



UNECA



AfDB



AUC

African Guide to writing metadata for development indicators

Improving the quality of reported statistics

Acronyms

ACS	African Centre for Statistics ACS
BIS	Bank for International Settlements
DDI	Data Documentation Initiative
ECA	Economic Commission for Africa
EBB	European Central Bank
ECE	Economic Commission for Europe
GSBPM	Generic Statistical Business Process Model
IAEG-MDG	The Inter-Agency Expert Group on MDG Indicators
ICSE	International Classification by Status in Employment
ILO	International Labour Organization
IMF	International Monetary Fund
MCV	Metadata Common Vocabulary
MDGs	Millennium Development Goals
OECD	Organisation for Economic Co-operation and Development
SDMX	Statistical Data and Metadata Exchange
Statcom-Africa III	Third Session of the Statistical Commission for Africa

Contents

Acronyms	i
Foreword.....	iii
Acknowledgements.....	v
About this guide	vi
1. The importance of metadata	1
1.1. What are metadata?	1
1.2. Metadata for different purposes	2
2. International metadata guidelines and standards.....	4
2.1. Guidelines on statistical metadata	4
2.2. Metadata standards and tools	6
2.3. Benefits of improving metadata management.....	7
3. Metadata for MDG indicators.....	9
3.1. Why are metadata often lacking in reports on development?	9
3.2. Role of national statistical offices	9
3.3. Determining metadata for MDG-related data.....	11
3.3.1. Metadata with a particular data value	11
3.3.2. Metadata about an overall data series	11
4. Writing metadata.....	13
4.1. Metadata for presentation with data.....	13
4.1.1. Mandatory metadata.....	13
4.1.2. Conditional metadata	14
4.1.3. Optional metadata	15
4.2. Metadata about an overall data series	16
4.3. Tips for writing metadata.....	18
5. Conclusion.....	21
6. Templates and checklists	22
6.1. Template for data series metadata.....	22
6.2. Metadata Checklist for Data Presentations	23
References	25

Foreword

The need for accurate and timely statistics for monitoring progress has compelled countries to produce statistics more frequently than ever before. Initiatives such as the Millennium Declaration and associated Millennium Development Goals (MDGs) have provided opportunities for strengthening national statistical systems to deliver reliable statistical information for evidence-based policy and programme formulation.

There have been many efforts at international, regional and national levels to improve the collection, compilation and dissemination of development-related data in African countries. Despite the efforts and progress made, the region faces ongoing challenges in timely and reliable production of data and reports on (MDG) indicators.

One challenge is reconciling the discrepancies that often exist between the data published by international and national sources. Varying concepts and definitions, as well as different compilation and estimation methodologies often contribute to the divergent results. Statistical metadata are needed to describe the data being presented so users can understand why differences exist and interpret the information correctly.

The Inter-Agency Expert Group on MDG Indicators (IAEG-MDG) – which is responsible for compilation and publication of MDG estimates at the international level - have recognized the need to improve transparency and coordination of processes for compiling international data. Another step involved in reconciling national and international MDG data is for countries to write clear and concise metadata for all data that they put in public domain.

“National statistical systems: to improve the compilation of metadata in countries and their accessibility by users”.

Recommendation from the International Conference on MDG Statistics, Manila, October 2011

Recognizing the need for action in this area, the Third Session of the Statistical Commission for Africa (Statcom-Africa III), held in Cape Town in January 2012, requested the Africa Group on Development Indicators to develop metadata standards on development indicators and improve capacity of countries to write proper metadata. This includes exploring the possibility of applying standards such as the Statistical Data and Metadata Exchange (SDMX).

The African Centre for Statistics (ACS) at the Economic Commission for Africa (ECA) is improving its database (Statbase) to make it now possible to display both national and international indicators in order to highlight discrepancies between these two sources. It will also provide metadata for each data point for country data so as to better understand and appreciate the differences in international and national data sets.

Well-written and accessible metadata is central to the African Charter for Statistics, which cites ‘clarity and understanding’ amongst its dissemination principles and emphasizes the need for sufficient metadata. This guide has been developed to assist the African countries in writing quality metadata for dissemination with development related indicators. Its contents served as the focus of a regional workshop on writing metadata, held in Lusaka, Zambia for both French and English speaking countries in July 2012.

The Guide is based on a publication commissioned by the United Nations Economic Commission for Europe (ECE), which is in draft form at the time of writing and awaiting publication, (working title: Making MDGs Meaningful: good practices in presenting metadata).

Both this guide and the ECE publication on which it is based have been developed as part of the United Nations Development Account project on “Strengthening statistical and inter-institutional capacities for monitoring the Millennium Development Goals through interregional cooperation and knowledge-sharing”.

Dr. Dimitri Sanga
Director ACS

Acknowledgements

This guide was written by statistical consultant, Jessica Gardner, under the overall supervision of Dr. Dimitri Sanga, Director of the African Centre for Statistics, the close supervision of Mr. Raj Gautam Mitra, Chief of the Demographic and Social Statistics Section and the coordination of Mr. Oumar Sarr, statistician in the said section.

Thanks to the collaboration between the Economic Commission for Africa, the African Union Commission and the African Development Bank, Such collaboration gave opportunities to African experts to gather in various occasions for the sake of reviewing the draft Guide.

About this guide

This is a practical guide to writing metadata to be disseminated with statistical indicators, such as those related to tracking progress towards the MDGs. It has been prepared as part of the United Nations Development Account Project on “Strengthening statistical and inter-institutional capacities for monitoring the Millennium Development Goals through interregional cooperation and knowledge-sharing”.

Although the guide focuses on metadata at the point of publication or dissemination, it takes the opportunity to raise awareness about the importance of managing metadata throughout the statistical production process. There are a number of valuable resources already published on this topic and this guide draws on the recommendations of those.

The guide is divided into the following sections:

The importance of metadata

This section explains what metadata are and why they are important, particularly in the dissemination of statistical information.

Standards and international guidelines

National statistical offices have been working together over recent decades to develop standards for producing, managing and publishing metadata. This section outlines some of the key standards and existing guidelines in this area.

Metadata for MDG indicators

Estimates for indicators relating to the MDGs and other internationally agreed goals are derived from multiple sources, and published at local, national, regional and international level. This section provides guidance on how to determine which metadata should accompany data presentations and the types of metadata products that should be developed.

Writing metadata

Explaining statistical concepts can be difficult. This section looks at techniques and examples for writing statistical metadata that are easy to understand.

Templates and checklists

Templates and checklists are provided to assist readers with implementing the recommendations in this guide.

1. The importance of metadata

Statistics provide essential information to measure progress in society, the economy and the environment. It is important that they are communicated clearly and succinctly so a wide range of audiences can understand and use the valuable information contained within.

Metadata are essential for interpreting development-related data and for making meaningful comparisons over time and between countries. They are produced and used at all stages of the statistical production process, both within the organization and by the eventual users of the data.

1.1. What are metadata?

So what are metadata? Quite simply, **metadata are data that define or describe other data.**¹ They are the information needed to explain and understand the data or values being presented.

Data labels, definitions, descriptions of methodology, legends, source information, footnotes, are all examples of metadata.

The **African Charter for Statistics** (4 February 2009) defines metadata as:

*“**Metadata:** the range of information, generally textual, that fosters understanding of the context in which statistical data have been collected, processed and analyzed with the objective of creating statistical information (legal and regulatory texts, methods and concepts used at all levels of information processing, definitions and nomenclatures, etc.)”*

Including sufficient metadata is particularly important for reporting on MDG indicators because:

- Indicators often have multiple data sources;
- Users of MDG reports and related data may have limited experience in interpreting statistics;
- Differences in MDG estimates often relate to the use of different definitions and concepts and varying practices in data collection and processing.

Metadata make it possible to understand these issues and gauge the limitations of a data point and its relation to other data. They allow the user to make judgments on the comparability of data from different sources and methods.

¹ Definition of metadata from the Metadata Common Vocabulary (MCV), 2009 version. Available at http://sdmx.org/wp-content/uploads/2009/01/04_sdmx_cog_annex_4_mcv_2009.pdf

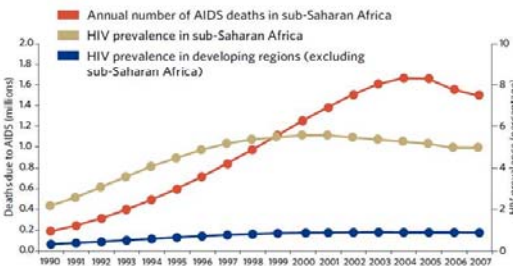
As the example in Figure 1 demonstrates, data make no sense without metadata. Metadata provide the information needed to understand what the values represent.

Figure 1 Making sense of MDG data: with or without metadata?

With metadata

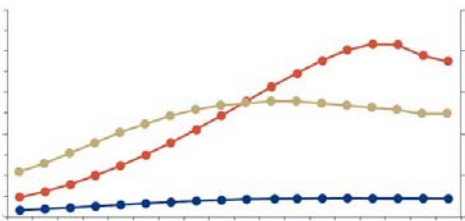
Despite small victories, AIDS continues to take a terrible toll, especially in sub-Saharan Africa

HIV prevalence in adults aged 15-49 years in developing regions and in sub-Saharan Africa (Percentage) and number of AIDS deaths in sub-Saharan Africa (Millions), 1990-2007



Without metadata

Despite small victories, AIDS continues to take a terrible toll, especially in sub-Saharan Africa



Source of original chart: United Nations, 2008, *The Millennium Development Goals Report 2008*

1.2. Metadata for different purposes

Metadata is not only relevant when data are disseminated. They are created throughout the statistical production process and have a variety of uses, such as documenting sampling methodology or identifying where data files are located.

According to international standards,² there are two types of metadata: structural metadata and reference metadata.

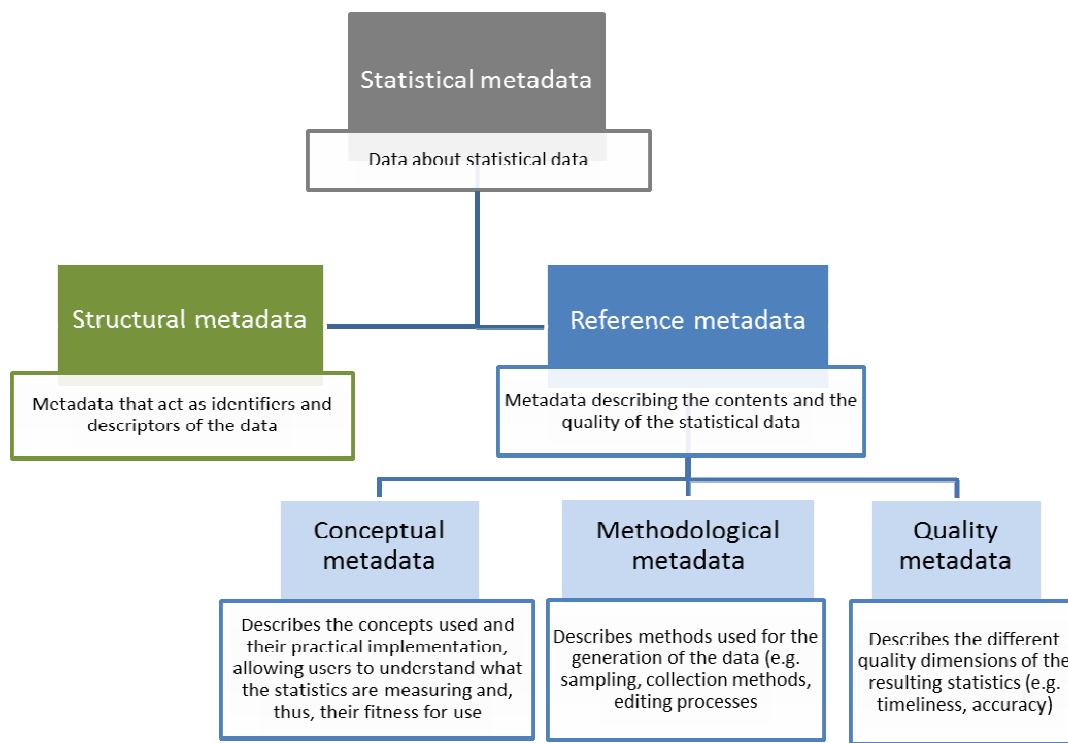
Structural metadata identify and describe data, so it can be identified and retrieved. For example, names of columns or dimensions of database cubes.

Reference metadata describe the contents and quality of the statistical data. There are three types of reference metadata:

- (a) Conceptual metadata, describing the concepts being measured;
- (b) Methodological metadata, describing the methods used to generate the data, such as sampling and collection methods; and
- (c) Quality metadata, describing the quality dimensions of the data, such as timeliness and accuracy.

This guide is more concerned with reference metadata – those describing the contents and quality of the statistics being presented.

Figure 2 **Types of statistical metadata**



Source: Diagramme from Making MDGs Meaningful: good practices in presenting metadata (ECE, forthcoming)

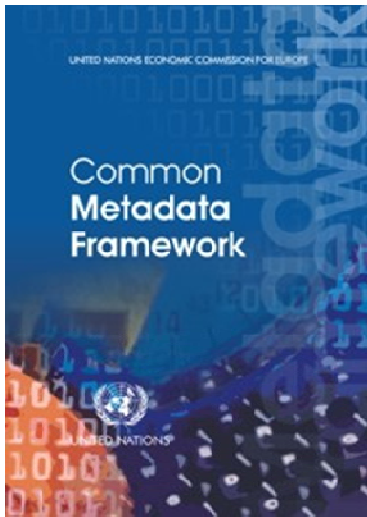
² Metadata Common Vocabulary (SDMX, 2009). Available at http://sdmx.org/wp-content/uploads/2009/01/04_sdmx_cog_annex_4_mcv_2009.pdf

2. International metadata guidelines and standards

National statistical offices face common challenges in managing statistical information and benefit from collaborating to develop standards, guidelines and tools to manage statistical metadata.

2.1. Guidelines on statistical metadata

A number of excellent resources already exist to guide countries in the management and presentation of metadata:



The **Common Metadata Framework**, maintained by ECE, is an evolving online reference to statistical metadata. It is a portal to a wide range of metadata standards and guidelines.

Contents include:

- Part A - Statistical Metadata in a Corporate Context
- Part B - Metadata Concepts, Standards, Models and Registries
- Part C - Metadata and the Statistical Business Process
- Part D - Implementation

Available in English

(<http://www1.unece.org/stat/platform/x/HYCE>)

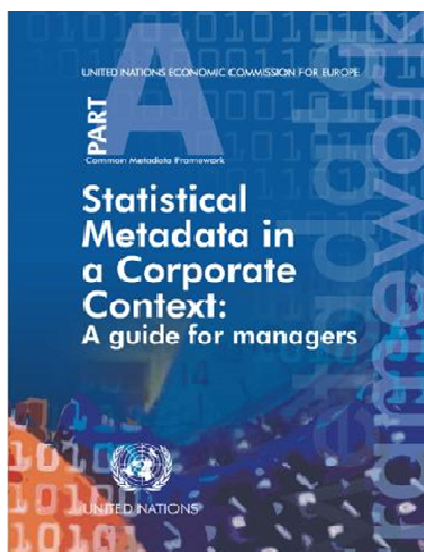


The Organisation for Economic Co-operation and Development (OECD) **Data and Metadata Reporting and Presentation Handbook** is a comprehensive guide to publishing metadata.

Contents include:

- Standards for the Reporting and Presentation of Statistics
- Importance of Use of Common Terminology
- Guidelines for the Reporting of Different Types of Data
- Guidelines for the Reporting of Different Forms of Time Series Data
- Guidelines for the Reporting and Dissemination of Metadata
- Guidelines on Reporting Practices

Available in English and French(www.oecd-ilibrary.org)



Statistical Metadata in a Corporate Context highlights the role of statistical metadata systems in a statistical organization. Focused on managerial issues related to the corporate governance of metadata systems, it will assist statistical organizations to make a business case for managing metadata effectively.

Contents include:

- The role of a statistical metadata system
- Users of a statistical metadata system
- Benefits for users
- Vision, strategy and implementation
- Core principles of metadata management
- Corporate governance models
- Case studies and experiences

Available in English and

Russian(www1.unece.org/stat/platform/display/metis/Part+A++Statistical+Metadata+in+a+Corporate+Context)

Guides on producing MDG indicators, such as the United Nations' *Indicators for Monitoring the Millennium Development Goals (2003)* and the International Labour Organization (ILO) *Guide to the new Millennium Development Goals Employment Indicators (2009)* (refer to Figure 3), outline the methodology for producing MDG estimates. These publications help not only producers, but also users of MDG-related data to understand and interpret the indicators correctly.

Such guides explain how each MDG indicator is calculated, provide definitions of the associated concepts and describe the limitations and typical data quality issues. They are excellent examples of how to present metadata, and form the basis for the recommendations in this guide on how national statistical systems should document and publish information on their approach to producing development statistics.

In statistics, it is good practice to conform to international standards, as emphasized in the

Figure 3 Guide to producing employment-related development indicators



Fundamental Principles of Official Statistics.³ Compliance with international standards can lead to greater consistency and interoperability within the organization. It will also help to exchange and share methods and tools with other organizations, both within the national statistical system and internationally.

There are several standards related to statistical metadata management that national statistical systems should at least be aware of, and possibly implement as part of their own systems and information management practices.

2.2. Metadata standards and tools

The international statistical community has made significant efforts to develop shared metadata standards and models. Key standards include:

Statistical Data and Metadata Exchange

The SDMX initiative sets technical standards and content-oriented guidelines to facilitate the exchange of statistical data and metadata.

Used by a large number of international and national statistical organizations, SDMX is maintained by a group of seven sponsors: the Bank for International Settlements (BIS), the European Central Bank (ECB), Eurostat, the International Monetary Fund (IMF), the Organization for Economic Cooperation and Development (OECD), the United Nations and the World Bank.

Data Documentation Initiative (DDI)

The DDI is a standard for technical documentation describing social science data. The current version (3.1) supports description of the full life cycle of a dataset or data collection. DDI is used by organizations to classify and manage information throughout the process of statistical production.

Metadata Common Vocabulary (MCV)

The MCV contains concepts and related definitions that are normally used by international organizations and national data producing agencies to describe statistical metadata. Terms such as census, estimate, footnote, measurement error, occupation, periodicity, quality and sample are all defined in the MCV.

The MCV is a valuable resource for establishing common terminology in the presentation of MDG data and metadata.

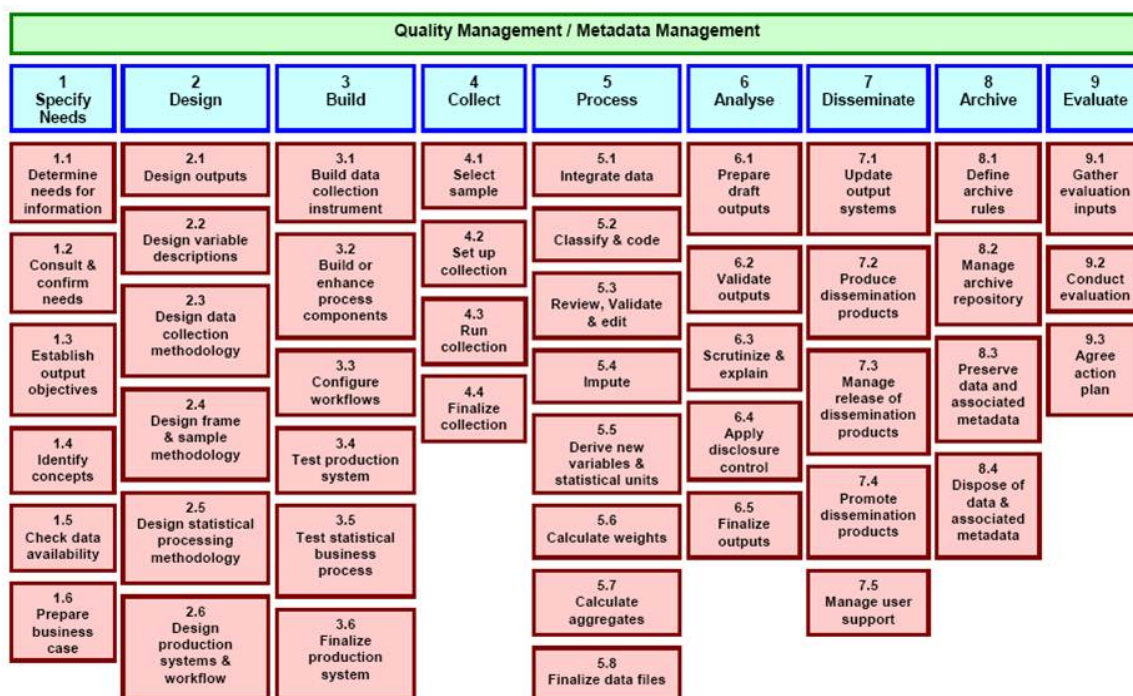
The statistical metadata community, which is most active in the ECE region through the Steering Group on Statistical Metadata (METIS),⁴ has been sharing knowledge and developing guidelines about managing metadata since the 1980s. Practical tools, such as the MCV emerged from the work of this group, as well generic models which represent the processes and information that are involved in the production of official statistics.

³ United Nations, 1994. *Fundamental Principles of Official Statistics* (<http://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>)

⁴ ECE Statistical Metadata Working Group (METIS) - www1.unece.org/stat/platform/display/metis/.

One such example is the Generic Statistical Business Process Model (GSBPM), pictured in Figure 4. The initial driver for developing GSBPM was to establish a common approach to describing the steps involved in producing statistics. The benefits of such a model include being able to identify and map at which points metadata are created and when they are reused.

Figure 4 Generic Statistical Business Process Model (GSBPM)



Source: ECE (2009) www.unece.org/stats/gsbpm.html

Capturing metadata at the source is one of the principles of good metadata management.⁵ For example, it is more efficient to document information about a data source (e.g. details of an administrative register) at the time of collection, and reuse this at the point of dissemination rather than having to go back and write documentation at a later stage. For those involved in managing metadata across a statistical organization, models such as GSBPM are valuable tools to develop systems and databases.

2.3. Benefits of improving metadata management

Managing metadata throughout the production process is a challenge for all agencies in the business of producing statistics. Good metadata management will lead to:

- Ensuring staff use up-to-date classifications and definitions
- Gaining resources previously spent on re-creating metadata unnecessarily.

⁵ Statistical Metadata in a Corporate Context: A guide for managers (ECE, 2009).

- High morale and productivity as staff can store and retrieve the information they need.
- Capitalising on lessons learned from past collections and feeding that into improvements in the next cycle.
- Data users encouraged by clear information needed to understand and interpret the data.
- Increased data use and trust in official statistics.

3. Metadata for MDG indicators

As outlined in the previous section, international recommendations on the minimum metadata that should accompany each data presentation have existed for a while now. Unfortunately, these recommendations are not always followed in MDG reports and other presentations of MDG-related data.

3.1. Why are metadata often lacking in reports on development?

There are many factors that contribute to the amount of metadata that are made available to users of MDG-related data. Reasons for the limited availability of metadata in reports and databases may include:

- MDG reports often produced by non-statisticians who may lack knowledge of metadata standards;
- Perception that metadata should be limited to appeal to a broad audience;
- A wide range of actors involved in producing MDG estimates;
- Low national capacity in producing and managing statistical metadata;
- Limited use of development-related data, therefore low demand for good metadata;
- Limited guidance on what metadata should be published.

In the past, national statistical offices focused the majority of their resources on the collection and production of statistics and less on analysing, disseminating and communicating the results. To stay relevant in the information age, statistical organizations are now placing greater emphasis on publishing data and metadata in a variety of forms to reach a broad and growing audience of data users.

3.2. Role of national statistical offices

Including sufficient metadata with MDG-related data may be hampered by the fact that national reports are often prepared by non-statisticians, who may not be familiar with the importance of good statistical metadata. This highlights the need for national and international statistical agencies to take a lead role in educating development practitioners on the importance of metadata, and checking that presentations align with good practices before they are published.

National statistical offices should first ensure they have efficient and effective procedures in place to manage and disseminate metadata. This may involve:

- Developing and implementing a glossary (or compendium) of statistical terms;

- Creating presentation guidelines that specify the form and elements of data and metadata presentations;
- Developing metadata management systems and tools, such as classification databases, metadata registries, and other centralized systems for consistently managing metadata.

A number of African countries have already made efforts to improve their metadata management. Further efforts must be made to ensure these practices are followed by others involved in the production and dissemination of MDG estimates.

Figure 5 Glossary of concepts and definitions developed in Uganda



Source: www.ubos.org/onlinefiles/uploads/ubos/pdf%20documents/compendium.pdf

3.3. Determining metadata for MDG-related data

Reports on MDG progress occur at the national and international level and cover a broad range of topics. As there are different processes and actors involved in these reporting streams, it is a challenge for developing countries to ensure consistent and coherent presentation of data and metadata.

Regardless of the sources and channel for reporting MDG-related data, good practices should be followed to ensure metadata are sufficient. It is helpful to think of different levels of detail when presenting metadata, from that which must be presented alongside the data, such as titles describing the data in tables, charts or maps, units of measurement and footnotes on breaks in series to more detailed information, such as definitions or guides on methodology, which may be provided in appendices or in a separate publication.

There are two kinds of metadata presentation that need attention:

3.3.1. Metadata with a particular data value

A guiding principle for publishing data is that tables, charts and maps should contain sufficient metadata so that they can “stand alone”, meaning readers can understand what is being presented without having to read the supporting text unless they are clearly directed to do so.

Tables, charts and maps should always contain enough metadata for the following reasons:

- It helps them to be understood at a glance;
- Readers are more likely to absorb and apply the findings correctly;
- Readers might copy and paste graphs or tables into their own reports or papers and it is important that information such as the data source remains with the data;
- Some readers may not read the analytical text explaining presentation of data.

3.3.2. Metadata about an overall data series

As emphasized above, MDG-related reports are usually designed to engage a broad audience. Data users vary in their knowledge of statistics from people who are unfamiliar, and often uncomfortable with data, to expert users and statisticians themselves.

Regardless of their experience in using statistics, many data users may need or wish to know more about how the estimates were compiled. Metadata, such as those published by the United Nations on the MDG indicators, can be compiled in a user guide for all national development indicators. This would include information that is common to a particular

data series, such as methods of calculation, definitions used, primary sources and frequency of release.

These guides can be referenced or linked to from presentations of the actual data.

The next section recommends mandatory, conditional and optional metadata that should accompany data values on MDG-related indicators, as well as what metadata to include about an overall data series.

4. Writing metadata

It takes skills to communicate the meaning of statistics and their level of quality. Resources such as the ECE *Making Data Meaningful* guides (2006, 2009) and *Guidelines for Statistical Metadata on the Internet* (2000), provide assistance on how to communicate statistics in tables, charts and maps and how to write about numbers in a way that appeals to a broad audience. It is vital that sufficient metadata exist as well as being worded in a way that makes them easy to understand.

There is some metadata that should always accompany a presentation of data (i.e. mandatory metadata). The amount of other more explanatory metadata will depend upon the target audience and the form in which the information is being published.

In MDG reports, metadata are often limited to a minimum of detail so as not to overwhelm or confuse the user with too much information. However, this approach risks disconnecting data with the information needed to interpret it correctly. By focusing on including sufficient metadata with tables, charts and maps, and further explanation in referenced user or compilations guides, national statistical systems can provide the appropriate level of details for all data users.

4.1. Metadata for presentation with data

4.1.1. Mandatory metadata

Should accompany all presentations of data.

- | | |
|---|---|
| 1. Title describing data being presented | <ul style="list-style-type: none">(a) Statistical unit: entity for which statistics are compiled (e.g. persons, households, events, enterprises);(b) Reference area: the country or geographic area to which the measured statistical phenomenon relates;(c) Reference period: the period of time or point in time to which the measured observation is intended to refer;(d) Unit of measure: the unit in which the data values are measured. |
| 2. Data provider | <p>Name of the organization which produced the data. The term “data provide” is based on the SDMX standard. Some countries may use this term to refer to respondents, but in the context of data dissemination, it refers to the agency or organization that is responsible for releasing the data.</p> |

3. Statistical concepts and definitions

Characteristics of data as defined by a statement that represents the essential nature of the term (e.g. “education level” is a concept and a definition used to explain what the concept means).

Definitions of statistical concepts and terms should be provided either in presentations of MDG-related data or references (e.g. links) given as to where they can be found. Knowing the precise definition used by the data provider is essential to understanding the data being presented.

4.1.2. Conditional metadata

Is provided where conditions require it, such as if anomalies or differences exist between data points that need to be explained.

4. Comparability

An explanation should be provided in a footnote where differences between statistics can be attributed to differences between the true values of statistical characteristics. These may include:

- (a) **Comparability - geographical** – degree of comparability between statistics measuring the same phenomenon for different geographical areas;
- (b) **Comparability over time** – degree of comparability between two or more instances of data on the same phenomenon measured at different points in time.

5. Source data⁶

Characteristics and components of the raw statistical data used for compiling statistical aggregates, i.e. type of primary source (e.g. survey, census, administrative records) and any relevant characteristics (e.g. sample size for survey data).

6. Symbols or abbreviations

Any symbols or abbreviations used in the presentation of data should be explained.

⁶ Many of the MDG indicators are rates or ratios comprising two or more component data series that may come from different sources (e.g. the ratio of boys to girls in primary education is calculated from enrolment data and population data). The optimum metadata would specify all primary source data used in deriving the estimates.

4.1.3. Optional metadata

There is a range of other information that would be helpful in guiding the user in their interpretation and use of MDG-related data. This metadata could be provided in an annex or another section/page of the MDG product. Where it is not practical to include this level of detail in the data product itself, links and references to where the information can be found can be provided.

- | | |
|---------------------------------------|--|
| 7. Accuracy | Metadata to describe the closeness of computations or estimates to the exact or true values that the statistics were intended to measure. This includes bias (systematic error) and variance (random error). It may be described in terms of major sources of error (e.g. coverage, sampling, non-response) or measures of accuracy. |
| 8. Contact information | Individual or organizational contact points for the data, including website, mail address, phone, email, etc. |
| 9. References / Relevant links | Further information and reading on data collection methods, related analytical reports or general information that may be of value to readers. |

Figure 6 Good example of presenting metadata

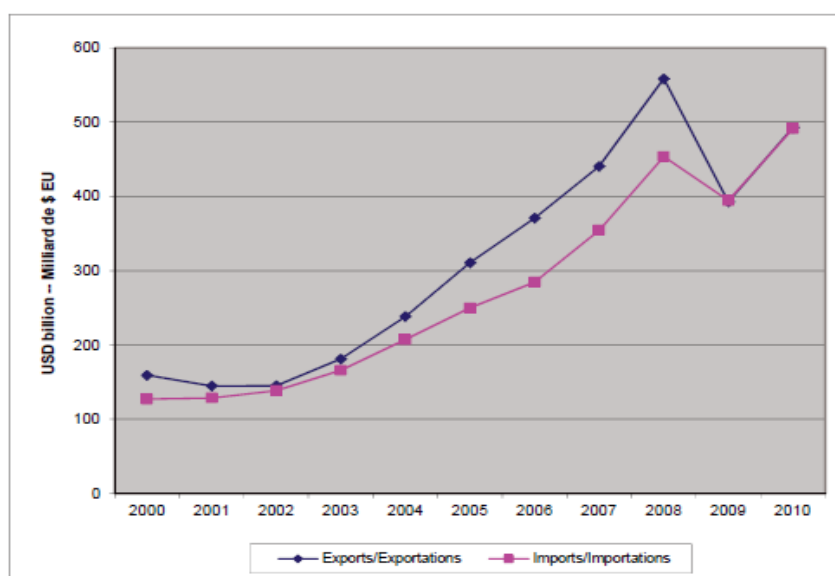


Figure 1: Trends in Africa's foreign merchandise trade, 2000–2010 (US\$ billions)

Source: ECA, African Statistical Yearbook, 2012

4.2. Metadata about an overall data series

The following metadata should be maintained for all national development indicators. The format is based on the one used by the United Nations for reporting on MDG indicators. The examples are taken from the ILO publication on employment-related MDG indicators, (2009).

Name of data series Title of the data series and reference to any names by which the indicator is known by.

For example: Proportion of own-account workers and contributing family workers in total employment (vulnerable employment rate)

Goal and target addressed Reference to the national and/or international development goal.

For example: MDG Goal 1: Eradicate extreme poverty & hunger
MDG Target 1B: Achieve full and productive employment and decent work for all, including women and young people.

Contact information Individual or organizational contact points for the data series, including website, mail address, phone, email, etc.

Method of computation The formula used to calculate the indicator and an explanation of how it is done.

For example: The vulnerable employment rate is calculated as the sum of own-account and contributing family workers as a proportion of total employment:

Vulnerable employment rate =

$$\frac{\text{number of own-account workers} + \text{number of contributing family workers}}{\text{workers}} * 100$$

Definition

Definitions should be provided for each statistical concept measured in the data series. According to the MCV (2009), a definition is “a statement that represents the concept and essential nature of the term”.

Where possible, definitions applied at the national level should align with international definitions, modified as appropriate to provide further elaboration and / or to meet national circumstances (OECD, 2007).

For example: The indicator is based on the measure of status in employment, as defined according to the International Classification by Status in Employment (ICSE), approved by the United Nations Statistical Commission in 1958 and revised at the 15th International Conference of Labour Statisticians in 1993... The 1993 ICSE defines own-account workers as: those workers who, working on their own account or with one or more partners, hold the type of jobs defined as a “self-employment jobs” (i.e. jobs where the remuneration is directly dependent upon the profits derived from the goods and services produced), and have not engaged on a continuous basis any employees to work for them...

Rationale

Explanation of why this indicator is relevant to national development.

For example: Vulnerable employment is a newly defined measure of persons who are employed under relatively precarious circumstances as indicated by the status in employment. Because contributing family workers and own-account workers are less likely to have formal work arrangements, access to benefits or social protection programmes and are more “at risk” to economic cycles, these are the statuses categorized as “vulnerable” ...

Sources and data collection

Describe the source(s) for these data and the method of data collection. Highlight any strengths or weaknesses of particular sources, especially when there is more than one source for the same indicator.

For example: Data could be obtained from population censuses, labour force or other household surveys, establishment surveys, administrative records and official estimates based on results from several of these sources. Labour force surveys can be designed to cover virtually the entire population of a country, all branches of economic activity, all sectors of the economy, and all categories of workers, including own-account workers, unpaid family workers and persons engaged in casual work or marginal economic activity. For this reason, household-based labour force surveys offer a unique advantage for obtaining information on the labour market of a country and its structure.

Gender issues

Highlight any issues that are relevant to gender equality.

For example: The indicator is highly gender sensitive since, historically, contributing family work is a status that is dominated by women. There is also a connection between vulnerable employment and poverty: if the proportion of vulnerable workers is sizeable, it may be an indication of widespread poverty. The connection arises because workers in the vulnerable statuses lack the social protection and safety nets to guard against times of low economic demand and often are incapable of generating sufficient savings for themselves and their families to offset these times.

Comments and limitations

Provide any other information about the indicator, in particular any limitations of the data.

For example: The indicator is not without its limitations; some wage and salaried workers might also carry high economic risk and some own-account workers might be quite well off and not vulnerable at all. But, despite the limitations, vulnerable employment is especially relevant for the less developed economies and regions, and the fact that a strong correlation has been established between high poverty rates for a region and high shares in vulnerable employment does substantiate the weight of the new indicator to measure progress towards the goal of decent employment for all.

Availability

Information on where the data are typically published (e.g. website address or name of recurring publication) and the frequency in which this data are usually made available.

For example: This data series is published in the annual Statistical Yearbook of country X and the website of the national statistical office (www.nso.gov.xx). New data on this indicator is usually released every 2-3 years.

4.3. Tips for writing metadata


In all cases, metadata should be clear and easy to understand by a broad audience, with technical terms either avoided or explained. Information on the limitations and reliability of data, such as sampling errors and non-sampling errors, should be provided in language that is understandable to non-specialists (OECD, 2007).

Some tips for writing metadata:

- Be aware of the target audience
- Use clear and simple language
- Keep sentences and paragraphs short
- Avoid technical terms, jargon and acronyms
- Ask colleagues to review the data and metadata – do they make sense?
- Use a standard glossary of terms for consistency
- Develop publication or style manuals that outline how data and metadata are to be presented in any statistical products

Figure 7 Good example of metadata on a data series

United Nations Statistics Division | Department of Economic and Social Affairs | United Nations
IAEG: Countries | Techgroup | English



Millennium Development Goals Indicators

The official United Nations site for the MDG Indicators

HOME DATA INDICATORS DOCUMENTS CAPACITY BUILDING LINKS CONTACTS

Select SeriesExpand

Series MetadataCollapse

Goal: Goal 6. Combat HIV/AIDS, malaria and other diseases

Target: Target 6.C: Have halted by 2015 and begun to reverse the incidence of malaria and other major diseases

Indicator: 6.7 Proportion of children under 5 sleeping under insecticide-treated bednets

Series: [Children under 5 sleeping under insecticide-treated bed nets, percentage](#)

Contact point in international agencyCollapse

Holly Newby
Statistics and Monitoring Section
Division of Policy and Planning
UNICEF
Postal Address: Three United Nations Plaza, New York, NY 10017
Telephone: 1 212-303-7932
Fax: 1 212-735-4411
E-mail: hnewby@unicef.org

DefinitionCollapse

Percentage of children aged 0-59 months who slept under an insecticide treated mosquito net the night prior to the survey.

Method of computationCollapse

The number of children aged 0-59 months who slept under an insecticide-treated mosquito net the night prior to the survey is expressed as a percentage of the total number of children aged 0-59 months included in the survey.

Comments and limitationsCollapse

Because of issues of date recall of last impregnation, this indicator may not provide reliable estimates of net retreatment status. Furthermore, the standard survey instrument does not collect information on whether the net was washed after treatment, which can reduce its effectiveness. Typically, estimates are provided for the national level, which may underestimate the level of coverage among subpopulations living in localized areas of malaria transmission.

Sources of discrepancies between global and national figuresCollapse

Because all nationally-representative data on insecticide treated mosquito net use are collected only through large-scale household surveys, and these figures are not modified, there would normally be no discrepancies between global and national figures. However, there could be discrepancies if national figures are calculated based on only those geographic areas with malaria transmission.

Process of obtaining data	Collapse
<p>Data from national-level household surveys, including Multiple Indicator Cluster Surveys (MICS), Demographic Health Surveys (DHS) and Malaria Indicator Surveys (MIS), are compiled in the UNICEF global databases.</p> <p>The data are reviewed in collaboration with Roll Back Malaria (RBM) partnership, launched in 1998 by the World Health Organization (WHO), the United Nations Children's Fund (UNICEF), the United Nations Development Programme (UNDP) and the World Bank. The RBM Partnership has expanded exponentially since its launch and is now made up of a wide range of partners — including malaria-endemic countries, their bilateral and multilateral development partners, the private sector, nongovernmental and community-based organizations, foundations, and research and academic institutions — who bring a formidable assembly of expertise, infrastructure and funds into the fight against the disease.</p>	
Treatment of missing values	Collapse
<p>There is no treatment of missing values. When the information needed to calculate the indicator is not available, the indicator is not estimated.</p>	
Data availability	Collapse
<p>Data are available for nearly all malaria endemic countries between the period of 1998 and 2006, for children under five years of age. The lag between the reference year and actual production of data series differs between surveys. Household surveys, such as DHS and MICS, are in general implemented every 3-5 years with results published within a year of field data collection.</p> <p>Data from national-level household surveys are compiled in the UNICEF global databases, which are reviewed in collaboration with Roll Back Malaria partners. Latest available estimates of these malaria prevention and treatment indicators are published annually by UNICEF in The State of the World's Children report, and are available at http://www.childinfo.org</p>	
Regional and Global estimates	Collapse
<p>Regional and global estimates are based on population-weighted averages weighted by the total number of children under five years of age. These estimates are presented only if available data cover at least 50% of total children under five years of age in the regional or global groupings.</p>	
Expected time of release	Collapse
<p>Latest available estimates are published annually, in December, by UNICEF in The State of the World's Children report, and are available at www.childinfo.org</p>	

Source: United Nations Millennium Development Goals website (mdgs.un.org)

5. Conclusion

Providing detailed metadata is essential for the accurate interpretation and use of official statistics. National statistical systems can increase the quality of their statistics by investing resources in improving the production and management of associated metadata. This involves developing consistent concepts and definitions, reviewing all stages of the production processes to create, capture and manage metadata effectively and publishing metadata in forms that can be easily accessed and understood by data users.

African countries would benefit from national and regional initiatives to improve metadata production and management. This guide suggests a regional approach to documenting and publishing metadata about MDG indicators, but the principles of good metadata management can be applied to all areas of statistics. Creating opportunities to share information and exchange experiences will enable the region to pool resources and develop consistent approaches.

Taking this work forward requires the support of senior statisticians and decision makers within the national statistical system. They will have to be convinced of the value of investing in metadata management systems. This guide links to valuable resources that include evidence of the benefits of good metadata management, which can support practitioners in raising awareness of the need for action.

6. Templates and checklists

6.1. Template for data series metadata

Name of data series	Title of the data series and reference to any names by which the indicator is known by.
Goal and target addressed	Reference to the national and/or international development goal.
Contact information	Individual or organizational contact points for the data series, including website, mail address, phone, email, etc.
Method of computation	The formula used to calculate the indicator and an explanation of how it is done.
Definition	<p>Definitions should be provided for each statistical concept measured in the data series. According to the MCV (2009), a definition is ‘a statement that represents the concept and essential nature of the term’.</p> <p>Where possible, definitions applied at the national level should align with international definitions, modified as appropriate to provide further elaboration and/or to meet national circumstances (OECD, 2007).</p>
Rationale	Explanation of why this indicator is relevant to national development.
Sources and data collection	Describe the source(s) for these data and the method of data collection. Highlight any strengths or weaknesses of particular sources, especially when there is more than one source for the same indicator.
Gender issues	Highlight any issues that are relevant to gender equality.
Comments and limitations	Provide any other information about the indicator, in particular any limitations of the data.
Availability	Reference to where the data series is usually published/made available, and the frequency in which it is usually released.

6.2. Metadata Checklist for Data Presentations

All presentations of data in tables, charts and maps should be checked against the list below to ensure that sufficient metadata are included before publication.

Mandatory	Title of the data presentation includes:		
	Name of the data series and population measured	<input type="checkbox"/>	
	Geographical area(s) covered	<input type="checkbox"/>	
	Reference period (time period that the data relate to)	<input type="checkbox"/>	
	Data provider identified		
	Name of the organization that produced the data	<input type="checkbox"/>	
	Link to web site (optional)	<input type="checkbox"/>	
	Definitions for statistical concepts provided	<input type="checkbox"/>	
	<i>Either directly with the presentation or a reference/link given as to where they can be found.</i>		
Conditional	Comparability issues noted		
	Exclusions or differences in geographic area explained	<input type="checkbox"/>	
	Different reference periods for particular values identified	<input type="checkbox"/>	
	Any anomalies for particular data points are noted	<input type="checkbox"/>	
	Primary source of the data specified		
	All details of the primary source are provided (i.e. type of data source (administrative, household survey or census, business survey or census), reference period, full official title of the series, full name of the source agency or institution)	<input type="checkbox"/>	
	Source(s) for the numerator and denominator specified	<input type="checkbox"/>	
Optional	Information on data accuracy provided		<input type="checkbox"/>

Contact information included

☐

Related publications and data sets are referenced/linked

☐

References

Government Statistical Service, United Kingdom (2009) *National Statistician's Guidance: Presentation and Publication of Official Statistics*. November 2009.
Available at: www.education.gov.uk/rsgateway/ns-nsg-presentation.pdf

Economic Commission for Europe (2012). *Common Metadata Framework*.
Available at: www.unece.org/stats/cmf.

_____ (2009a), *Making Data Meaningful Part 2: a guide to presenting statistics*.
Available
at: www.unece.org/fileadmin/DAM/stats/documents/writing/MDM_Part2_English.pdf

_____ (2009b). *Generic Statistical Business Process Model (GSBPM)*.
Available at: www.unece.org/stats/gsbpm.

_____ (2009c). *Statistical Metadata in a Corporate Context: A guide for managers*.
Available at: <http://www1.unece.org/stat/platform/x/LICE>

_____ (2000a). *Guidelines for Statistical Metadata on the Internet*, Conference of European Statistician Statistical Standards and Studies No. 52.
Available at: www.unece.org/fileadmin/DAM/stats/publications/metadata.pdf.

_____ (2000b). *Terminology on Statistical Metadata*, Conference of European Statistician Statistical Standards and Studies No. 53. Available
at: <http://www.unece.org/fileadmin/DAM/stats/publications/53metadaterminology.pdf>.

Eurostat, (2010). *Analysis of the 10 Core MDG Indicators: Methodology and Quality Issues* – Countries of ECOWAS (including UEMOA) Regional Organisation.
Available at:
http://epp.eurostat.ec.europa.eu/portal/pls/portal/!PORTAL.wwpob_page.show?_docname=2358277.PDF

International Labour Organization (2009). *Guide to the new Millennium Development Goals Employment Indicators*. Available at: http://www.ilo.org/wcmsp5/groups/public/---ed_emp/documents/publication/wcms_110511.pdf

Organisation for Economic Co-operation and Development (2007). *Data and Metadata Reporting and Presentation Handbook*.
Available at: www.oecd.org/dataoecd/46/17/37671574.pdf

Statistical Data and Metadata Exchange,(2009a). *SDMX Content-Oriented Guidelines Annex 1: Cross-Domain Concepts*. Available at:
http://sdmx.org/wp-content/uploads/2009/01/01_sdmx_cog_annex_1_cdc_2009.pdf

_____(2009b). *SDMX Content-Oriented Guidelines Annex 4: Metadata Common Vocabulary*. Available at:http://sdmx.org/wp-content/uploads/2009/01/04_sdmx_cog_annex_4_mcv_2009.pdf

United Nations Development Programme (2007). *Measuring Human Development: a primer*. Available at:hdr.undp.org/en/media/Primer_complete.pdf.

United Nations (2011). *United Nations, Millennium Development Goal 8 - The Global Partnership for Development: Time to Deliver, MDG Gap Task Force Report 2011*.

_____(2003). *Indicators for Monitoring the Millennium Development Goals*. Available at:
<http://mdgs.un.org/unsd/mdg/Resources/Attach/Indicators/HandbookEnglish.pdf>