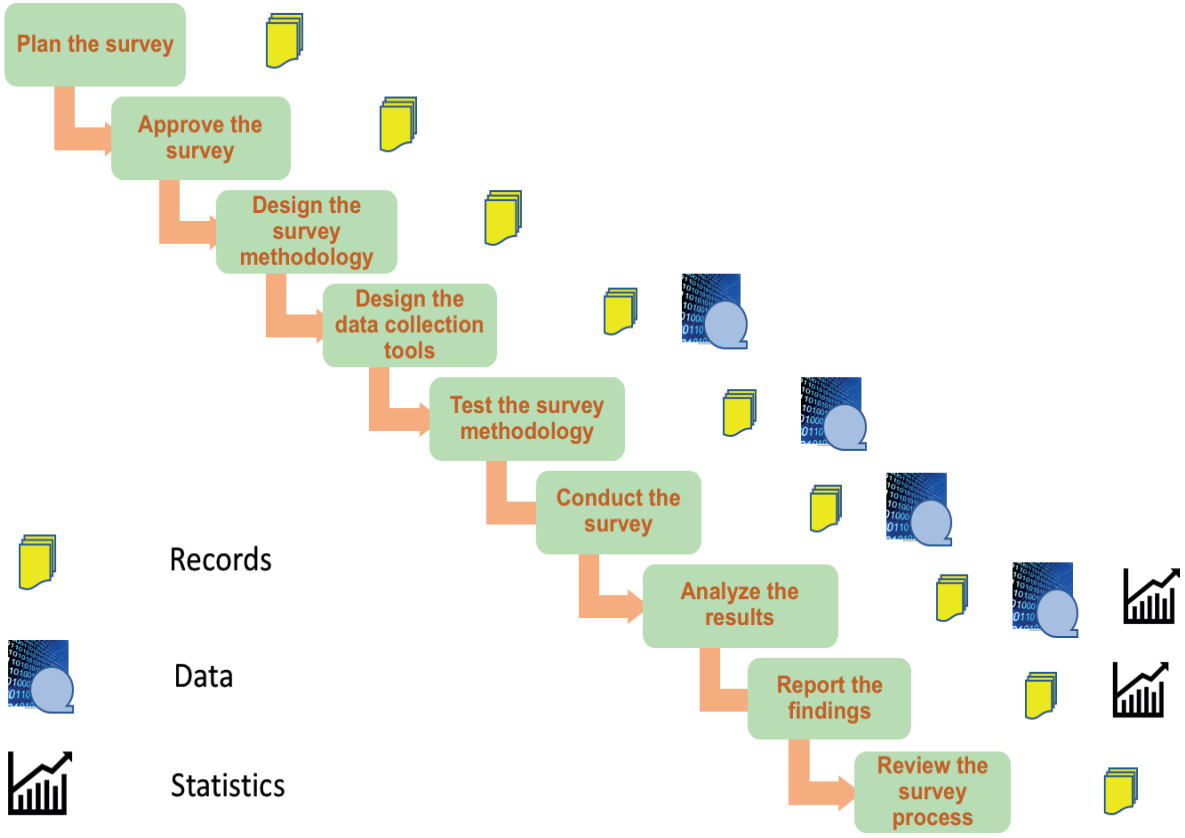
Appendix A

Concepts

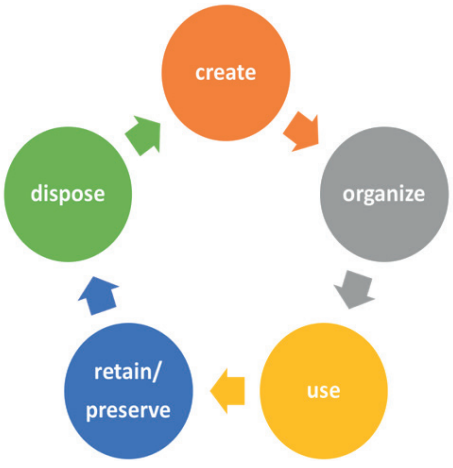
The graphics that follow illustrate the way that a map can be created showing data, records and statistics in relation to the process that generated them. For instance, the tasks in the process supporting a typical survey are:



Data, statistics and records are generated at each stage of the process to serve specific purposes. If the purpose is to support decision making, provide evidence of actions, decisions or accountability, or confirm individual rights and entitlements, it must be possible to demonstrate that the data, statistics or records are authentic and reliable, have integrity and can serve as trusted sources of information.

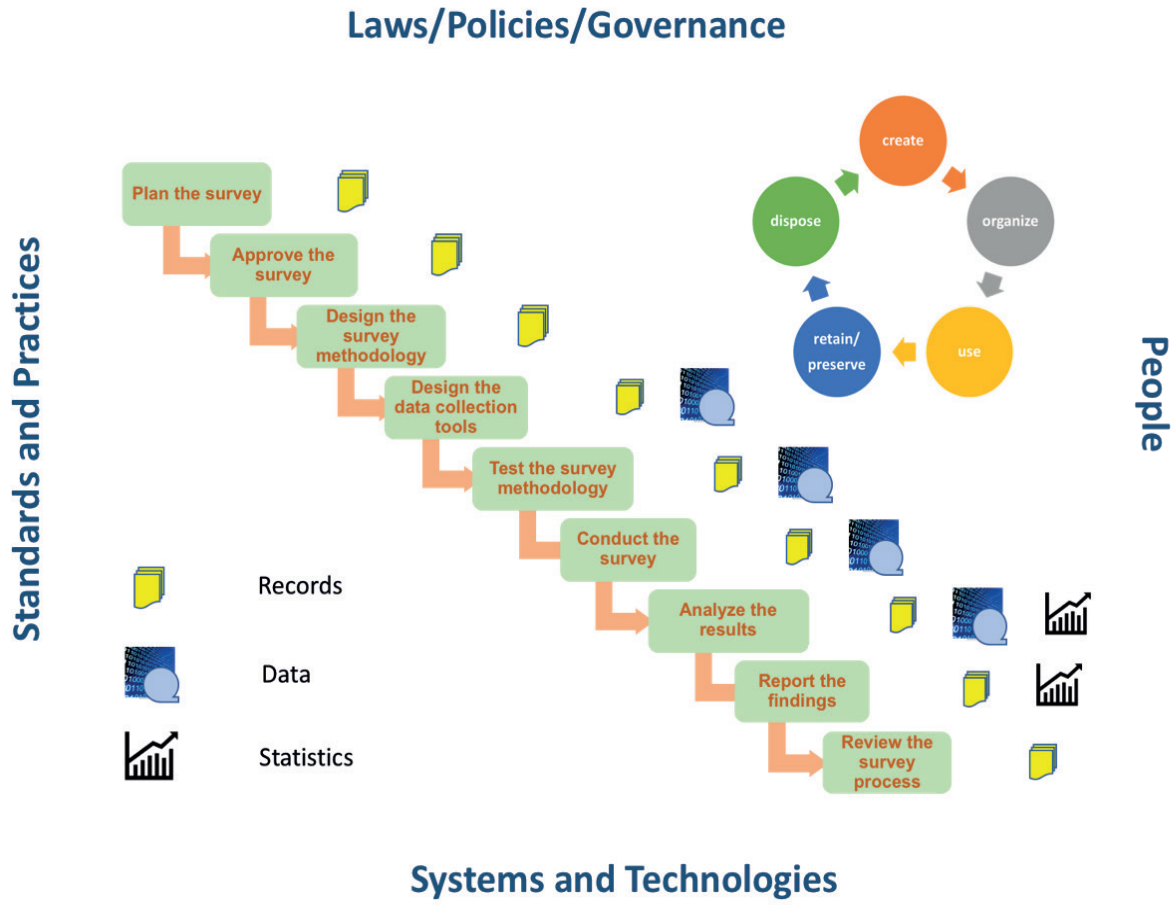


Data, statistics and records proceed through a **life cycle** beginning with their creation and ending with their final disposition, either through deletion/destruction or transfer to an archive. Their authenticity, reliability and trustworthiness must be protected for as long as they are needed to support their multiple purposes.



Metadata is what makes it possible to relate data, statistics and records to one another. It captures information about the data, statistics and records, for example how they were created, the systems and processes used to generate and manage them and the activities they support. Metadata can identify, authenticate and contextualise data, statistics and records. It is the basis for managing information about them through time, providing specific information about their content, the relationships among them, their technical features (such as formatting, supporting software, hardware) and their preservation.

If data, statistics and records and their supporting metadata are to serve multiple purposes, to be capable of being interrelated and to retain qualities of integrity and trustworthiness throughout their life cycle, they must be managed through a structured and consistent management framework of laws, policies, governance structures, standards and practices, enabling systems, technologies and qualified/trained people who know the importance of reliable information for achieving goals and priorities and demonstrating accountability.



Appendix B

Additional resources and standards for managing data, statistics and records

Several international and national organisations have developed guidelines, codes of practice, models (generic and national) and other resources that can help those responsible for managing data, statistics and records as a basis for measuring the SDG indicators. In some cases, these organisations also have developed authoritative, widely accepted standards that can provide a foundation for developing standards-based strategies, consistent in design and application. This appendix describes key resources and standards that can help manage the quality and integrity of data, statistics and records. Although by no means exhaustive, it provides a representative view of the range of standards and resources available and underlines the importance of adopting a multi-disciplinary approach to a multi-disciplinary set of issues.

It is important to understand how these standards and resources were produced. Organisations, committees and groups concerned with managing data, statistics and records have established goals and priorities that have shaped the development of standards and resources. For instance, all three disciplines have developed metadata standards and resources. Certain areas of the data management community, such as the those involved in managing geo-spatial data, have developed authoritative and sophisticated metadata standards, models and guides, including implementation guides, that can augment the efforts of other disciplines developing similar metadata standards and resources.

From the perspective of the field of statistics, the processes supporting data collection, the production of statistics and the need to ensure that processes underpinning the statistics can demonstrate a high level of quality and integrity, has led those in the statistics field to develop a large number of standards and resources to address this important issue. These can be helpful to those in other disciplines who are struggling to enhance the quality and integrity of work processes that may be similar in design to those in the statistics field.

With regard to preservation, the records management/archives community has been more active than the other disciplines in developing digital preservation standards and resources with a focus on preserving the quality, integrity and trustworthiness of digital records over the long term. The records community has long understood the need to maintain the characteristics of complex record types and their metadata even as software, hardware, format and storage media have changed. This has resulted in an extensive suite of standards and resources that can benefit those in other disciplines facing the same digital preservation challenges.

Data management

Standards

Certified Documentation and Data Management Association
Cross industry standards: https://cdma-ca.com/Cross-Industry-Standards

DDI Alliance
Data Documentation Initiative: https://www.ddialliance.org

International Organization for Standardization

- ISO/IEC JTC 1/SC32 Data management and interchange:
  https://www.iso.org/committee/45342/x/catalogue/
- ISO 8000: data quality:
  https://www.iso.org/standard/50798.html
- ISO/IEC 25012
  data quality model:
  https://iso25000.com/index.php/en/iso-25000-standards/iso-25012
- ISO Open Geospatial Consortium, TC 211
  Guide to the role of standards in geospatial information management:
  http://ggim.un.org/meetings/GGIM-committee/8th-Session/documents/Standards_Guide_2018.pdf

Professional Petroleum Data Management Association
Standards and best practices:
https://ppdm.org/ppdm/PPDM/Standards/PPDM/PPDM_Standards.aspx?hkey=fcd52d61-9bd5-4f27-9c91-a4b9fda3d2e8

United States, US Geologic Survey
Geo-spatial standards:

- USGS National Geospatial Program Standards and Specifications
- USGS Core Science Analytics and Synthesis (CSAS) Metadata Program Standards
- FGDC National Data Standards Publications
- FGDC Standards Working Group
- U.S. Integrated Taxonomic Information System (ITIS)
- U.S. National Vegetation Classification (USNVC)

Resources

Bank of England, Statistics and Regulatory Data Division
Data quality framework:
https://www.bankofengland.co.uk/-/media/boe/files/statistics/data-quality-framework

Canadian Institute for Health Information
Data and information quality:
https://www.cihi.ca/en/submit-data-and-view-standards/data-and-information-quality

Data Management Association International
Body of knowledge: https://dama.org/content/body-knowledge

Digital Curation Centre
Resources: http://www.dcc.ac.uk/resources

International Aid Transparency Initiative
Data quality:
https://iatistandard.org/en/guidance/publishing-data/data-quality/

International Association for Information and Data Quality
Publications
https://www.iqint.org/publications/

International Association for Social Science Information Services and Technology
Resources:
http://www.iassistdata.org/resources

New South Wales, Digital.NSW
Managing data and information:
https://www.digital.nsw.gov.au/support-services/data-information/managing-data-information

New Zealand, Digital Government New Zealand
Data management:
https://www.digital.govt.nz/standards-and-guidance/data-2/data-management/

Science Europe
Guidance document presenting a framework for discipline-specific research data management:
https://www.scienceeurope.org/wp-content/uploads/2018/01/SE_Guidance_Document_RDMPs.pdf

UK Data Service
Prepare and manage data:
https://www.ukdataservice.ac.uk/manage-data

United States, Food and Drug Administration
Data integrity and compliance with CGMP guidance for industry:
https://www.fda.gov/downloads/drugs/guidances/ucm495891.pdf

United States, Health and Human Services, Office of Research Integrity
Guidelines for responsible data management in scientific research:
https://ori.hhs.gov/images/ddblock/data.pdf

United States, Library of Congress
Data integrity ways and means:
http://www.digitalpreservation.gov/meetings/documents/othermeetings/3-4_Clarke-SUN-Data_Integrity.pdf

United States US Geological Survey
Data management plan, stewardship:
https://www2.usgs.gov/datamanagement/plan/stewardship.php

World Health Organization
Guidance on good data and records management:
http://apps.who.int/medicinedocs/en/m/abstract/Js22402en/

Management of statistics

Standards

European Commission, Eurostat
European Statistics Code of Practice:
https://ec.europa.eu/eurostat/documents/64157/4392716/Revised_CoP_Nov_2017.pdf

International Monetary Fund
Data Quality Assessment Framework (DQAF):
https://www.imf.org/external/np/sta/dsbb/2003/eng/dqaf.htm

UNECE Conference of European Statisticians (CES)
Standards developed by the Conference of European Statisticians Standards and Models:
https://www.unece.org/leginstr/stat.html

Resources

Australia, Workplace Gender Equality Agency
Data management policies:
https://www.wgea.gov.au/sites/default/files/data-management-policies.pdf

Canada, Statistics Canada
A compendium of management practices for statistical organizations from statistics Canada's international statistics fellowship programme:
https://www150.statcan.gc.ca/n1/pub/11-634-x/11-634-x2016001-eng.htm

Data quality toolkit:
https://www.statcan.gc.ca/eng/data-quality-toolkit

Dubai Statistics Center
National framework of statistical data quality:
https://www.dsc.gov.ae/en-us/DSC-News/Pages/National-Framework-of-Statistical-Data-Quality-is-Launched-.aspx

European Commission, Eurostat
Tools and standards:
https://ec.europa.eu/eurostat/web/quality/quality-reporting

International Monetary Fund
Overarching strategy on data and statistics at the fund in the digital age:
https://www.imf.org/~/media/Files/Publications/PP/2018/pp020918-overarching-strategy-on-data-and-statistics-at-the-fund-in-the-digital-age.ashx

International Household Survey Network
Quality management: http://www.ihsn.org/node/214

National Science Foundation
Data quality guidelines: https://www.nsf.gov/policies/docs/nsfinfoqual.pdf

Statistics Netherlands
Quality: https://www.cbs.nl/en-gb/about-us/organisation/quality

Organization for Economic Cooperation and Development
Quality dimensions and guidelines for OECD statistical activities:
http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=std/qfs(2011)1&doclanguage=en

Uganda Bureau of Statistics
Plan for national statistical development 2013/2014-2017/2018 theme: enhancing data quality and use:
https://www.ubos.org/onlinefiles/uploads/ubos/pdf%20documents/PNSD2.pdf

United Nations Economic Commission for Africa (ECA)
African Centre for Statistics:
https://www.uneca.org/acs

United States, Office of Management and the Budget
Standards and guidelines for statistical surveys:
https://unstats.un.org/unsd/dnss/docs-nqaf/USA_standards_stat_surveys.pdf

United States, Environmental Protection Agency
How EPA manages the quality of its environmental data: https://www.epa.gov/quality

Records management and archives

Standards

Association of Records Managers and Administrators
Generally accepted record-keeping principles:
https://cdn.ymaws.com/www.arma.org/resource/resmgr/files/Learn/2017_Generally_Accepted_Reco.pdf

International Council on Archives
Standards and tools:
https://www.ica.org/en/standards-and-tools-0

International Organization for Standardization
Records management standards:
https://www.iso.org/committee/48856/x/catalogue/

- ISO 15489:2001 Records Management
- ISO 14721 Space data and information transfer systems - open archival information system (OAIS) - reference model
- ISO 16175-1:2010 Information and documentation - principles and functional requirements for records in electronic office environments - part 1: overview and statement of principles
- ISO 16175-2:2011 Information and documentation - principles and functional requirements for records in electronic office environments - part 2: guidelines and functional requirements for digital records management systems
- ISO 16175-3:2010 Information and documentation - principles and functional requirements for records in electronic office environments - part 3: guidelines and functional requirements for records in business systems
- ISO 16363:2012 Space data and information transfer systems - audit and certification of trustworthy digital repositories
- ISO 23081 Information and documentation - records management processes - metadata for records
- ISO 26122 Information and documentation - work process analysis for records
- ISO/IEC 27001:2013 Information technology - security techniques - information security management systems – requirements
- ISO 30300:2011 Information and documentation - management systems for records - fundamentals and vocabulary
- ISO 31000: 2012 Risk management - principles and guidelines

Resources

Archives New Zealand
Digital record keeping:
http://archives.govt.nz/advice/public-offices/digital-recordkeeping

Australasian Digital Recordkeeping Initiative
Products: http://www.adri.gov.au/content/products/

Digital Preservation Coalition
Digital preservation handbook: http://www.dpconline.org/

International Records Management Trust
http://irmt.org/

- Assessment tools
- Education and training materials
- Research reports

Interpares Trust
Research studies:
https://interparestrust.org/trust/about_research/studies

Library and Archives Canada
Information management – products, services, tools:
http://www.collectionscanada.gc.ca/government/index-e.html

Minnesota State Archives
Electronic records management resources:
http://www.mnhs.org/preserve/records/electronicrecords.htm

National Archives of Australia
Digital records:
http://www.naa.gov.au/records-management/agency/digital/index.aspx

National Archives of Malaysia
Guidelines for storing and preserving electronic records in the public sector:
http://www2.arkib.gov.my/english/elektronik.html

European Commission: ICT Policy Support Framework Programme
The European archival records and knowledge preservation: E-ARK project:
http://www.eark-project.com/

National Archives of Norway
Noark 5:
https://www.arkivverket.no/forvaltning-og-utvikling/noark-standarden/noark-5

New South Wales, State Archives and Records
Digital continuity and digital archives:
https://www.records.nsw.gov.au/recordkeeping/advice/digital-continuity-and-digital-archives

Open Planets Foundation
Services:
http://openpreservation.org/technology/services/

Public Records Office Victoria
Victorian electronic records strategy:
http://prov.vic.gov.au/government/vers

The National Archives (UK)
Information and records management:
http://www.nationalarchives.gov.uk/information-management/projects-and-work/information-records-management.htm

US National Archives and Records Administration
Records management: http://www.archives.gov/records-mgmt/

Institute of Commonwealth Studies
School of Advanced Study
Senate House
Malet Street
London
WC1E 7HU

Email: ics@sas.ac.uk
Web: commonwealth.sas.ac.uk

Follow the School of Advanced Study:

facebook.com/schoolofadvancedstudy

twitter.com/@SASNews
twitter.com/@ICwS_SAS

sas.ac.uk

ISBN 978-1-912250-17-2
90000
9 781912 250172