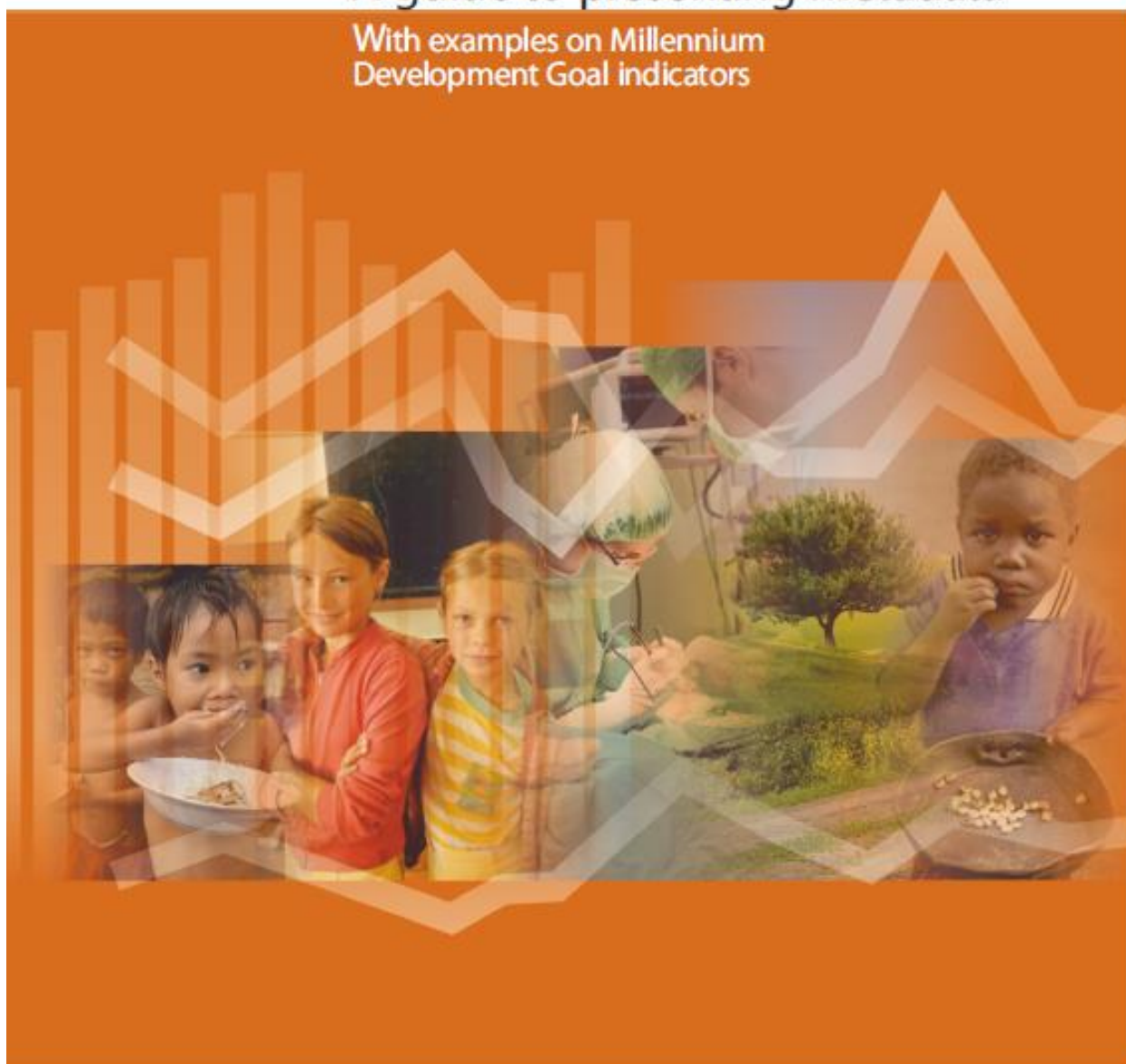

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

Getting the Facts Right

A guide to presenting metadata

With examples on Millennium
Development Goal indicators



UNITED NATIONS

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE

Getting the Facts Right

A guide to presenting metadata

With examples on Millennium Development Goal indicators



UNITED NATIONS
New York and Geneva, 2013

NOTE

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area, or of its authorities, or concerning the delimitation of its frontier or boundaries.

ECE/CES/29

Preface

The Millennium Declaration was unanimously adopted by 152 heads of State and Government at the General Assembly in 2000. The main outcome of this summit was the establishment of the Millennium Development Goals (MDGs). Monitoring progress towards the MDGs is at the heart of the framework. This is done through over sixty internationally agreed indicators and additionally many others are used for national monitoring. To support the process of monitoring MDGs, the United Nations Economic Commission for Europe (UNECE) has received a mandate to produce a regional database for MDG indicators (ECOSOC: E/2006/15/Add.1), which was launched in 2011.

It is well known that discrepancies exist between data used nationally and internationally. Metadata on definitions, methodology and primary data sources can explain these differences and evaluate the comparability of data between countries. However, the work on the UNECE database has revealed that the metadata provided in official national and international MDG publications is insufficient. It was therefore decided to produce this handbook on presenting metadata with examples from the UNECE region. The examples are based on MDG indicators, but the guidelines are applicable to any statistical data. The guide will therefore remain relevant also after the target year of 2015 of the MDGs.

The handbook is prepared in the framework of the United Nations Development Account project “Strengthening statistical and inter-institutional capacities for monitoring the MDGs through interregional cooperation and knowledge sharing” coordinated by the United Nations Economic Commission for Latin America and the Caribbean (ECLAC). Under guidance from UNECE staff, the handbook was drafted by Jessica Gardner, an external consultant. It benefitted greatly from the comments of a group of experts at the Interregional MDG indicators meeting “Sharing knowledge to improve MDG monitoring and reporting” (Santiago, Chile, 15-17 May 2012), the Conference of European Statisticians Steering Group on Statistical Metadata (METIS) and the Inter-agency and Expert Group on MDG Indicators. UNECE is grateful to all the experts who contributed to this publication.

Contents

Preface	iii
List of Figures	vi
Acronyms	ix
Executive Summary	1
I.Introduction	7
II.What are metadata?	9
Types of metadata	10
International collaboration relating to statistical metadata.....	11
III.Producing and managing metadata	13
Managing metadata: challenges and initiatives	14
Statistical metadata systems	16
Tools for managing metadata	16
Tools for disseminating metadata	17
Metadata standards	17
IV.Presenting metadata	21
Data should stand alone.....	21
Different presentations for different audiences	22
New possibilities for presenting metadata.....	24
User Guides	26
Providing guidelines for statisticians.....	28
V.Metadata for tracking development progress	31
Determining metadata requirements for MDG indicators.....	31
Recommendations for presenting metadata with MDG-related data.....	35
Specific requirements for presenting MDG-related data.....	37
VI.Specific metadata and examples of current practice	39
Indicator 1.1 – Population below \$1 (PPP) per day, percentage	39
Indicator 1.5 – Employment-to-population ratio	43
Indicator 1.7 – Proportion of own-account and contributing family workers in total employment (vulnerable employment rate).....	46
Indicator 1.8 – Prevalence of underweight children under five years of age	49
Indicator 2.1 – Net enrolment rate in primary education.....	52
Indicator 3.1 – Ratios of girls to boys in primary, secondary and tertiary education (Gender Parity Index).....	55
Indicator 3.3 – Seats held by women in national parliament	58

Indicator 4.2 – Infant mortality rate.....	59
Indicator 5.1 – Maternal mortality	62
Indicator 5.3 – Contraceptive prevalence rate	64
Indicator 5.5 – Antenatal care coverage	66
Indicator 6.3 – Proportion of population aged 15-24 years with comprehensive correct knowledge of HIV/AIDS.....	68
Indicator 6.9 – Incidence, prevalence and death rates associated with tuberculosis	70
Indicator 7.8 – Proportion of population using an improved drinking water source	73
Indicator 8.16 – Internet users per 100 population	76
VII.Recommendations	79
References	81

Annexes

I.Euro-SDMX Metadata Standard (ESMS)	85
II.Official list of MDG indicators	87

List of Figures

Figure 1. Making sense of MDG data: with or without metadata?	9
Figure 2. Example from MDG reports that lack sufficient metadata.....	10
Figure 3. Types of statistical metadata.....	11
Figure 4. Generic Statistical Business Process Model version 4.0 (UNECE, 2009b)	13
Figure 5. Metadata capture and dissemination tool developed by International Household Survey Network	17
Figure 6. The definition of “Data” in the Metadata Common Vocabulary	19
Figure 7. MDG Report of Bosnia and Herzegovina (2010) provides metadata in an annex	22
Figure 8. Different levels of detail when presenting metadata	23
Figure 9. Online tools like Gapminder make it possible to visualize and interact with MDG-related data and metadata	24
Figure 10. Metadata in the UNECE Statistical Database	25
Figure 11. Cover of the UNDP guide on measuring human development.....	26
Figure 12. Helping users in Latvia to understand technical terms.....	26
Figure 13. Presenting metadata in the OECD Statistical Database	27

Figure 14. The Australian Bureau of Statistics provides a range of guides for data users ..	27
Figure 15. Example of publication guidelines	28
Figure 16. Standard components of statistical tables	28
Figure 17. Metadata in a statistical table published in the Kazakhstan MDG Report 2005	29
Figure 18. Metadata included with a chart in the MDG Report of Slovakia of 2004	29
Figure 19. Extract from the United Nations Handbook on producing MDG estimates	32
Figure 20. Electronic version of the updated handbook on Indicators for Monitoring the Millennium Development Goals	33
Figure 21. Presenting poverty data on the World Bank website (data.worldbank.org) ..	41
Figure 22. Presentation of MDG indicator 1.1 in the official United Nations MDG Database (mdgs.un.org)	42
Figure 23. Kyrgyzstan explains their approach to poverty measurement.....	42
Figure 24. Inadequate presentation: Employment-to-population ratio in a national MDG report.....	44
Figure 25. Presentation of employment/population ratios in an OECD publication.....	45
Figure 26. Employment data presented in the ILO LABORSTA online database	47
Figure 27. Inadequate metadata: Economic status in a national MDG report.....	48
Figure 28. Using a map to present underweight prevalence data and metadata	50
Figure 29. A glossary in the MDG Report of Kazakhstan (2007) provides an explanation of terms used throughout the report	50
Figure 30. Inadequate metadata: Proportion of underweight children in the national MDG report published in 2010	51
Figure 31. Net enrolment rate presented in the Global Education Digest (UNESCO Institute for Statistics, 2011)	53
Figure 32. Inadequate metadata: Gross enrolment ratio presented in a national MDG report.....	54
Figure 33. Imperfect presentation: Gender parity index in a national MDG report.....	56
Figure 34. Explaining the gender parity index (UNESCO Institute for Statistics, 2011) ..	57
Figure 35. Presenting estimates of women’s share of parliamentary seats.....	58
Figure 36. Presenting child and infant mortality.....	60
Figure 37. Child mortality statistics in United Nations reports.....	61
Figure 38. Maternal mortality estimates in the MDG report of Armenia.....	63
Figure 39. Presenting estimates from multiple sources in Georgia	63
Figure 40. Inadequate presentation: Contraceptive prevalence rate in a national MDG report.....	65
Figure 41. Contraceptive prevalence rate in Australia	65

Figure 42. Inadequate metadata: Antenatal care statistics in a national MDG report	67
Figure 43. Antenatal care coverage in the World Health Report 2005	67
Figure 44. Providing users with the question used in data collection can aid understanding.....	69
Figure 45. Prevalence of tuberculosis in Hungary.....	71
Figure 46. Explanatory notes in the MDG Report of Hungary of 2004	71
Figure 47. Inadequate metadata: Tuberculosis mortality rate in national MDG report ..	72
Figure 48. Inadequate example: Access to improved drinking water.....	74
Figure 49. Presenting international estimates of the proportion of population with access to an improved water source.....	75
Figure 50. Inadequate presentation: The number of internet users in a national MDG report	77
Figure 51. Presentation of data on mobile phone and internet use.....	77

Acronyms

CES	Conference of European Statisticians
DDI	Data Documentation Initiative
ECLAC	Economic Commission for Latin America and the Caribbean
ESMS	Euro-SDMX Metadata Standard
GSBPM	Generic Statistical Business Process Model
HLG	High-Level Group for the Modernisation of Statistical Production and Services
IAEG-MDG	Inter-agency and Expert Group on Millennium Development Goal Indicators
ILO	International Labour Organization
ITU	International Telecommunications Union
MCV	Metadata Common Vocabulary
MDG	Millennium Development Goal
METIS	UNECE Working Group on Statistical Metadata
MICS	Multiple Indicator Cluster Survey
NADA	National Data Archive
OECD	Organisation for Economic Cooperation and Development
PPP	Purchasing Power Parity
SDMX	Statistical Data and Metadata Exchange
UNDP	United Nations Development Programme
UNESCO	United Nations Education, Science and Culture Organization
UNECE	United Nations Economic Commission for Europe
UNFPA	United Nations Population Fund
UNICEF	United Nations Children's Fund
UNSD	United Nations Statistics Division
UN Women	United Nations Entity for Gender Equality and the Empowerment of Women
WHO	World Health Organization

Executive Summary

Metadata are essential for interpreting data and making meaningful comparisons over time and between countries. This is particularly so for reporting on Millennium Development Goal (MDG) indicators, which often have multiple data sources and typically attract a wide audience of users who may have limited background in interpreting statistics.

Dissimilarities in data within and between countries may reflect real differences or be due to varying definitions, estimation and calculation methods or data collection and compilation issues. Therefore, the sources and methods for producing statistics need to be clearly explained, so the data can be better understood.

What are metadata?

Metadata are data that define or describe other data¹. They are the information needed to explain and understand the data or values being presented. A number of excellent resources already exist to guide countries in the management and presentation of metadata (OECD, 2007; UNECE, 2000a, 2009a, 2009b). They make clear recommendations on the minimum metadata that should be provided with each data presentation. Unfortunately, these recommendations are rarely followed in MDG reports and other presentations of MDG-related data.

Metadata are produced and used at all stages of the statistical production process. Effective metadata management within statistical organizations allows this information to be available and re-used whenever it is needed. Good metadata management leads to:

- gaining resources previously spent on re-creating metadata unnecessarily
- producing accurate metadata and data, thereby increasing the quality of statistics
- capitalising on lessons learned from past collections and feeding that into improvements in the next cycle
- higher morale and productivity as staff can store and retrieve the information they need
- encouraging data use by providing clear information needed to understand and interpret the data
- increasing transparency and trust in official statistics.

Producing and managing metadata

Managing metadata throughout the production process is a challenge for those in the business of producing statistics. Countries have benefited from collaborating to develop standards, guidelines and tools to manage statistical metadata. In the UNECE region,

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

this work has been carried out through the METIS (Statistical Metadata) group. A Common Metadata Framework provides a portal to shared standards and resources on metadata management (www.unece.org/stats/cmf). It includes links to information on:

- Principles of metadata management
- Metadata systems and standards
- Exchange of experiences
- Tools for managing and disseminating metadata.

Metadata standards, models and guidelines form a valuable basis for statistical organizations to develop their information management systems. Following 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.

Presenting metadata

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 remain 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.

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. Ensuring presentations contain all the metadata needed means the information can be understood at a glance and users are much more likely to absorb and apply the findings correctly.

Sufficient metadata would include:

- **A clear title** that describes the data series, population, coverage and reference period
- **Labels** to describe the data, such as variable names and units of measurement, using words that can be easily understood
- **Footnotes** that include information needed to interpret the data accurately, such as definitions, excluded populations and other exceptions
- **Source of the data**, such as the collection method, the organization that conducted it and the dates of collection (e.g. Labour Force Survey 2006).

The extent to which detailed metadata are included in the presentation of data will depend upon the target audience and the form in which the information is being published. Data users vary in their knowledge of statistics from people who are unfamiliar and often uncomfortable with data, to expert users and statisticians themselves.

The Internet and associated technologies have had a huge impact on the way that official statistics are now disseminated and used. Data can be published online quickly and cheaply like never before. Similarly, the presentation of metadata has been revolutionised by the Internet, with the possibility to link to searchable glossaries, hover over terms for instant definition and even provide videos to describe statistical methodology and tools. However, disseminating statistical information online provides new challenges as well as possibilities. Data producers must ensure that metadata are continually updated and that it remains with the data as it is downloaded and transformed into different formats.

Statistical organizations should have policies and guidelines that instruct staff on how to present statistical data and metadata in the reports and other products they release. Such guidelines need to prescribe the format data and metadata must be in before it is published, for example, the layout of tables, charts and maps and metadata that must be included. Publication guidelines, templates and the processes that ensure they are followed, will lead to complete and consistent metadata being presented with all statistics published.

Metadata for tracking development progress

Countries face particular challenges in ensuring sufficient metadata are included with reports on progress towards development goals. MDG-related data comes from numerous sources and is often compiled into reports by non-statisticians, who may be unfamiliar with standards for data and metadata presentation.

Furthermore, the demands for development-related data are high, but national coordination systems in developing countries are often weak. This leads to multiple and inconsistent sources for the same indicators, and a lack of adequate metadata to explain the discrepancies.

Guidelines on producing MDG estimates are provided by members of the Inter-Agency and Expert Group on MDG Indicators (IAEG-MDG)². These guidelines do not usually prescribe how metadata should be presented with each MDG indicator, but they do provide valuable guidance to countries on the types of metadata that are most relevant.

Each MDG indicator is based on different sources and methodologies and is usually compiled by different organizations in the national statistical system. As measurement issues vary from indicator to indicator, different metadata are needed. Fourteen MDG indicators were carefully chosen by UNECE to reflect the diversity of metadata requirements. Recommendations and examples of current practices in metadata presentation are provided for each of the following indicators:

² Coordinated by the United Nations Statistics Division (UNSD), the IAEG-MDG comprises representatives of international agencies responsible for collating and producing reports on national progress towards MDG indicators. Members include: UNICEF, UNFPA, WHO, ILO, World Bank, ITU, UNDP, UN Women, OECD, UNESCO, United Nations Regional Commissions. Refer to mdgs.un.org/unsd/mdg/Host.aspx?Content=IAEG.htm for more details.

- 1.1 – Population below \$1 (Purchasing Power Parity) per day, percentage
- 1.5 – Employment-to-population ratio
- 1.7 – Proportion of own-account and contributing family workers in total employment (vulnerable employment rate)
- 1.8 – Prevalence of underweight children under five years of age
- 2.1 – Net enrolment ratio in primary education
- 3.1 – Ratios of girls to boys in primary, secondary and tertiary education (Gender Parity Index)
- 3.3 – Seats held by women in national parliament
- 4.2 – Infant mortality rate
- 5.1 – Maternal mortality
- 5.3 – Contraceptive prevalence rate
- 5.5 – Antenatal care coverage
- 6.3 – Proportion of population aged 15-24 years with comprehensive correct knowledge of HIV/AIDS
- 6.9 – Incidence, prevalence and death rates associated with tuberculosis
- 7.8 – Proportion of population using an improved drinking water source.

Recommendations

Recommendations for national statistical organizations and MDG reporting bodies to enhance the availability and presentation of MDG-related metadata:

- Develop presentation guidelines for MDG data and metadata
- Adopt a common terminology
- Develop metadata management systems.

Recommendations to international agencies:

- Agree on international standards for MDG data and metadata presentation
- Emphasise the importance of good metadata
- Develop capacity in metadata management and presentation of statistics.

Recommendations on which metadata should be considered mandatory, conditional and optional:

Mandatory metadata

Data presented in tables, charts and maps in MDG reports, online databases, or other formats, should always be accompanied by the following metadata elements³:

³ The descriptions of each element are based on definitions found in the Metadata Common Vocabulary (SDMX, 2009b).

1. Title describing data being presented includes the following:

- 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: *organization which produced the data.*

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.

Conditional metadata

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. Comparability issues can be broken into:*

- a. **Geographical comparability** – *degree of comparability between statistics measuring the same phenomenon for different geographical areas.*
- b. **Comparability over time** – *degree of comparability between two or more data points on the same phenomenon in a time series.*

5. Source data⁴: *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.

Optional metadata

There is a range of other information that will be helpful in guiding the user in their interpretation and use of MDG-related data. These metadata could be

⁴ Many of the MDG indicators are rates or ratios comprised of 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.

provided in an annex or other 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 should be provided.

- 7. Accuracy** – *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 information on how to reach the contact points (e.g. website, mail address, phone, e-mail).*
- 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.*

I. Introduction

Metadata are essential to interpreting development-related data and to 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.

The clear targets and measurable indicators set in September 2000 by the countries of the United Nations⁵ to reach the Millennium Development Goals (MDGs), has put the spotlight on official statistics and been a catalyst for increased investment in data production and dissemination. This emphasis on quantitative measures has led to improvements in data collection through surveys and censuses, and better dissemination and use of data in policy and decision-making. This publication examines the availability of metadata – information about the data – in reports on MDG progress.

The United Nations Statistics Division (UNSD) oversees the collation and reporting of MDG-related data at the international level through the work of the international agencies mandated to monitor each indicator. These agencies work directly with countries to collect national estimates and produce coherent and comparable data at the international level. Developing countries also initiate their own reporting on progress in regular MDG reports, focusing on a selection of nationally relevant indicators which often differ from those on the official list of MDG indicators.

Dissimilarities in data between countries can reflect real differences in country achievements on different indicators. They can also reflect differences in definitions, estimation and calculation methods applied and data collection and compilation issues. These latter differences need to be clearly explained, so the data can be better understood.

This publication provides an overview on current practices in presenting metadata with MDG-related data. It aims to highlight the importance of metadata and the essential role they play in communicating and understanding data. This guide should contribute to building capacity in producing and using statistics, adding to the valuable guidance that UNSD and organizations of the Inter-agency and Expert Group on MDG Indicators (IAEG-MDG) provide to assist countries in the production, dissemination and use of MDG-related statistics.

Providing sufficient metadata with publication of MDG-related data will help to explain differences between estimates based on various data sources. For example, the discrepancies that often exist between national and international estimates due to adjustments made for comparability purposes.

⁵ Refer to the United Nations official MDG database and website for more information: mdgs.un.org/unsd/mdg/Host.aspx?Content=Indicators/About.htm.

This publication is divided into the following sections:

What are metadata?

A definition and brief overview of the role metadata play in managing and interpreting statistical information.

Producing and managing metadata

Explains the important role metadata play throughout the statistical production process and outlines some of the common challenges faced by statistical organizations in the management of metadata. An overview of international collaboration and examples of metadata management tools and standards is provided.

Presenting metadata

Presentations of data in tables, charts and maps should be accompanied by the basic information needed to understand the data, with more details provided in an annex or other products. The impact of the Internet on the presentation of data and metadata is explored. Providing presentation guidelines for producers of statistics is essential to increasing the quality of statistical products.

Metadata for tracking development progress

This section presents the basis for determining metadata required for correct interpretation and understanding of MDG-related data. Recommendations for minimum and optimum metadata to accompany all presentations of MDG-related statistics are provided.

Examples of current practice

Comparability issues and metadata requirements are provided for a selection of official MDG indicators, along with examples of current practice in presenting data and metadata.

Recommendations

This publication provides several recommendations to improve the quality of metadata in MDG-related reports and products.

II. What are metadata?

Metadata are data that define or describe other data⁶. 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.

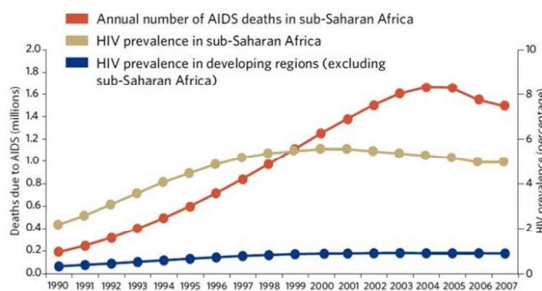
As the example in Figure 1 demonstrates, data make no sense when presented on their own. Metadata provide the information needed to understand what the values represent.

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

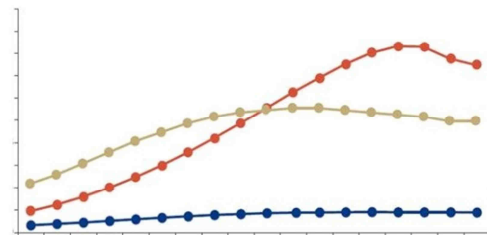
HIV prevalence in adults aged 15-49 years in developing regions and in sub-Saharan Africa (Percentage) and number of AIDS deaths in sub-Saharan Africa (Millions), 1990-2007



Source of original chart: United Nations, 2008.

Without metadata

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



The example “without metadata” in Figure 1 may be a little extreme. However, it is easy to find presentations that have essential pieces of information missing, such as no axis labels on charts, no source specified or use of technical terms and abbreviations that are not defined. Missing metadata impacts on the user’s ability to interpret and use what is being presented and it impacts on the value and trust placed in official statistics. In fact, a study of availability and comparability of MDG data in West Africa, revealed that the lack of metadata was one of the main weaknesses of national statistical systems in that region (Eurostat, 2010).

Including sufficient metadata is particularly important for reporting on MDG indicators, which often have multiple data sources and typically attract a broad audience who may have limited background in interpreting statistics.

Differences in MDG estimates compiled by countries often relate to the use of different definitions and concepts and varying practices in data collection and processing. Even where these differences are minimal, the resulting data might be quite dissimilar

⁶ Definition of metadata from the Metadata Common Vocabulary (SDMX, 2009b).

(OECD, 2007). Metadata make it possible to understand the limitations of a data point and its relation to other data. They allow the user to make judgements on the comparability of data from different sources and methods. For example, metadata should be sufficient to enable a user to compare estimates for an indicator that come from two different primary data sources (e.g. census and household survey) and compare those estimates between countries and over time.

Figure 2. Example from MDG reports that lack sufficient metadata

Indicator 21: Mortality related to external causes in different age groups (ages 0-4, 5-14, 15-19)

1995	2000	2002	2003
0-4/5-14/15-19 67/118/159	0-4/5-14/15-19 32/64/121	0-4/5-14/15-19 43/60/103	0-4/5-14/15-19 43/33/86

Source: [REDACTED]

The example above shows mortality statistics from a national MDG report that are both difficult to understand, due to the presentation style, and to compare, given the lack of metadata on primary data source and calculation method.

Types of metadata

According to terminology agreed to describe types of statistical metadata (SDMX, 2009b), there are two types of metadata: *structural metadata* and *reference metadata*.

Structural metadata identify and describe data, so they can be found 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: *conceptual metadata*, describing the concepts being measured; *methodological metadata*, describing the methods used to generate the data, such as sampling and collection methods; and finally, *quality metadata*, describing the quality dimensions of the data, such as timeliness and accuracy.

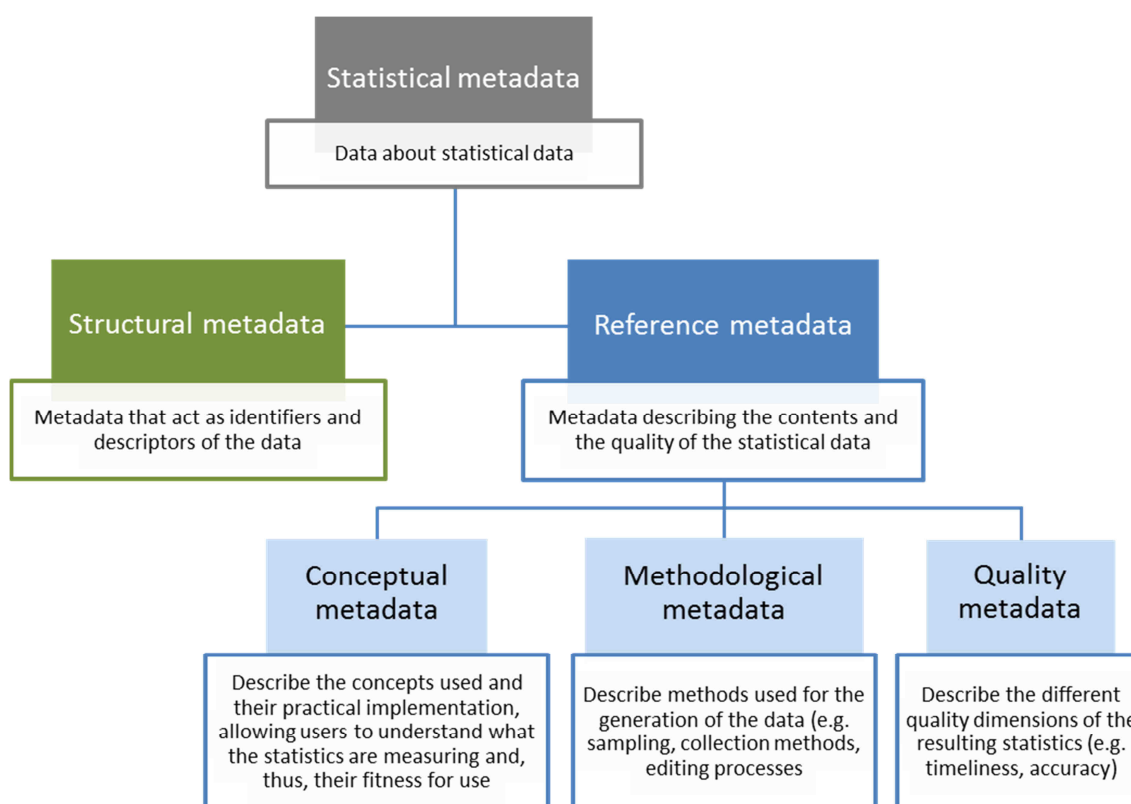
This publication is mainly concerned with reference metadata and how they are presented to users when data are published.

The presentation of metadata is often restricted to the name or broad definition of the indicator presented, the year to which the data refer, the units in which they are expressed and the source. It is important to note that metadata are more. As shown in Figure 3 below, reference metadata include description of concepts, methods and quality dimensions. This information makes it possible to understand what the statistics are measuring. Metadata place data into a context and help users to judge their comparability and reliability.

A number of excellent resources already exist to guide countries in the management and presentation of metadata (OECD, 2007; UNECE, 2000a, 2009a, 2009b). They make clear recommendations on the minimum metadata that should be provided with each data

presentation. Unfortunately, these recommendations are not always followed in MDG reports and other presentations of MDG-related data. The nature of MDG reporting by a wide range of actors within and outside the national statistical system and the relatively low statistical capacity in developing countries may contribute to poor metadata quality.

Figure 3. Types of statistical metadata



Source: *Metadata Common Vocabulary (SDMX, 2009b)*.

This publication provides guidance to countries on the minimum and optimal metadata to be presented with MDG estimates. The recommendations are based on existing metadata standards and agreed terminology.

Managing metadata is a field of expertise that cuts across all areas of statistics production. Therefore, this publication also takes the opportunity to provide some background on standards and international developments in this area.

International collaboration relating to statistical metadata

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

In the UNECE region, collaborative efforts in this area has been facilitated through the working group on Statistical Metadata – known as METIS – since the 1980s. The work and strategic direction of this group is managed by the High-Level Group on the Modernisation of Statistical Production and Services (HLG), under the Conference of European Statisticians (CES).

The METIS group has developed the *Common Metadata Framework* (UNECE, 2012), which provides a portal to information related to managing statistical metadata throughout the statistical production process.

More information on the common challenges and solutions to producing and managing metadata is provided in the next section.

III. Producing and managing metadata

Metadata are produced and used at all stages of the statistical production process. As the Generic Statistical Business Process Model (GSBPM) below demonstrates, producing statistics involves a number of generic steps or processes, regardless of subject matter. At each of these steps, metadata are created and re-used to drive, inform and monitor the production process.

Figure 4. Generic Statistical Business Process Model version 4.0 (UNECE, 2009b)

Quality Management / Metadata Management								
1 Specify Needs	2 Design	3 Build	4 Collect	5 Process	6 Analyse	7 Disseminate	8 Archive	9 Evaluate
1.1 Determine needs for information	2.1 Design outputs	3.1 Build data collection instrument	4.1 Select sample	5.1 Integrate data	6.1 Prepare draft outputs	7.1 Update output systems	8.1 Define archive rules	9.1 Gather evaluation inputs
1.2 Consult & confirm needs	2.2 Design variable descriptions	3.2 Build or enhance process	4.2 Set up collection	5.2 Classify & code	6.2 Validate outputs	7.2 Produce dissemination products	8.2 Manage archive repository	9.2 Conduct evaluation
1.3 Establish output objectives	2.3 Design data collection methodology	3.3 Configure workflows	4.3 Run collection	5.3 Review, Validate & edit	6.3 Scrutinize & explain	7.3 Manage release of dissemination products	8.3 Preserve data and associated metadata	9.3 Agree action plan
1.4 Identify concepts	2.4 Design frame & sample methodology	3.4 Test production system	4.4 Finalize collection	5.4 Impute	6.4 Apply disclosure control	7.4 Promote dissemination products	8.4 Dispose of data & associated metadata	
1.5 Check data availability	2.5 Design statistical processing methodology	3.5 Test statistical business process		5.5 Derive new variables & statistical units	6.5 Finalize outputs	7.5 Manage user support		
1.6 Prepare business case	2.6 Design production systems & workflow	3.6 Finalize production system		5.6 Calculate weights				
				5.7 Calculate aggregates				
				5.8 Finalize data files				

For example, national legislation may mandate a national statistical office to conduct a labour force survey to ascertain the share of women and men in wage employment in the non-agricultural sector (MDG indicator 3.2), amongst other things. This would be represented by step 1.1 in GSBPM and associated metadata would include reference to the Statistics Act that established the need for this information. Then, during the “design” phase (phase 2 in the GSBPM), each concept being measured must be clearly defined and the collection methodology must be developed and documented.

Further along the process, data collection (phase 4) takes place. Important metadata captured during this stage of production would include dates of data collection and non-response rates. Towards the end of the statistical production process, the results are disseminated (phase 7). Here a range of metadata is needed to explain what has been produced. Details captured earlier, such as definitions, methodology and non-response rates, would be re-used and presented to assist the reader in understanding the final output.

No matter what statistics are being produced, systems and processes are needed to effectively manage metadata throughout the production cycle. International

recommendations have been developed to guide statistical organizations in this area, for example, on metadata standards, management principles and systems, such as central metadata repositories.

Managing metadata: challenges and initiatives

Effective metadata management within statistical organizations allows the metadata to be available and re-used whenever it is needed. 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
- producing accurate metadata and data, thereby increasing the quality of statistics
- streamlining the design and build of collection mechanisms, by maintaining tested and previously used questions in a single location
- capitalising on lessons learned from past collections and feeding that into improvements in the next cycle
- high morale and productivity as staff can store and retrieve the information they need
- data users encouraged by clear information needed to understand and interpret the data
- increased data use and trust in official statistics.

As part of the Common Metadata Framework (UNECE, 2012), the METIS group has developed some governing principles to good metadata management. These include:

Capture metadata at the source: given the number and variety of steps involved in producing statistics, it is essential to capture metadata **as soon as it is available**. For example, information about the data source(s) should be presented when the statistics are published, allowing the data user to understand and interpret it accurately. If this information is not captured and stored effectively when the data are being collected, time will be wasted going back to find and accurately document it at a later stage. This wastes time, is frustrating for the statistician and important details may be lost.

Single source: the same metadata will be captured and used by different people across the organization. The latest, authoritative source of metadata should be maintained in a single location so that everyone is using the correct version. For example, the same definition for household would be used by all subject-matter statisticians that use household surveys as a basis for collection.

Version control: earlier versions of metadata should be kept along with information about the changeover, such as date and reason. For example, when classifications such as the one for identifying occupations are revised, a copy or link to the previous version should be stored. Reference to this should be kept together with the new classification, as well as details such as date of change, reasons, description of the main change(s)

introduced by the revised classification and notes on any break in series. This will enable metadata users to access important details to produce and use statistics that use this classification.

All 16 core principles for metadata management are detailed in Box 1.

Box 1. Core Principles for Metadata Management

Handling	Statistical business process model: Manage metadata with a focus on the overall statistical business process model (www.unece.org/stats/gsbpm).
	Active not passive: Make metadata active to the greatest extent possible. Active metadata are metadata that drive other processes and actions. Treating metadata this way will ensure they are accurate and up-to-date.
	Reuse: Reuse metadata where possible for statistical integration as well as efficiency reasons
	Versions: Preserve history (old versions) of metadata.
Authority	Registration: Ensure the registration process (workflow) associated with each metadata element is well documented so there is clear identification of ownership, approval status, date of operation, etc.
	Single source: Ensure that a single, authoritative source ('registration authority') for each metadata element exists.
	One entry/update: Minimize errors by entering once and updating in one place.
	Standards variations: Ensure that variations from standards are tightly managed/approved, documented and visible.
Processes	Integrity: Make metadata-related work an integral part of business processes across the organization.
	Matching metadata: Ensure that metadata presented to the end-users match the metadata that drove the business process or were created during the process.
	Describe flow: Describe metadata flows within and between statistical business processes (alongside data flows and business logic).
	Capture at source: Capture metadata at their source, preferably automatically as a bi-product of other processes.
	Exchange and use: Exchange metadata and use them for informing both computer based processes and human interpretation. The infrastructure for exchange of data and associated metadata should be based on loosely coupled components, with a choice of standard exchange languages, such as XML.
Users	Identify users: Ensure that users are clearly identified for all metadata processes, and that all metadata capturing will create value for them.
	Different formats: The diversity of metadata is recognised and there are different views corresponding to the different uses of the data. Different users require different levels of detail. Metadata appear in different formats depending on the processes and goals for which they are produced and used.
	Availability: Ensure that metadata are readily available and useable in the context of the users' information needs (whether an internal or external user)

Statistical metadata systems

Statistical metadata systems allow metadata to be captured and stored for retrieval when required. Ideally, they will comprise central metadata repositories, where metadata are stored and maintained in one location and used by subject matter experts when required.

“Statistical metadata systems play a fundamental role in statistical organizations. Such systems comprise the people, processes and technology used to manage statistical metadata.”

Statistical Metadata in a Corporate Context: A guide for managers (UNECE, 2009c)

The challenge lies in providing efficient systems that ensure people involved in each step of the process create and re-use metadata and embrace this as an essential and useful part of production processes.

Tools for managing metadata

Statistical organizations usually have a range of tools for managing metadata. These may include a *concepts management system*, such as the one developed by Statistics Portugal (2009). This system provides a central database to hold definitions of statistical concepts used across the organization and establish links between them. This database is not only valuable as an internal metadata management tool and a system to harmonize definitions across collections, but also forms the basis for providing metadata to users through the Statistics Portugal website.

Another common tool for managing metadata is a *statistical classification system*, where classifications and code lists are maintained in a central location for use across the organization. For example, the national statistical office of the Czech Republic has developed SMS-CLASS, as central system based on the Neuchatel model of statistical classifications. “It allows creation, storage, update and use of statistical classifications, which are necessary for data processing. There is basic meta-information kept on each classification including its history, e.g. the title and coordinator of classification, validity and contents of classification/code-list in language versions” (Czech Statistical Office, 2009).

Initiatives to facilitate sharing of software between statistical organizations are ongoing. Intergovernmental meetings on statistical metadata and the management of statistical information systems provide a valuable forum for exchanging experiences and ideas.

The UNECE Conference of European Statisticians has also established a Sharing Advisory Board that monitors strategic developments around collaboration and the sharing of tools for statistical production, including metadata management tools⁷.

⁷ More information is available from www1.unece.org/stat/platform/display/msis/Software+Sharing.

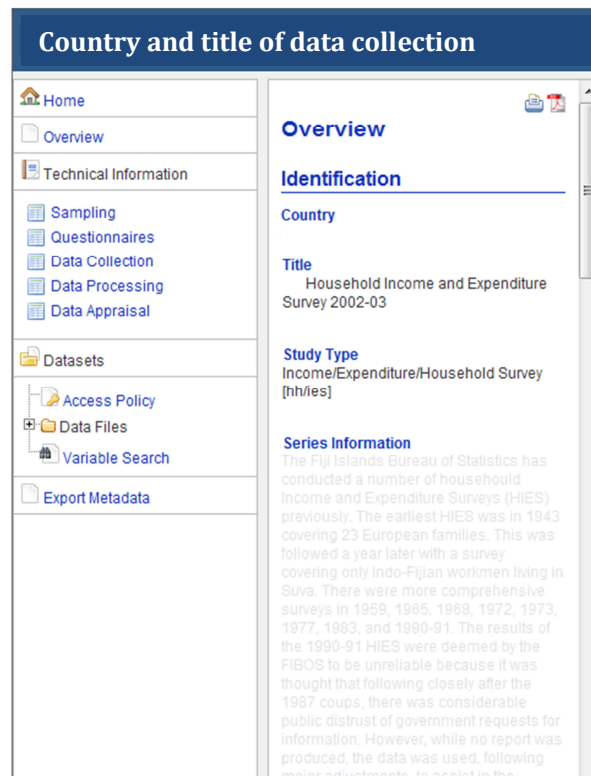
Tools for disseminating metadata

The International Household Survey Network (www.ihsn.org) has developed a tool to catalogue and disseminate survey and census metadata together with the resulting microdata sets.

The National Data Archive (NADA) system has been implemented by a number of national and international statistical organizations (Figure 5 shows a screen shot of one implementation of NADA in the Pacific region). NADA can store standard metadata about each data collection, such as:

- **Sampling:** procedure followed; deviations from sample design; response rate; and weighting
- **Questionnaires:** structure of the questionnaire used and copies of the actual forms
- **Data collection:** dates and reference periods; collection mode; notes on the process followed; data collectors; supervision arrangements
- **Data Processing:** editing approach and other relevant information
- **Data Appraisal:** estimates of sampling error.

Figure 5. Metadata capture and dissemination tool developed by International Household Survey Network



Source: Secretariat of the Pacific Community (SPC) (www.spc.int/prism/reports/data-catalog).

NADA is a freely available, open source tool, compliant with metadata standards Data Documentation Initiative (DDI) and Dublin Core (Dupriez, O. and Greenwell, G., 2007). Statistical organizations can use it as a catalogue of their data collection activities and as a publishable online database of associated metadata. Links can then be provided so interested data users can find the information they need to interpret the data.

In addition to capturing metadata, NADA is designed to be used as a microdata dissemination tool.

Metadata standards

National and international statistical organizations face common challenges in managing their metadata, and there has been significant effort to develop shared metadata standards and models.

These include:

- Statistical Data and Metadata Exchange (SDMX)
- Data Documentation Initiative (DDI)
- Metadata registries ISO 11179
- Metadata Common Vocabulary (MCV)
- Neuchâtel Model.

A brief description of these examples is provided below. The UNECE Common Metadata Framework (www.unece.org/stats/cmf) (UNECE, 2012) provides further details of these and other metadata standards, models and guidelines.

Statistical Data and Metadata Exchange (SDMX)

The Statistical Data and Metadata Exchange (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, the European Central Bank, Eurostat, the International Monetary Fund, the Organisation for Economic Cooperation and Development (OECD), the United Nations and the World Bank.

Data Documentation Initiative

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 registries ISO 11179

This is a standard for describing and managing the meaning and representation of data. The basic semantic unit is a concept. Both DDI and SDMX are based on ISO/IEC 11179 for their descriptions of data and their use of concepts as a basic semantic unit.

Metadata Common Vocabulary (MCV)

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 MCV.

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

Figure 6. The definition of “Data” in the Metadata Common Vocabulary

76. Data	
Definition:	Characteristics or information, usually numerical, that are collected through observation.
Context:	Data are the physical representation of information in a manner suitable for communication, interpretation, or processing by human beings or by automatic means (Economic Commission for Europe of the United Nations (UNECE), "Terminology on Statistical Metadata", Conference of European Statisticians Statistical Standards and Studies, No. 53, Geneva, 2000). Statistical data are data derived from either statistical or non-statistical sources, which are used in the process of producing statistical products.
Source:	The International Statistical Institute, "The Oxford Dictionary of Statistical Terms", edited by Yadolah Dodge, Oxford University Press, 2003
Hyperlink:	
Related Terms:	Characteristic Coverage Data analysis Data presentation Metadata Periodicity Special Data Dissemination Standard, SDDS Timeliness

Source: *Metadata Common Vocabulary (SDMX, 2009b)*.

Neuchâtel Model

Version 2.1 of the Neuchâtel Terminology Model Classification (2004) provides a common language and perception of the structure of classifications and the links between them. In 2006 the model was extended with variables and related concepts. The discussion includes concepts like object types, statistical unit types, statistical characteristics, value domains, populations etc. These two models together claim to provide a more comprehensive description of the structure of statistical information embodied in data items.

Metadata standards, models and guidelines form a valuable basis for statistical organizations to develop their data and metadata management systems. Compliance with international standards leads 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.

Effectively managing metadata throughout the statistical production process is the first step in ensuring sufficient information can be provided to data users. With adequate metadata now available, the next challenge is to ensure it is consistently presented in a way that can be easily understood.

IV. Presenting 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 broad range of audiences can understand and use the valuable information contained within.

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 remain relevant in the information age, statistical organizations are now placing greater emphasis on publishing data and metadata in a variety of forms to reaching a growing audience of data users.

A number of resources already exist to guide national statistical organizations in their presentation of metadata. The OECD handbook on *Data and Metadata Reporting and Presentation* (2007) provides detailed recommendations on the types of metadata that should be presented when reporting data. The UNECE *Making Data Meaningful* guides (2006, 2009) and *Guidelines for Statistical Metadata on the Internet* (2000a), help organizations to communicate statistics in tables, charts and maps and writing about numbers for a broad audience.

There are also several guides on producing MDG indicators (United Nations, 2003; WHO, 2006; ILO, 2009) that help, not only producers, but users of MDG-related data to understand and interpret the estimates correctly. These guides explain how each MDG indicator is calculated, provide definitions of the associated concepts and describe the limitations and typical data quality issues.

A particular challenge for including sufficient metadata with MDG-related data is that national reports are usually prepared by non-statisticians, who may not be familiar with how to produce and present 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 practice before they are published.

Existing guidelines and resources form the basis for the recommendations in this handbook. These are aimed at all organizations involved in reporting on progress towards the MDGs, both within and outside national statistical systems.

Data should stand alone

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. Ensuring presentations contain all the information needed to interpret the data means they can be understood at a glance and users are much more likely to absorb and apply the findings correctly.

Sufficient metadata would include:

- **A clear title** that describes the data series, population, coverage and reference period
- **Labels** to describe the data, such as variable names and units of measurement, using words that can be easily understood
- **Footnotes** that include information needed to interpret the data accurately, such as definitions, excluded populations and other exceptions
- **Source of the data**, such as the collection method, the organization that conducted it and the dates of collection.

Different presentations for different audiences

The extent to which detailed metadata are included in the presentation of data will depend upon the target audience and the form in which the information is being published. Data users vary in their knowledge of statistics from people who are unfamiliar, and often uncomfortable with data, to expert users and statisticians themselves.

Figure 7. MDG Report of Bosnia and Herzegovina (2010) provides metadata in an annex

64 Progress towards the Realization of the MILLENNIUM DEVELOPMENT GOALS in BH 2010

Objective/Target/Indicator	Baseline 2000/2001	2007		2009 or latest available data	2015 Targets MDGs (unless otherwise indicated)*	Sources	Comments	Progress towards 2015 Targets**
		Targets as in PRSP	Achieved					
a	b	c	d	e	f	g	h	i
2.2. Percentage of the number of students enrolled into 1st grade reaching the 5th grade	total	99.0	99.8	99.8	100.0	b) MICS 2000 d) MICS 2006 e) Ibid.	Data in 2007 column refer to 2005-2006 period; indicator 57 according to MICS. The missing target for 2015 is suggested here, taking into consideration other targets and data trend.	likely
	men	99.2	99.6	99.6				
	women	98.8	100.0	100.0				
	total	99.6	n/a	99.2				
2.3.1. Literacy rate of the 15-24 age group - %						BH Common Country Study Development	Data in 2007 column refer to 2005-2006 period; indicator 60 according to MICS. The missing target for 2015 is suggested here, taking into consideration other targets and data trend.	likely
Additional indicator						2003 data	d1) 79.3% refers to net attendance rate - MICS indicator 56, page 106. Net primary school completion rate is 86.6% while the primary/secondary school transfer rate is 92.7%. e) Enrolment rate for 2009 has been calculated based on a number of persons that enrolled secondary school and the population age structure according to the 2009 Labour Force Survey.	potentially
2.4. Secondary rate %						Attendance Rate, FBH/SI, RSSI and n.	e2) Revised estimate based on change in	likely

Source: Progress towards the realization of the Millennium Development Goals in Bosnia and Herzegovina 2010 (Bosnia and Herzegovina Ministry of Finance and Treasury and United Nations Country Team in Bosnia and Herzegovina, 2010).

Reports on progress towards MDGs tend to be aimed at a broad audience of users not necessarily familiar with statistics. Metadata are often limited to a minimum of detail so as not to overwhelm or confuse the user with too much information. However, this

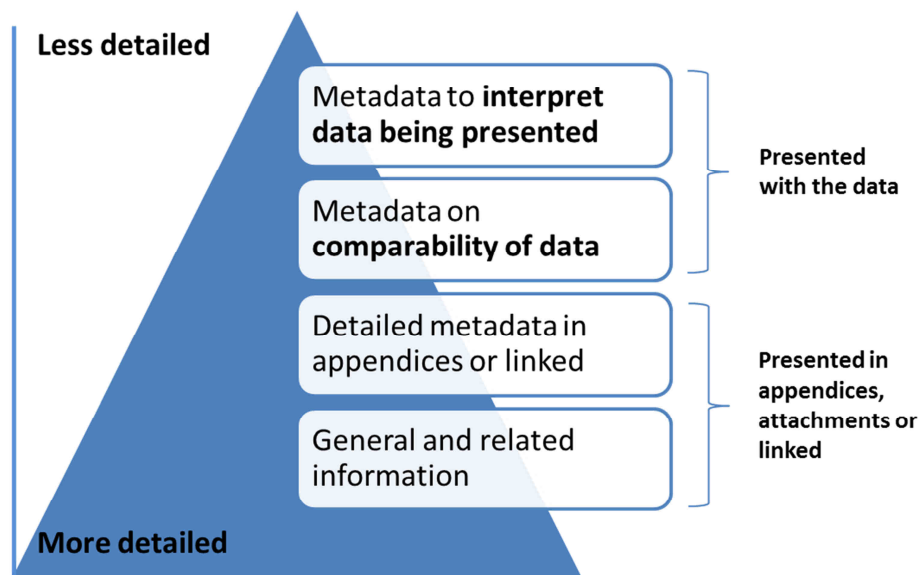
approach risks disconnecting data with the information needed to interpret it correctly. A common solution to this dilemma is to include metadata in an annex to the report, or provide references and links to explanatory materials.

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.

Metadata can be divided into four levels of detail:

1. Metadata needed to interpret the data presented in a table, graph or text
 - Title, units, reference period, etc.
 - Important information about comparability, e.g. break in series/change in definition of data source that significantly influences the comparability.
2. Metadata needed for comparability with other data for the same indicator (from other countries or other data from the country itself that are not shown) or needed for the interpretation of the data in wider context. This also refers to information regarding the reliability of the data. For example, where reference periods of geographical areas differ for particular data values.
3. More detailed metadata that are relevant but that do not have an influence in the interpretation or the comparability of the data.
4. Other general information related to the data series being presented.

Figure 8. Different levels of detail when presenting metadata



Distinguishing between each level of metadata and whether it should be presented alongside the data or in appendices and other publications is a subjective decision.

Much will depend on the type of data being presented and the target audience(s). The aim should be to aid understanding, prevent misuse but not overwhelm users with details.

The amount of detail may increase as the level of metadata deepens, although as the information becomes more detailed it would likely become more general in nature. For example, level 1 metadata would relate specifically to the data values being presented, whereas level 4 metadata may relate generally to all data in that series, such as sampling information or collection methods.

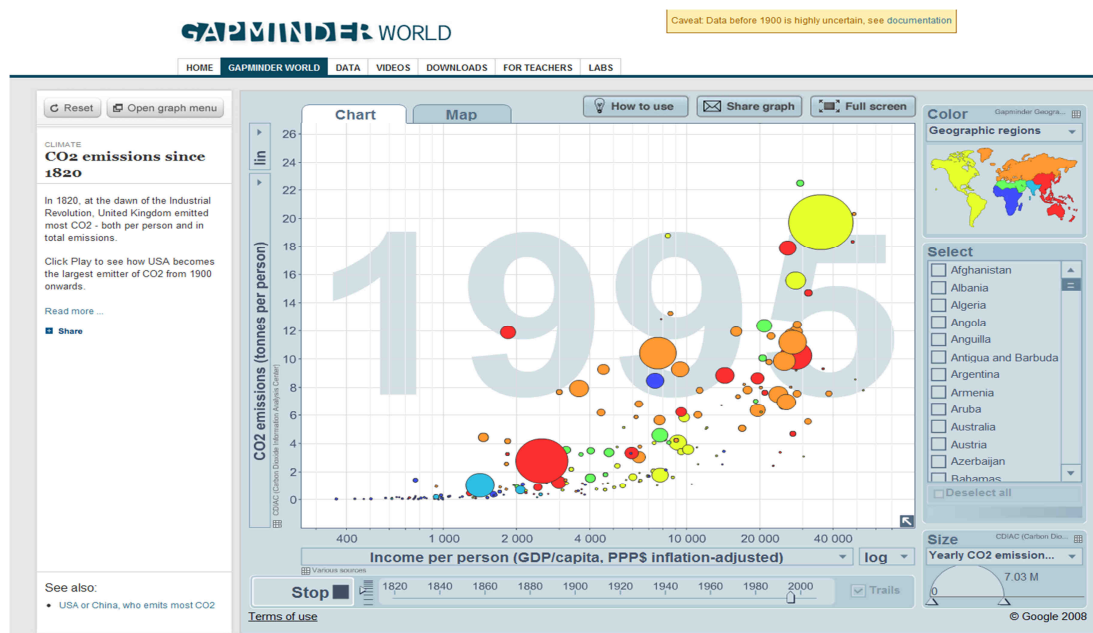
New possibilities for presenting metadata

The Internet and associated technologies have had a significant impact on the way that official statistics are now disseminated and used. Data can be published online quickly and cheaply like never before.

Online databases – that allow users to build their own queries and extract data in the form that they need it – are relatively inexpensive and easy to establish. So too are tools that visualize data in animated charts and maps, making it both fun and interesting to use statistics.

Similarly, the presentation of metadata has been revolutionised by the Internet, with the possibility to link to searchable glossaries, hover over terms for instant definition and even provide videos to describe statistical methodology and tools.

Figure 9. Online tools like Gapminder make it possible to visualize and interact with MDG-related data and metadata



Source: Gapminder World (www.gapminder.org).

However, disseminating statistical information online provides new challenges as well as possibilities. Data producers must ensure that metadata are continually updated and that they remain with the data as they are downloaded and transformed into different formats.

For example, additional metadata, such as “date of last change” (reflecting the last date and time the online data set was updated) become important for online data presentation. Furthermore, footnotes and source information need to be clear to the data user on the screen, as well as in any spread sheets or formats that can be downloaded from the database.

The UNECE Statistical Database provides metadata on definitions and sources to its users as they generate and view data on the screen. When this information is downloaded into spread sheets and other formats, the same metadata are included.

Figure 10. Metadata in the UNECE Statistical Database

The screenshot displays the UNECE Statistical Database interface. The main content area shows data for 'Hunger by Reporting level, Indicator, Country and Year' for the years 2000 and 2005. The data is presented in a table with columns for the indicator, country, and year. The table includes data for International, Albania, Azerbaijan, Armenia, and Bosnia and Herzegovina. The values for 2000 and 2005 are shown in the rightmost columns of the table.

	2000	2005
International		
Children under 5 moderately or severely underweight (%)	17.0	6.6
Albania	14.0	
Azerbaijan	2.6	
Armenia	4.2	
Bosnia and Herzegovina		
Children under 5 severely underweight (%)		
Albania	7.8	
Azerbaijan	5.8	
Armenia	0.5	
Bosnia and Herzegovina	1.2	

The metadata section on the right provides detailed information about the indicators and data sources. It includes a 'General note on the UNECE MDG Database' and a 'Definition of the indicators' section. The definition for 'Children under 5 moderately or severely underweight (%)' is provided, along with a reference to the WHO Child Growth Standards. The metadata section is highlighted with an orange box, and an arrow points from the caption to this section.

Source: UNECE Statistical Database (w3.unece.org/pxweb).

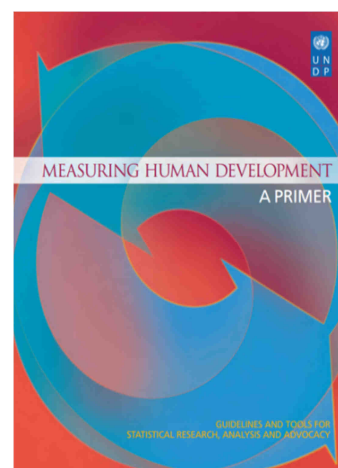
User Guides

Statistics are complicated. Data users often appreciate clear guidance on how to interpret and use the information contained within. Providing a “user guide” can be a valued contribution to ensuring MDG-related data are interpreted correctly and used in policymaking.

For example, the United Nations Development Programme (UNDP) – *Measuring Human Development: a primer* (2007) – provides more than 150 pages of information to guide users of the UNDP Human Development Reports. It includes information on “common pitfalls in comparing data”; “a variety of data sources”; and “constructing composite indices”.

Regardless of their level of expertise, explaining technical terms used in the presentation of statistics is helpful for all data users – confirming for more experienced users what the national definitions of familiar terms are, and guiding those who may be seeing the term used for the first time. In 2005 MDG report of Latvia, such explanations were provided in a separate box within the report (Figure 12).

Figure 11. Cover of the UNDP guide on measuring human development



Source: UNDP, 2007.

Figure 12. Helping users in Latvia to understand technical terms

Box 6.1. Explanation of terms

External causes of death – injuries, accidents or violence, which have resulted in or contributed to a person’s death. External causes of death include suicides, falls, traffic accidents, alcohol poisoning, overdosing of narcotics or other harmful substances.

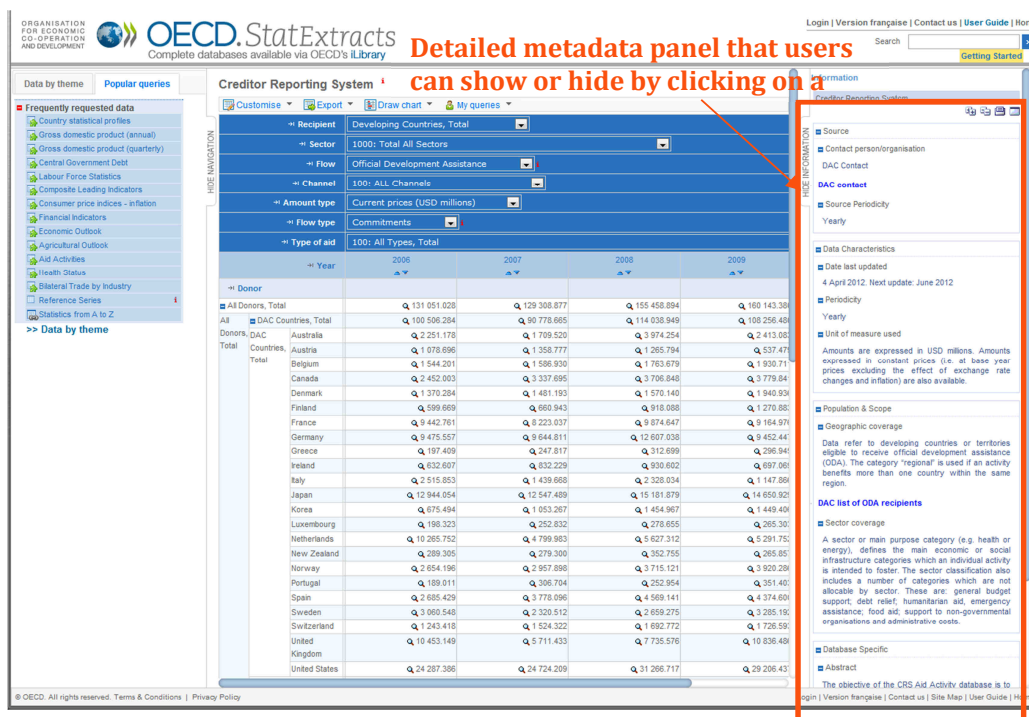
Diseases resulting from an unhealthy lifestyle – diseases caused by an unhealthy diet, working and living conditions, stress, sedentary lifestyle, and other harmful habits.

Source: *Life in 2015: the Latvia MDG Report* (Ministry of Foreign Affairs of the Republic of Latvia and the United Nations Development Programme, 2005).

The online statistics database of OECD – which manages the MDG data relating to official development assistance (Goal 8) – provides detailed metadata in a side panel that users show or hide with a simple click. Metadata include: definitions, source information, contact details, periodicity of data update, units of measure, geographic and population coverage, an abstract explaining the purpose of the data, and links to more detailed information (Figure 13).

Just as data users come from a range of backgrounds, with varying abilities to understand statistics, user guides are needed in a variety of formats. The Australian Bureau of Statistics provides its users with a range of resources to help them understand statistics (Figure 14). These include an online course on basic statistical concepts, video tutorials, a guide for policymakers, frequently asked questions, and more.

Figure 13. Presenting metadata in the OECD Statistical Database



Source: OECD StatExtracts (stats.oecd.org).

Figure 14. The Australian Bureau of Statistics provides a range of guides for data users



Knowing how to understand and use statistics is vital for informed decision-making. This is why the ABS is committed to increasing the level of statistical literacy in the community. This section contains a variety of tools and resources to assist you to understand, interpret and evaluate statistical information.

- Why statistics matter**
What does it mean to be statistically literate and why is it more important than ever?
- Introduction to statistics**
An online course covering basic concepts and principles of statistical thinking.
- ABS training courses**
Check out the ABS Training Courses available in your State or Territory.
- Online resources**
Resources on a range of topics to assist you to analyse and understand ABS data.
- Online video tutorials**
Online tutorials to help you to understand and use ABS information.
- Online presentations**
A range of presentations on a range of topics.
- Information for you**
Information and resources to assist those working in Education, Media, Business and Gov't.
- Statistical Language**
Learn about a range of statistical concepts
- Statistically Speaking Blog**
Check out our blog - it's all about statistics.

New! Revised Guide for Using Statistics for Evidence Based Policy

New! Frequently Asked Questions

New! Interact with our Data

Source: Australian Bureau of Statistics
abs.gov.au/websitedbs/a3121120.nsf/home/Understanding%20statistics

Providing guidelines for statisticians

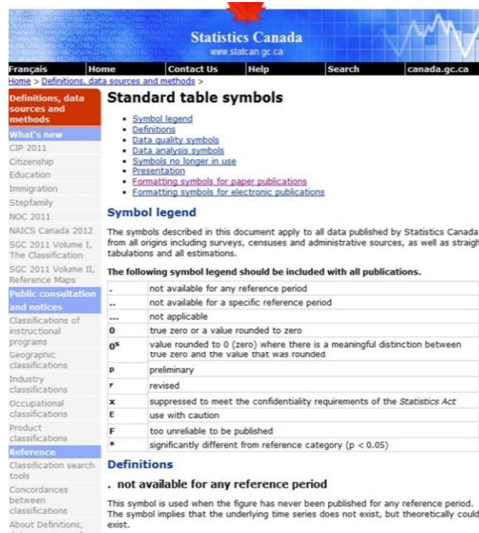
Statistical organizations should have policies and guidelines that instruct staff on how to present statistical data and metadata in the reports and products they release. Such guidelines need to prescribe the format data and metadata must be in before it is published, for example, the layout of tables, charts and maps and metadata that must be included. Templates may also be used to standardise the presentation of information.

Publication guidelines, and the processes that ensure they are followed, will lead to complete and consistent metadata being presented with all statistics published.

The UNECE guidelines on presenting statistics – *Making Data Meaningful Part 2: a guide to presenting statistics* (UNECE, 2009a) – suggest that all tables include at least:

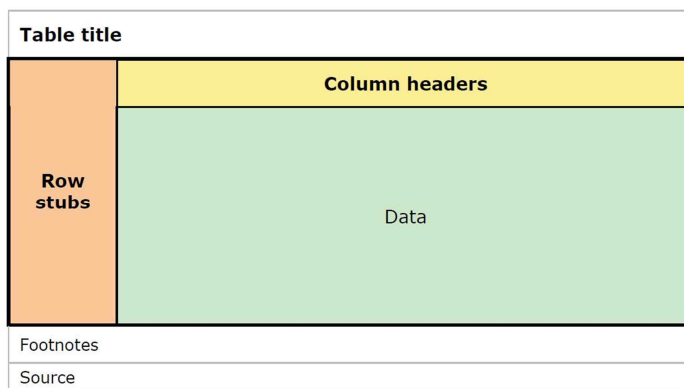
- a title (giving a clear and accurate description of the data, answering the questions “what”, “where” and “when”)
- column and row labels
- footnotes and
- the data source(s).

Figure 15. Example of publication guidelines



Source: Policy on use of standard table symbols (Statistics Canada, 2012). (www.statcan.gc.ca/concepts/definitions/guide-symbol-signes-eng.htm).

Figure 16. Standard components of statistical tables



Source: *Making Data Meaningful Part 2: a guide to presenting statistics* (UNECE, 2009a).

Figure 17. Metadata in a statistical table published in the Kazakhstan MDG Report 2005

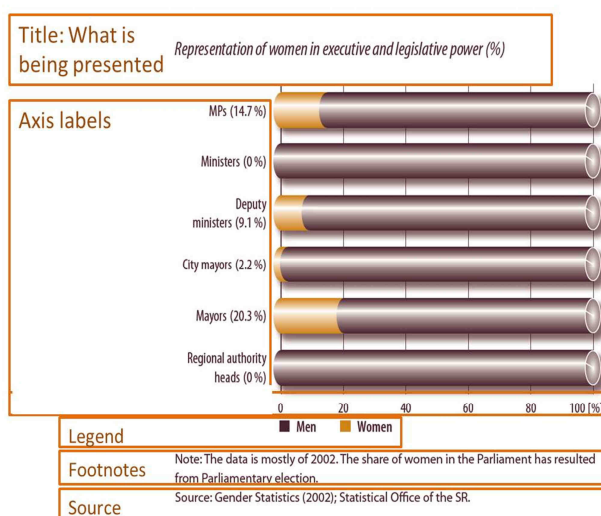
What is being presented		Where		When				
Table 1.5. Proportion of people with income below the food basket in Kazakhstan 1997-2004, %								
	1997	1998	1999	2000	2001	2002	2003	2004
Proportion of people with income below the food basket	12.7	16.2	14.5	11.7	11.7	8.9	6.3	4.3
Source: Agency on Statistics of the Republic of Kazakhstan ³³								
Source								
³³ Poverty Monitoring Indicators in Kazakhstan. Almaty: Agency on Statistics of the Republic of Kazakhstan, UNDP, 2003. Living Standards and Poverty in Kazakhstan. Statistical Monitoring. Almaty: Agency on Statistics of Kazakhstan, UN TG on Poverty Alleviation 2004. Living Standards in the Republic of Kazakhstan. Statistical Monitoring. Ed. by Y.K. Shokamanov. Almaty, 2005.								
Footnotes								

Source: Millennium Development Goals in Kazakhstan, 2005.

Similarly, charts (or graphs) should contain the following essential metadata:

- **Title** – to “what”, “where” and “when” do the data relate
- **Axis labels** – identify the values displayed in the chart
- **Axis titles** – identify the units of measure on each axis
- **Legend and data labels** – identify the symbols, patterns or colours used in the chart. A legend is not necessary when only one series of values is represented in the chart. Data labels displayed on or next to the data components (e.g. lines, bars) may be used in place of a legend.
- **Footnotes** – provide additional information needed to understand the data
- **Source** – the organization that produced the data, the data collection method (e.g. type of survey) and date of collection.

Figure 18. Metadata included with a chart in the MDG Report of Slovakia of 2004



Millennium Development Goals: Reducing Poverty and Social Exclusion. Slovak Republic. UNDP. 2004.

Statistical organizations that lack their own guidelines can use existing ones, like the UNECE *Making Data Meaningful* guides, to develop policies for the presentation of data with sufficient metadata.

V. Metadata for tracking development progress

Comparing MDG indicators across time and between countries is an integral part of the MDG framework. An estimate for a certain indicator for a specific year can often be based on more than one primary data source and use different estimation methods.

For monitoring MDGs, more data sources and a larger variety of methods tend to be used than in annual publications of the National Statistics Office. Demands for metadata are therefore particularly high. This becomes apparent, for example, when there are discrepancies between estimates from different sources. Metadata should explain the reasons for this.

Countries therefore face particular challenges in ensuring that sufficient metadata are included with reports on progress towards development goals. MDG-related data are produced by numerous sources and are often compiled into reports by non-statisticians, who may be unfamiliar with standards for data and metadata presentation. Furthermore, the demands for development-related data are high, but national coordination systems in developing countries are often weak, leading to multiple and inconsistent sources for the same indicators, and a lack of adequate metadata to explain the discrepancies.

Reports on MDG progress occur through two main streams:

- questionnaires circulated by international agencies responsible for maintaining international estimates of MDG indicators
- national reports compiled from available sources on the official and nationally relevant MDG indicators.

As there are many actors involved in both streams, at national and international levels, it is a challenge for developing countries to manage the statistical information and ensure consistent and coherent presentation of data and metadata.

Regardless of the sources and reporting channel for MDG-related data, there should be standards in place to ensure metadata are sufficient. This section looks at current practices and recommends mandatory, conditional and optional metadata that should accompany data reported on MDG-related indicators.

Determining metadata requirements for MDG indicators

Guidelines on producing MDG estimates are provided by IAEG-MDG⁸ which was established as an inter-agency working group of the United Nations Development Group. Directed by IAEG-MDG, the United Nations Development Group first published a

⁸ Coordinated by the UNSD, the IAEG-MDG comprises representatives of international agencies responsible for collating and producing reports on national progress towards MDG indicators. Members include: UNICEF, UNFPA, WHO, ILO, the World Bank, ITU, UNDP, UN Women, OECD, UNESCO, the United Nations Regional Commissions. Refer to mdgs.un.org/unsd/mdg/Host.aspx?Content=IAEG.htm for more details.

comprehensive handbook on producing MDG estimates in 2003. Entitled *Indicators for Monitoring the Millennium Development Goals* (United Nations, 2003).

The handbook provides the following metadata for each of the MDG indicators:

1. Definition.
2. Goal and target addressed.
3. Rationale.
4. Method of computation.
5. Data collection and source.
6. Periodicity of measurement.
7. Gender issues.
8. Disaggregation issues.
9. Comments and limitations.
10. References and international data comparisons.

Although it does not explicitly prescribe how metadata should be presented with MDG-related data values, this handbook does provide important guidance to countries on the types of metadata that are most relevant.

This handbook is currently under revision to incorporate new MDG indicators that were introduced in 2008. The electronic version of the updated handbook is published on a Wikipedia platform and it will be regularly updated to include new information when it becomes available.

Recommendations on metadata presentation for MDG-related data can be guided by the dissemination practices of agencies responsible for reporting on MDG indicators. The UNECE MDG Database, United Nations MDG Database (mdgs.un.org) and the UNDP website for the Human Development Report are good examples to consider.

Figure 19. Extract from the United Nations Handbook on producing MDG estimates

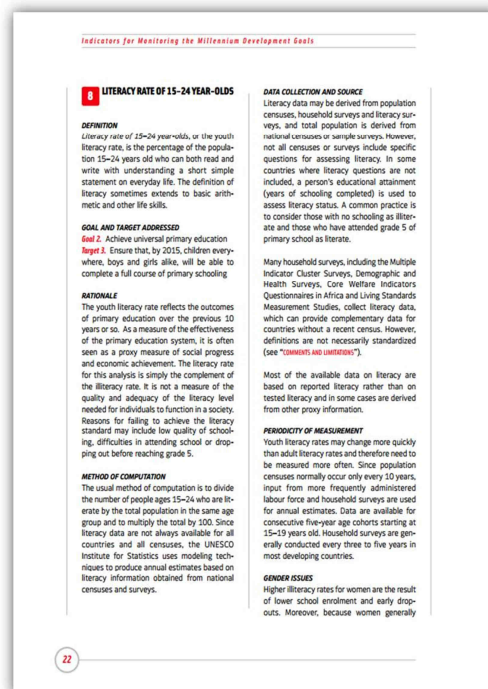


Figure 20. Electronic version of the updated handbook on Indicators for Monitoring the Millennium Development Goals

Source: mdgs.un.org/unsd/mi/wiki/

As Table 1 shows, the approaches to presenting metadata differ from agency to agency, with a different elements used by each, as well as varying terms to describe them. This report recommends international agencies involved in reporting on MDG progress consider agreeing on using a common approach to presenting metadata.

Table 1. Metadata presented with MDG-related data in United Nations databases

	Presented with data <i>Metadata presented alongside the data (i.e. in tables, charts, maps)</i>	Available via links <i>Metadata presented in appendices, accompanying documents and via links</i>
UNECE MDG Database w3.unece.org	<ol style="list-style-type: none"> 1. Series name 2. Footnotes relating to specific data values 3. General note on the UNECE MDG Database 4. Definitions 5. Latest update 6. Source 7. Unit of measurement 8. Data type (e.g. average) 	Glossary of terms

	Presented with data <i>Metadata presented alongside the data (i.e. in tables, charts, maps)</i>	Available via links <i>Metadata presented in appendices, accompanying documents and via links</i>
United Nations MDG Database mdgs.un.org	<ol style="list-style-type: none"> 1. Series name (usually including units of measurement) 2. Date last updated 3. Footnotes relating to specific data values where anomalies exist 4. General note on data adjustments performed by international agencies 	<p>Metadata available for each indicator series:</p> <ol style="list-style-type: none"> 1. Goal 2. Target 3. Indicator 4. Series name 5. Contact point 6. Definition 7. Method of computation 8. Comments and limitations 9. Process of obtaining data 10. Treatment of missing values 11. Data availability 12. Regional and Global estimates 13. Expected time of release
UNDP Human Development Report hdr.undp.org	<ol style="list-style-type: none"> 1. Series name (units of measurement) 2. Definitions 3. Main Data Source (year) 4. Date last updated (website only) 5. Footnotes (Notes) relating to specific data values 6. Explanation of symbols used 	<p>About the data:</p> <ol style="list-style-type: none"> 1. Data sources and contacts 2. Readers' guide 3. Technical notes 4. A range of papers and materials to assist with understanding the data <p>Readers' guide:</p> <ol style="list-style-type: none"> 1. Comparisons over time and across editions of the Report 2. Discrepancies between national and international estimates 3. Symbols <p>Technical notes:</p> <ol style="list-style-type: none"> 1. How the indices are calculated 2. Definition of regional groupings 3. Statistical references
World Health Organization (WHO) Global Health Observatory Data Repository apps.who.int/ghodata	<ol style="list-style-type: none"> 1. Series name 2. Unit of measure 	<p>Footnotes relating to specific data values.</p> <p>For each data series:</p> <ol style="list-style-type: none"> 1. Rationale 2. Definition 3. Associated terms 4. Unit of measure 5. Preferred data sources 6. Other possible data sources 7. Method of measurement 8. Method of estimation Method of estimation of global and regional aggregates 9. Disaggregation 10. Links

	Presented with data <i>Metadata presented alongside the data (i.e. in tables, charts, maps)</i>	Available via links <i>Metadata presented in appendices, accompanying documents and via links</i>
United Nations Children’s Fund (UNICEF) ChildInfo www.childinfo.org	<ol style="list-style-type: none"> 1. Last update 2. Series name 3. Units of measure 4. Source 5. Description of acronyms 6. Footnotes on comparability issues 	<ol style="list-style-type: none"> 1. Methodology 2. Disparities 3. Related publications and background reading
International Labour Organization (ILO) LABOURSTAT Database laborsta.ilo.org	<ol style="list-style-type: none"> 1. Series name 2. Units of measure 3. Source (for each country) 4. Notes 	<ol style="list-style-type: none"> 1. Definitions 2. Classifications (e.g. ISIC-Rev3.) 3. Abbreviations and symbols 4. Sources and methods (detailed information for each country)

Recommendations for presenting metadata with MDG-related data

Monitoring of progress towards development goals will be enhanced with guidance to countries on which metadata should be included or referenced in their MDG reports and databases.

Metadata standards should play an important role in recommending what information should be presented to describe MDG-related data. Being focused on dissemination, SDMX – the international standard for data and metadata exchange – is particularly relevant in this regard. Conforming to SDMX involves using terminology that is consistent with the SDMX cross domain concepts (SDMX, 2009a).

For many countries in the UNECE region, the Euro-SDMX Metadata Standard (ESMS) is the emerging standard for presenting metadata. As the title suggests, the standard is based on SDMX, recommending a set of metadata elements using terms from a harmonized list.

“Considerable efficiency gains can be reached when the reference metadata are produced on the basis of a harmonised list of statistical concepts...”

Commission Recommendation of 23 June 2009 on Reference Metadata for the European Statistical System (2009/498/EC)

There must be flexibility in the presentation of metadata, so that statistical products have the level of detail appropriate for the target audience(s). This report recommends metadata that are mandatory (always required); conditional (understood as required under certain specified conditions); and optional (permitted but not required)⁹. These

⁹ This classification of mandatory, conditional and optional metadata is based on the ISO 11179: Metadata registries standard.

recommendations are based on metadata standards and use terms that are consistent with SDMX and ESMS.

Mandatory metadata

Data presented in tables, charts and maps in MDG reports, online databases, or other formats, should always be accompanied by the following metadata elements¹⁰:

- 1. Title describing data being presented includes the following:**
 - **Statistical unit:** *entity for which statistics are compiled (e.g. persons, households, events, enterprises)*
 - **Reference area:** *the country or geographic area to which the measured statistical phenomenon relates*
 - **Reference period:** *the period of time or point in time to which the measured observation is intended to refer*
 - **Unit of measure:** *the unit in which the data values are measured.*
- 2. Data provider:** *organization which produced the data.*
- 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.

Conditional metadata

- 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. Comparability issues can be broken into:*
 - 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 data**¹¹: *characteristics and components of the raw statistical data used for compiling statistical aggregates, i.e. type of primary source (e.g. survey, census,*

¹⁰ The descriptions of each element are based on definitions found in the Metadata Common Vocabulary (SDMX, 2009b).

¹¹ Many of the MDG indicators are rates or ratios comprised of 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

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.*

Optional metadata

There is a range of other information that will be helpful in guiding the user in their interpretation and use of MDG-related data. This metadata could be provided in an annex or other 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 should be provided.

- 7. Accuracy** – *quality 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 information on how to reach the contact points (e.g. website, mail address, phone, e-mail).*
- 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.*

In all cases, metadata should be clear and easy to understand for 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).

Specific requirements for presenting MDG-related data

The metadata above recommended are applicable to any data. Comparability is important for all data, but especially for MDG-related data. Improvement over time is at the core of the MDG goals and targets. Changes in definitions, methods and primary data sources have to be explained by metadata and it is essential to specify deviations from the definition and methodology recommended by the United Nations agencies (e.g. United Nations, 2003; ILO, 2009).

Each MDG indicator is based on different sources and is often compiled by different organizations in the national statistical system. As illustrated by the Generic Statistical Business Process Model (see page 13), several departments and persons work on the

enrolment data and population data). The optimum metadata would specify all primary source data used in deriving the estimates.

production of the data within these agencies. Ideally, at each step in the process, the competent specialist has recorded the metadata and indicated the relevance of the information. If this is not available, identifying the issues associated with estimating each MDG indicator is a good guide to determining the type of metadata that is relevant to correct interpretation. For example, the issues highlighted in the “Comments and Limitations” section of the United Nations MDG Handbook (2003) point to common barriers to comparability where more detailed metadata would be useful.

The next section provides examples of recommended metadata and current practices in presenting a range of MDG indicators.

VI. Specific metadata and examples of current practice

This section focuses on 14 MDG indicators that were carefully chosen by UNECE to reflect the diversity of metadata requirements. Information to guide metadata requirements, such as related concepts and definitions, is provided for each indicator.

Comparability issues highlighted in the United Nations MDG Handbook and other guides on the production of these indicators are used to emphasise the specific metadata that are relevant to each indicator. This information complements the mandatory, conditional and optional metadata recommendations above.

The precise metadata to be provided depends on the exact definition, methodology, data source used, and specific national issues. This leads to a large number of possibilities. The examples are not intended to be exhaustive, but serve as illustrations of the principles of the process. All basic information to understand the data is mandatory. The rule of thumb for conditional and optional metadata is that all issues that might influence the reliability or comparability and thus the interpretation of the data, should be covered. Depending on the type of publication, this can be provided alongside the data, added as a footnote or an appendix, or covered through a reference or hyperlink in the case of data published in electronic form.

Indicator 1.1 – Population below \$1 (PPP) per day, percentage

(also referred to as *Poverty headcount ratio at \$1.25 a day (PPP) (per cent of population)*)

Definition This indicator is defined as the percentage of the population living in households below the *poverty line* where the average daily consumption (or income) per person is less than \$1.25 a day measured at 2005 prices adjusted for *purchasing power parity* (PPP).

Concepts Poverty line; Purchasing Power Parity (PPP)

Agency responsible for global reporting Development Data Group, the World Bank
iresearch.worldbank.org/PovcalNet/index.htm?0,2

Metadata considerations

Measuring the level of poverty is complex. At the international level the standard measure of \$1.25 (PPP) per day is used, measured in 2005 prices. In the past, \$1.00 per day and \$1.08 per day, in 1993 prices, have been used. These changes show the importance of including such metadata alongside the data, even if standard definitions are used. In an MDG report, further details, such as the basket used for the PPP conversion, can be covered by a reference.

At national level, poverty lines, are set using various definitions and methods. Using terms such as “relative poverty”, “absolute poverty”, “severe poverty” or “extreme poverty” should be avoided unless the exact definitions are provided. Metadata are

needed to explain how national poverty lines were determined, ensuring users can interpret the information correctly.

For example, in the case of a poverty line based on a certain food basket, the minimum food-energy intake (in kilocalories) is relevant information. If different values are used for urban and rural or for adults and children, this information as well as details such as the items in the basket, are needed for cross-country comparison. National MDG reports can provide this in an appendix or through a reference to a methodological report if it is the same for all data presented. If it is significantly different for individual data points, however, such conditional metadata should be made available more directly with the data.

Other examples of conditional metadata may be:

- how household level data is weighted for conversion to adult equivalents
- whether poverty is measured through income or through consumption.

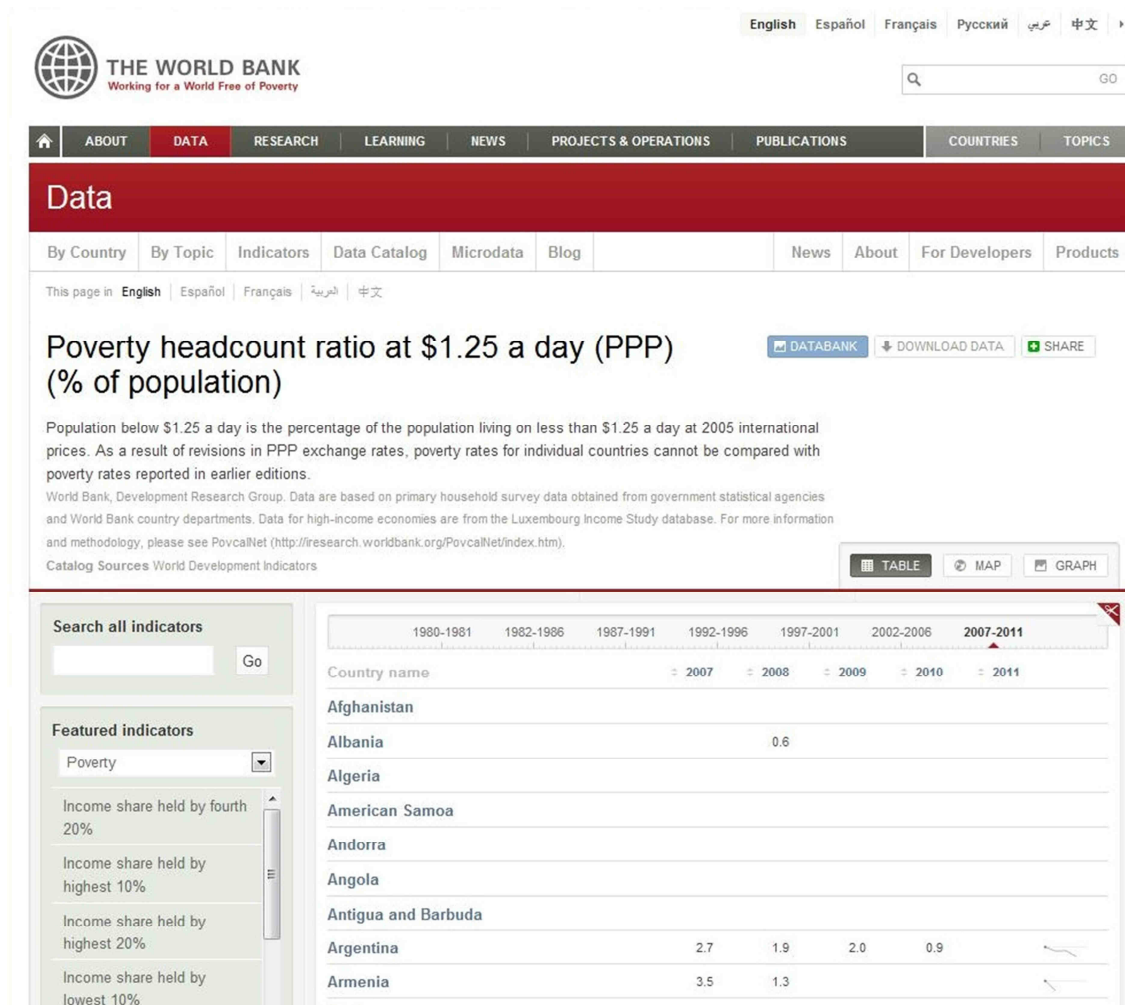
Examples of current practice

The World Bank is the international agency responsible for maintaining comparable data on poverty levels around the world. Their online database provides a number of poverty-related indicators, including MDG indicator 1.1, referred to as *Poverty headcount ratio at \$1.25 a day (PPP) (per cent of population)*.

As can be seen from Figure 21, **mandatory metadata** are clearly visible in the title of the table, with the statistical unit (population), reference area (country names), reference period (range of years as shown in column headings) and unit of measure (per cent of population), all clearly stated.

A definition of the indicator and a summary of where the source data comes from are shown under the title. A link to more detailed information (**optional metadata**) is provided for those users who want to know more about the methodology (iresearch.worldbank.org/PovcalNet/index.htm).

Figure 21. Presenting poverty data on the World Bank website (data.worldbank.org)



In the official United Nations MDG Database, examples of **conditional metadata** can be found in footnotes (Figure 22). These notes point out differences in methodology, such as where estimates are based on consumption, as is the case for most data values in the table below, or income, as was the case for Turkmenistan in 1993.

The 2009 MDG report of Kyrgyzstan clearly explains the national approach to measuring poverty in a separate box (Figure 23), drawing the attention of interested readers to more detailed metadata without complicating the analytical text in the report.

Figure 22. Presentation of MDG indicator 1.1 in the official United Nations MDG Database (mdgs.un.org)

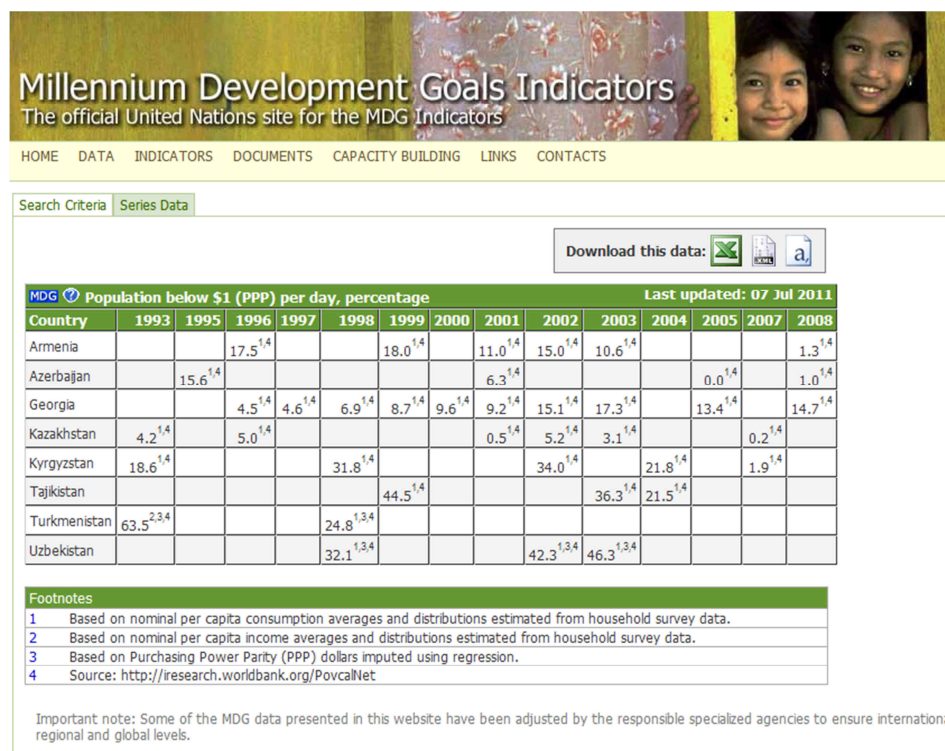
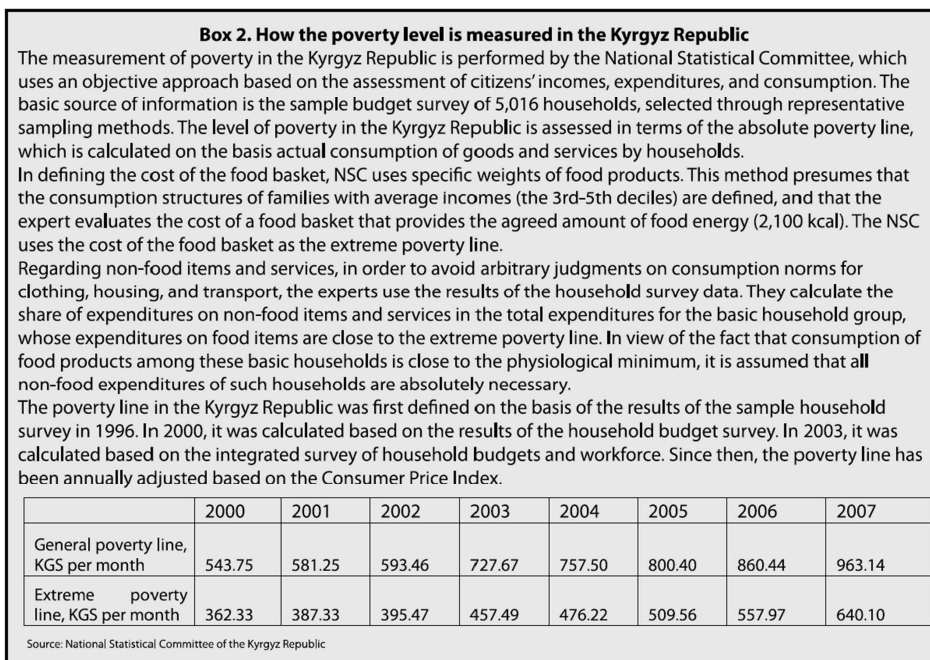


Figure 23. Kyrgyzstan explains their approach to poverty measurement



Source: The second periodic progress report on the millennium development goals in the Kyrgyz Republic, 2009.

Indicator 1.5 – Employment-to-population ratio

<i>Definition</i>	The employment-to-population ratio is the proportion of a country's working-age population that is employed. The employment-to-population ratio is expressed in units of percentage.
<i>Concepts</i>	Working-age population; employed
<i>Agency responsible for global reporting</i>	International Labour Organization (ILO) www.ilo.org/trends
<i>Relevant classifications</i>	Not applicable

Metadata considerations

The national definition for employment, in particular the number of hours worked to be considered employed, and differing upper and lower age limits for working-age population, should be explained. Information on the estimated coverage of the informal sector is relevant when it accounts for a significant proportion of total employment and income generation. In a national MDG report, details on how it is assured that employed persons who only work for a few hours per week, are in unpaid employment or that work near or in their home are included could be listed in a box or appendix. Even if standard labour force surveys were used such information is beneficial as surveys methods can change over time. Metadata can also educate the users of the data. Other cases where conditional metadata may be needed to highlight comparability issues are:

- the inclusion or exclusion of members of the armed forces, mental, penal or other types of institutions
- the age limits of the working age population, especially if it does not refer to the persons aged 15 years and older
- the impact of seasonal employment not captured by data collection method.

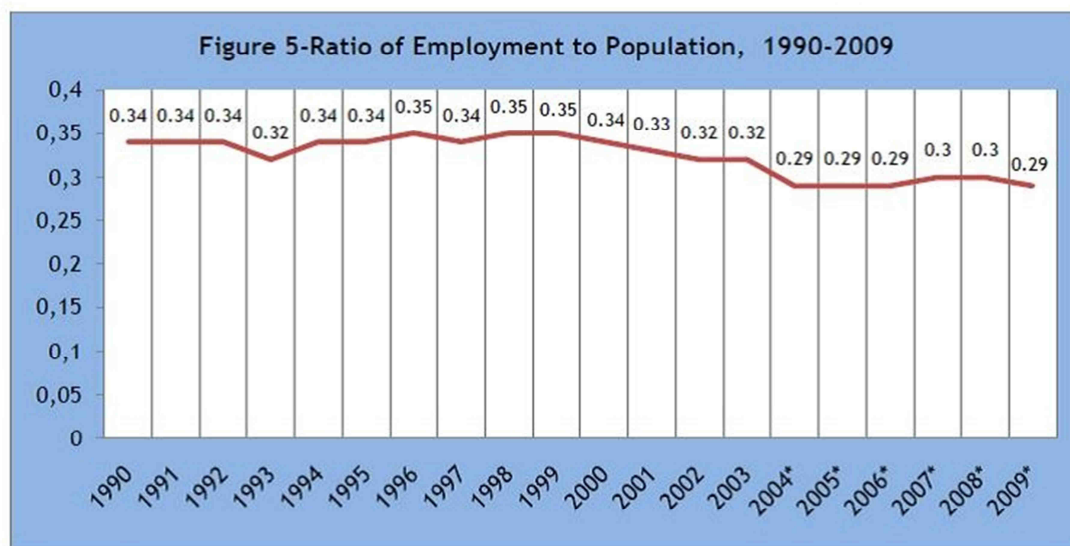
More details can be found in the *Guide to the Millennium Development Goals Employment Indicators* (ILO, 2013) and *Key Indicators of the Labour Market* (ILO, 2011).

Examples of current practice

Presentation of the employment-to-population ratio in a national MDG report of 2010 (Figure 24) lacks **mandatory metadata** needed to interpret the figures, such as the age range of the population used in the calculating the estimates. Comparing these values to those in the United Nations MDG Database for the country, there are significant differences (the United Nations figures are between 8-18 per cent higher than national estimates). Without a definition of the working-age population, it is unclear why there

might be such a discrepancy. The example correctly mentions information regarding a break in series.

Figure 24. Inadequate presentation: Employment-to-population ratio in a national MDG report



Source: [redacted]

* These figures were revised according to the population data in the 2008 [redacted] and the results of the 2008 [redacted] Population and Health Survey conducted by the [redacted]. Therefore the data pertaining to 2004, 2005, 2006, 2007, 2008 and 2009 are not comparable with previous years.

The OECD *Employment Outlook* (2006) publication provides a good example of metadata that should be included with estimates of employment to population ratios:

- the age of the population (15-64) is clearly indicated in the table heading
- the method of calculation is provided in a footnote, guiding the reader as to what these values represent
- anomalies and breaks in series are included in the footnotes
- source is indicated.

OECD provide further **conditional and optional metadata** on sources, definitions, symbols used and breaks in series in another section of this publication.

Figure 25. Presentation of employment/population ratios in an OECD publication

Table B. Employment/population ratios, activity and unemployment rates^a (cont.)
Age limits of working age population

Women aged 15-64 years (percentages)

	Employment/population ratio						Labour force participation rate						Unemployment rate					
	1994	2001	2002	2003	2004	2005	1994	2001	2002	2003	2004	2005	1994	2001	2002	2003	2004	2005
Australia	56.9	61.7	62.1	62.9	63.1	64.7	62.8	66.0	66.3	67.1	66.9	68.4	9.5	6.5	6.3	6.2	5.7	5.3
Austria	58.8	59.8	61.0	61.5	60.7	62.0	61.3	62.2	63.5	64.2	64.2	65.6	4.0	3.8	3.9	4.2	5.4	5.5
Belgium	44.8	50.7	51.1	51.4	53.0	54.1	51.2	54.5	55.4	55.8	57.7	59.5	12.5	6.9	7.8	8.0	8.3	9.0
Canada	61.1	65.9	67.0	67.9	68.4	68.3	67.8	70.8	72.1	73.2	73.4	73.1	9.8	6.9	7.2	7.2	6.9	6.5
Czech Republic	61.0	57.0	57.1	56.3	56.0	56.3	64.4	63.2	62.8	62.5	62.2	62.4	5.2	9.9	9.1	9.9	10.0	9.8
Denmark	67.1	71.4	72.6	70.5	72.0	70.8	73.8	75.0	75.9	74.8	76.1	75.1	9.0	4.8	4.4	5.8	5.5	5.6
Finland	58.7	65.4	66.1	65.7	65.5	66.5	69.1	72.5	72.8	72.1	71.9	72.9	14.9	9.7	9.1	8.9	9.0	8.7
France	50.8	55.2	55.8	56.4	56.7	56.9	59.3	61.8	62.1	63.4	63.8	63.8	14.4	10.8	10.1	11.0	11.1	10.9
Germany	54.7	58.7	58.8	58.7	59.2	59.6	60.9	63.8	64.2	64.5	65.8	66.9	10.1	8.0	8.4	8.9	10.1	11.0
Greece	37.1	41.2	43.1	44.5	45.5	46.2	43.2	48.8	51.0	52.1	54.1	54.6	14.0	15.6	15.4	14.5	16.0	15.3
Hungary	47.8	49.8	49.8	50.9	50.7	51.0	52.7	52.4	52.7	53.9	54.0	55.1	9.3	5.0	5.4	5.6	6.1	7.5
Iceland ^b	74.6	81.1	79.8	81.2	79.4	81.2	79.1	83.1	82.2	83.9	81.8	83.5	5.7	2.5	2.9	3.1	3.0	2.7
Ireland	38.9	54.0	55.2	55.4	55.8	58.0	45.8	56.0	57.3	57.6	58.0	60.3	15.2	3.5	3.7	3.9	3.7	3.8
Italy	35.4	41.1	42.0	42.7	45.2	45.3	41.9	47.3	47.9	48.3	50.6	50.4	15.5	13.1	12.3	11.7	10.6	10.1
Japan	56.5	57.0	56.5	56.8	57.4	58.1	58.3	60.1	59.7	60.0	60.1	60.8	3.1	5.1	5.4	5.1	4.7	4.4
Korea	49.8	50.9	52.0	51.1	52.2	52.5	50.8	52.8	53.5	52.9	54.1	54.5	2.0	3.5	2.9	3.5	3.5	3.6
Luxembourg	44.9	50.8	51.5	52.0	51.9	53.7	47.0	52.0	53.5	54.5	55.9	57.0	4.3	2.2	3.6	4.6	7.1	5.9
Mexico	36.2	39.4	39.9	39.4	41.3	41.5	38.1	40.4	41.0	40.5	42.8	43.1	4.9	2.4	2.5	2.7	3.5	3.7
Netherlands	52.6	63.9	64.5	64.9	65.0	..	57.3	66.1	66.9	67.9	68.5	..	8.1	3.3	3.5	4.4	5.1	..
New Zealand	59.9	64.8	65.3	65.7	66.5	68.0	64.9	68.4	69.1	69.2	69.6	70.8	7.7	5.3	5.4	5.1	4.5	4.1
Norway ^b	67.5	73.8	73.9	72.7	72.7	72.0	70.9	76.4	76.6	75.8	75.7	75.4	4.8	3.4	3.7	4.0	3.9	4.4
Poland	51.9	47.8	46.4	46.2	46.4	47.0	62.1	59.9	58.9	58.4	58.2	58.3	16.4	20.2	21.2	20.8	20.2	19.4
Portugal	55.0	61.0	60.8	60.6	61.7	61.7	60.0	64.5	65.0	65.6	67.0	67.9	8.3	5.4	6.5	7.7	8.0	9.2
Slovak Republic	52.6	51.8	51.4	52.2	50.9	50.9	61.3	63.8	63.2	63.5	62.9	61.5	14.1	18.8	18.7	17.8	19.1	17.2
Spain ^b	31.5	43.8	44.9	46.8	49.0	51.9	46.3	51.6	53.7	55.7	57.7	59.1	31.8	15.3	16.4	16.0	15.1	12.2
Sweden ^b	70.7	73.5	73.4	72.8	71.8	..	77.0	77.1	77.1	76.8	76.6	..	8.2	4.7	4.7	5.2	6.2	..
Switzerland	64.9	70.7	71.5	70.7	70.3	70.4	68.2	73.3	73.9	74.1	73.9	74.3	4.4	3.5	3.2	4.6	4.8	5.2
Turkey	30.4	26.3	26.6	25.2	24.3	23.7	33.1	28.5	29.5	28.1	27.0	26.5	8.3	7.8	9.8	10.5	10.0	10.6
United Kingdom ^b	62.1	66.0	66.3	66.4	66.6	66.8	67.1	68.9	69.3	69.2	69.6	69.7	7.4	4.2	4.4	4.1	4.3	4.1
United States ^b	65.2	67.1	66.1	65.7	65.4	65.6	69.4	70.4	70.1	69.7	69.2	69.2	6.1	4.7	5.7	5.7	5.5	5.2
EU-15 ^c	49.4	55.1	55.6	56.1	57.1	57.8	56.5	60.3	61.0	61.6	62.8	63.5	12.7	8.7	8.8	9.0	9.1	9.0
EU-19 ^c	49.9	54.3	54.6	55.0	55.9	56.5	57.2	60.2	60.7	61.2	62.2	62.8	12.8	9.8	9.9	10.1	10.1	10.0
OECD Europe ^c	48.0	51.2	51.5	51.6	52.1	52.5	54.8	56.5	57.0	57.2	57.9	58.3	12.3	9.5	9.7	9.9	10.0	9.9
Total OECD ^c	52.9	55.4	55.3	55.3	55.8	56.1	57.8	59.4	59.6	59.7	60.1	60.4	8.4	6.8	7.2	7.3	7.2	7.0

a) Ratios refer to persons aged 15 to 64 years who are in employment or in the labour force divided by the working age population, or in unemployment divided by the labour force.
 b) Refers to persons aged 16 to 64.
 c) Averages for 2005 include estimates for the Netherlands and Sweden, based on annual growth rates of series taken from the European Union Labour Force Survey.
 Source: OECD database on Labour Force Statistics (see URLs at the beginning of the Annex). For Belgium, Denmark, Greece and Luxembourg data are from the European Union Labour Force Survey.
 .. Data not available
 . Decimal point
 | Break in series
 Statlink: <http://dx.doi.org/10.1787/282807675356>

Footnotes give the data source, anomalies, breaks in series and method of calculation

Source: Employment Outlook 2006 (OECD, 2006) (available from www.oecd.org/dataoecd/53/15/36900060.pdf).

Indicator 1.7 – Proportion of own-account and contributing family workers in total employment (vulnerable employment rate)

<i>Definition</i>	<p>The proportion of own-account workers and contributing family workers in total employment is defined as the proportion of workers in self-employment who do not have employees, and unpaid family workers in total employment.</p> <p>This indicator is expressed in units of percentage.</p>
<i>Concepts</i>	<p>Own-account workers; contributing family workers (also known as unpaid family workers); self-employment, employees; employment; vulnerable employment</p>
<i>Agency responsible for global reporting</i>	<p>International Labour Organization (ILO) www.ilo.org/trends</p>
<i>Relevant classifications</i>	<p>International Classification by Status in Employment (ICSE), 1993</p>

Metadata considerations

As with MDG indicator 1.5 above, clarifying the definition for employment is also relevant for this indicator. Other areas where **conditional metadata** may be needed to highlight comparability issues are:

- Non-standard groupings of employment status, e.g. members of producers' cooperatives included with wage and salaried workers rather than self-employed by some countries
- Exclusion of members of the armed forces, mental, penal or other types of institutions, may impact on relative shares of employment, particularly in countries with large armed forces
- Inconsistent coverage of rural and urban areas.

More details can be found in the *Guide to the Millennium Development Goals Employment Indicators* (ILO, 2009) and *Key Indicators of the Labour Market* (ILO, 2011).

Examples of current practice

The ILO LABOURSTA online database provides a good example of how to present metadata with this indicator. Text in the table heading (Figure 26) includes links to a detailed definition of the indicator and employment status categories, as well as explanation of abbreviations and symbols. The topic title clarifies the unit of measurement (thousands, rather than percentage of total population) and the source of the data (Labour force survey). A footnote indicates a break in series that occurred in 2004. The definition of the working-age population is also made clear in a footnote (aged 15 years and over).

Figure 26. Employment data presented in the ILO LABORSTA online database

>> **Topic:** Employment - 2D Total employment, by status in employment (Thousands)

>> **List of countries selected:** MACEDONIA, THE FORMER YUGOSLAV REP. OF |

>> **Tips:**

- Rolling the mouse over each classification category will display the corresponding title
- Abbreviations and symbols

>> Download

MACEDONIA, THE FORMER YUGOSLAV REP. OF [List of countries selected]

Source: (BA) Labour force survey

Total coverage
ICSE-1993 ¹

Total men and women	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Total	545.2	549.8	599.3	561.3	545.1	523.0 ²	545.3	570.4	590.2	609.0
1 Employees	406.0	408.7	419.8	396.7	396.5	394.3 ²	391.7	403.6	426.7	437.5
2 Employers	43.4	23.7	38.2	38.4	41.4	30.8 ²	31.3	33.9	32.7	30.1
3 Own-account workers	39.7	57.7	55.0	57.6	44.8	53.3 ²	65.5	70.8	71.2	78.8
5 Contributing family workers	56.1	59.8	86.2	68.7	62.4	44.6 ²	56.8	62.2	59.7	62.6
Men	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Total	338.0	339.6	357.3	342.8	327.3	320.6 ²	332.2	352.0	358.8	373.5
1 Employees	245.0	249.4	250.9	240.3	235.0	231.5 ²	230.0	241.0	248.6	261.1
2 Employers	37.2	19.8	29.3	29.6	32.8	23.8 ²	25.1	26.9	25.0	22.9
3 Own-account workers	33.3	45.2	43.5	46.1	36.6	44.9 ²	55.7	59.1	60.0	65.8
5 Contributing family workers	22.5	25.1	33.5	26.7	23.0	20.4 ²	21.3	25.0	25.2	23.6
Women	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Total	207.2	210.3	242.0	218.6	217.8	202.4 ²	213.1	218.4	231.4	235.5
1 Employees	161.0	159.3	168.9	156.3	161.5	162.9 ²	161.7	162.6	178.0	176.4
2 Employers	6.2	3.9	8.9	8.8	8.7	7.0 ²	6.1	7.0	7.7	7.2
3 Own-account workers	6.4	12.5	11.5	11.5	8.2	8.3 ²	9.8	11.7	11.2	13.0
5 Contributing family workers	33.6	34.7	52.7	42.0	39.4	24.2 ²	35.5	37.2	34.5	39.0

Notes: ¹Persons aged 15 years and over. ²Prior to 2004: April.

Source: LABORSTA (laborsta.ilo.org¹²).

The table on economic status of men and women in a national MDG report (Figure 27) shows the proportion of men and women for each category of employment. The presentation would be improved with additional metadata:

- clarifying that the unit of measure is a percentage of each employment category
- providing definitions of employed, employer, self-employed and unpaid family worker
- explaining why the standard category of “Members of producers’ cooperatives” is not included in the table would clarify whether this category has been amalgamated with another, not applicable or left out for some reason
- providing information on the source data, such as whether the data are based on population census, labour force survey, or another source.

The information would be more meaningful if data on the overall proportion of unpaid family workers in total employment were also provided in the report.

¹² LABORSTA will be migrated to ILOSTAT database at www.ilo.org/ilostat.

Figure 27. Inadequate metadata: Economic status in a national MDG report

Table 3.3.: Economic Status of Men and Women – Proportion

	2003		2006		2007	
	women	men	women	men	women	men
Employed	41	59	40	60	42	58
Employer	21	79	21	79	24	76
Self-employed	18	82	17	83	16	84
Unpaid family worker	63	37	60	40	58	42
Total	40	60	38	62	39	61

Source: State Statistical Office

Indicator 1.8 – Prevalence of underweight children under five years of age

<i>Definition</i>	This indicator is defined as the percentage of children aged 0-59 months whose weight is less than two standard deviations below the median weight for age of the international reference population.
<i>Concepts</i>	International reference population; moderately and severely underweight
<i>Agency responsible for global reporting</i>	United Nations Children’s Fund (UNICEF) www.childinfo.org
<i>Relevant classifications</i>	World Health Organization Child Growth Standards, 2006 www.who.int/childgrowth/en/

Metadata considerations

The comparability of this indicator can be impacted by the quality of data collection – coverage, accuracy of age, weight and height measurements – as well as change to the reference population with the release of the WHO Child Growth Standards in 2006. Metadata are needed to describe these and other conditions that may impact on the interpretability of the data.

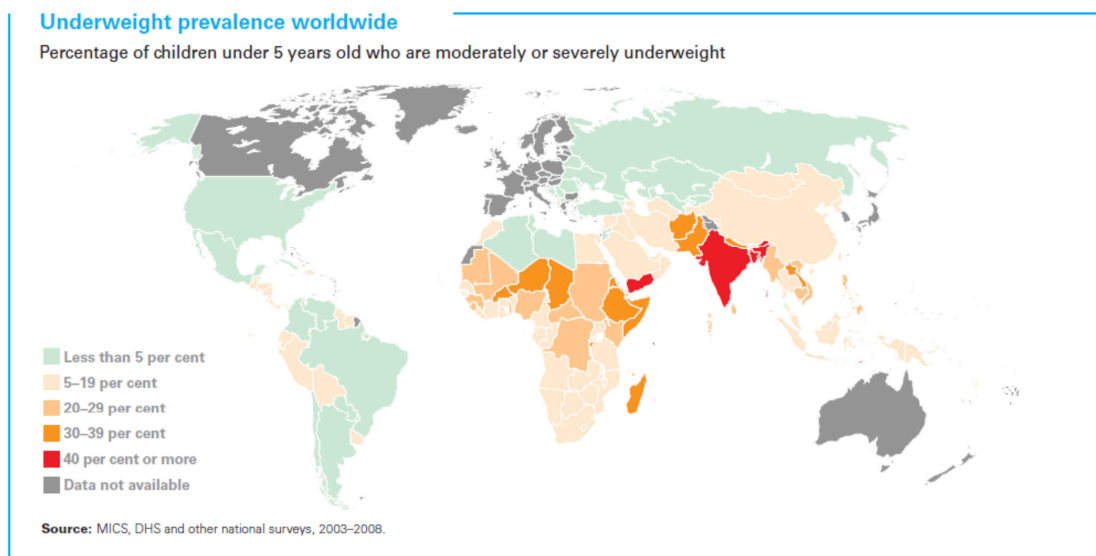
Mandatory metadata would be the reference to the primary data source. If non-standard surveys are used to obtain the data, it is important to report on the methodology and the sampling procedure. Depending on how different results would be compared to standard methodology, it should be either provided with the data or through a reference.

Examples of current practice

In their thematic report on child nutrition, UNICEF (2009) presents data on the prevalence of underweight children (aged under five years old) using a map. As Figure 28 below shows, the **mandatory metadata** are provided in the title, with source data specified beneath the map (i.e. Multiple Indicator Cluster Survey (MICS), Demographic Health Surveys (DHS) and other national surveys, 2003-2008). The legend confirms the percentage range that each colour corresponds to.

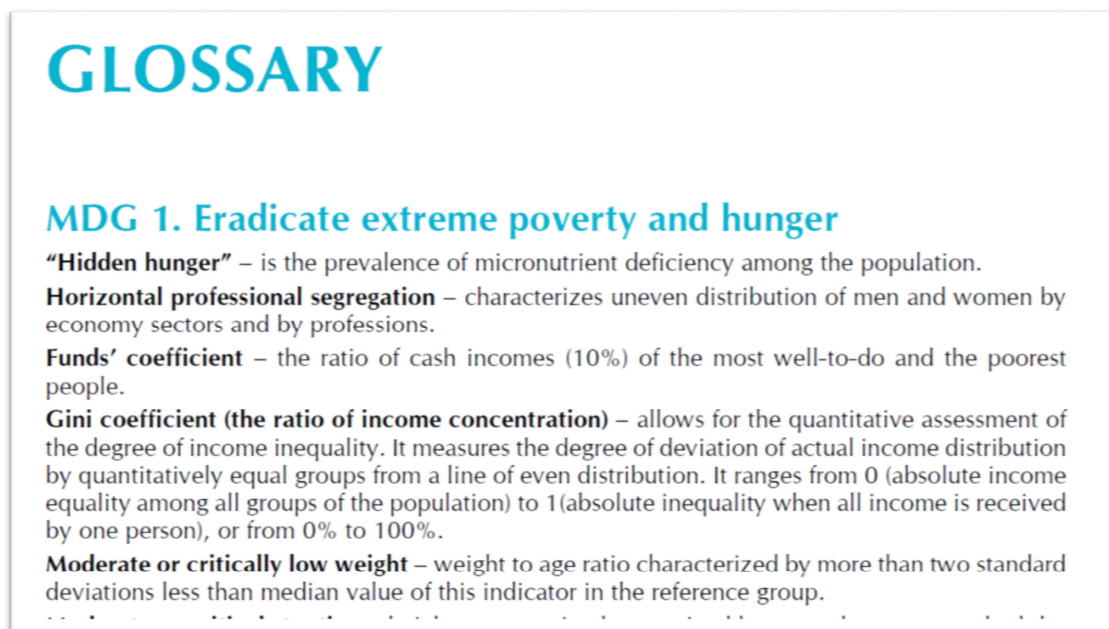
It is useful to include **optional metadata** in a glossary, such as that found in Kazakhstan’s 2007 MDG Report, which defines the statistical concepts being measured, such as “moderate or critically low weight” (Figure 29).

Figure 28. Using a map to present underweight prevalence data and metadata



Source: *Tracking Progress on Child and Maternal Nutrition: A survival and development priority* (UNICEF, 2009).

Figure 29. A glossary in the MDG Report of Kazakhstan (2007) provides an explanation of terms used throughout the report



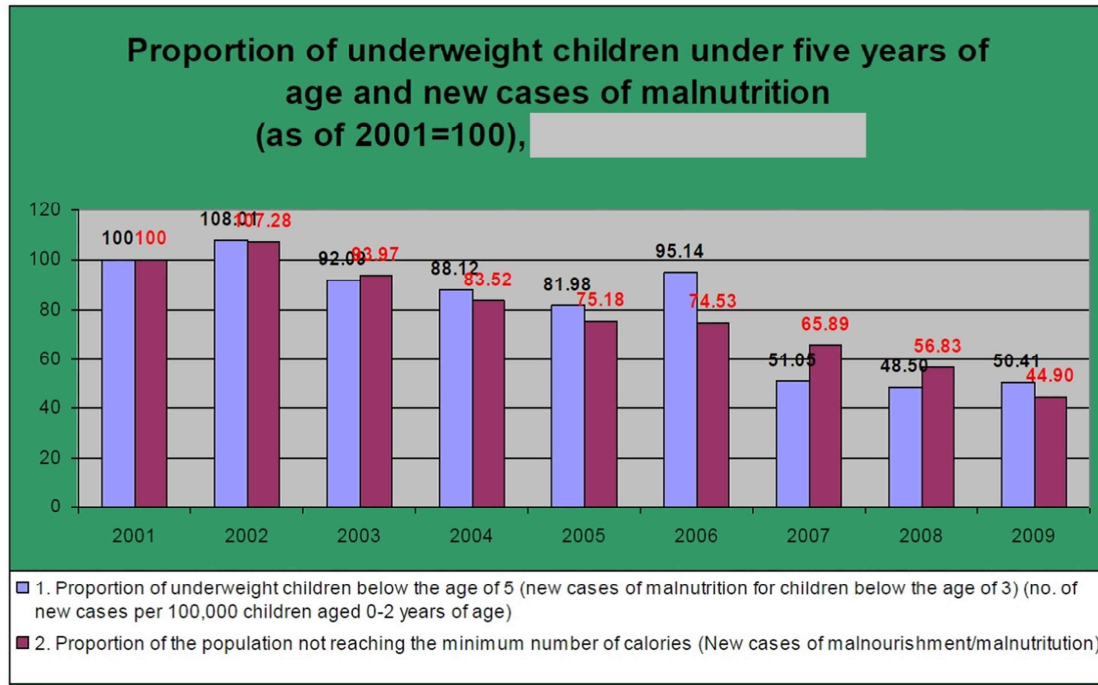
Source: *Millennium development goals in Kazakhstan, 2007*.

The presentation of indicator 1.8 in a national MDG report release in 2010 is an example of where **mandatory metadata** are lacking. The main issue here is that the data are presented as an index, showing the relative change based on the proportion of underweight children aged five years or less in 2001. To readers less familiar with

statistics, it would be difficult to know this is the case and the presentation could be confused as indicating the number of cases each year.

Metadata should be included to explain how the index was produced. Furthermore, the descriptions in the chart's legend could be clearer and a definition of underweight could be given in the chart, analytical text or the report's appendices.

Figure 30. Inadequate metadata: Proportion of underweight children in the national MDG report published in 2010



Source: [Redacted]

Indicator 2.1 – Net enrolment rate in primary education

<i>Definition</i>	<p>The net enrolment rate (NER) in primary education is the number of children of official primary school age who are enrolled in primary education expressed as a percentage of the total population of children of official primary school age.</p> <p>United Nations Education, Science and Culture Organization (UNESCO) recommends using an adjusted NER in primary education, calculated as the number of children of official primary school age who are enrolled either in primary or secondary education expressed as a percentage of the total population of children of official primary school age. This adjusted NER is used for international MDG monitoring.</p>
<i>Concepts</i>	Children of official primary school age, primary education
<i>Agency responsible for global reporting</i>	United Nations Education, Science and Culture Organization (UNESCO) Institute for Statistics www.uis.unesco.org
<i>Relevant classifications</i>	International Standard Classification of Education (ISCED). Currently the revision of 1997 is in use. A new revision (ISCED 2011) is expected to be in use for international data reporting from 2014 (collection) or 2015 (dissemination)

Metadata considerations

Education systems differ between and sometimes within countries. Net enrolment in primary education is calculated based on the official age for primary school in any given country or area. Provided that the correct age range is used for both the numerator and the denominator¹³, differences in official age ranges between countries and areas does not affect comparability for this indicator. However, metadata should indicate the official age range used, particularly for comparability over time should the official primary age range change in the future.

In order to ensure cross-national comparability, the levels of education in national systems are mapped to ISCED. Any shift from the definition or the intended coverage of the indicator should be clearly indicated in accompanying metadata.

If administrative data are used to estimate the children in education, population estimates of the children of official primary school age are used as a denominator. Errors in the latter can significantly influence the indicator. It is therefore important to list the source of the population estimates as well. Any issues with the accuracy of population estimates should be covered by metadata.

¹³ The numerator is the number of official primary school age children currently attending primary school and the denominator is all children of official primary school age.

Other examples of situations where conditional metadata would be required are:

- when children of primary ages that are attending pre-primary education are included, it should be indicated
- information on over- or underreporting in surveys and administrative data.

Examples of current practice

The presentation of net enrolment rate in the UNESCO Institute for Statistics annual flagship publication – *Global Education Digest 2011* – illustrates how to present **conditional metadata** in footnotes.

Figure 31. Net enrolment rate presented in the Global Education Digest (UNESCO Institute for Statistics, 2011)

REGION	Net enrolment rate (adjusted)							
	1999				2009			
	MF (12)	M (13)	F (14)	GPI (15)	MF (16)	M (17)	F (18)	GPI (19)
Country or territory								
Georgia	100
Kazakhstan	95 ^{**,+1}	94 ^{**,+1}	96 ^{**,+1}	1.02 ^{**,+1}	100 ⁺¹
Kyrgyzstan	88 [*]	89 [*]	87 [*]	0.99 [*]	91	91	91	1.00
Mongolia	96	94	97	1.03	100
Tajikistan	97 ⁺¹	98 ⁻¹	99 ⁻¹	96 ⁻¹	0.96 ⁻¹
Turkmenistan
Uzbekistan	90	91	89	0.98

The symbols used in the table above refer to general notes available in the accompanying Reader's Guide:

...	No data available
*	National estimation
**	For country data: UIS estimation For regional averages: Partial imputation due to incomplete country coverage (between 33% to 60% of population)
-	Magnitude nil or negligible
.	Not applicable
x(y)	Data are included in column (y) of the table
+n	Data refer to the school or financial year n years after the reference year
-n	Data refer to the school or financial year n years prior to the reference year

This Reader's Guide also provides explanation of reference year, data sources, statistical concepts (e.g. population), and technical notes. The latter includes notes on geographical coverage. For example:

“Serbia: Education data do not cover Kosovo, whereas UN Population Division’s data do. Therefore, population data used for the calculation of indicators were provided by Eurostat and include Kosovo.”

The publication includes a glossary and definitions of indicators amongst its appendices. These provide more information to help the reader understand the indicator:

“Enrolment. The number of pupils or students officially enrolled in a given grade or level of education, regardless of age.”

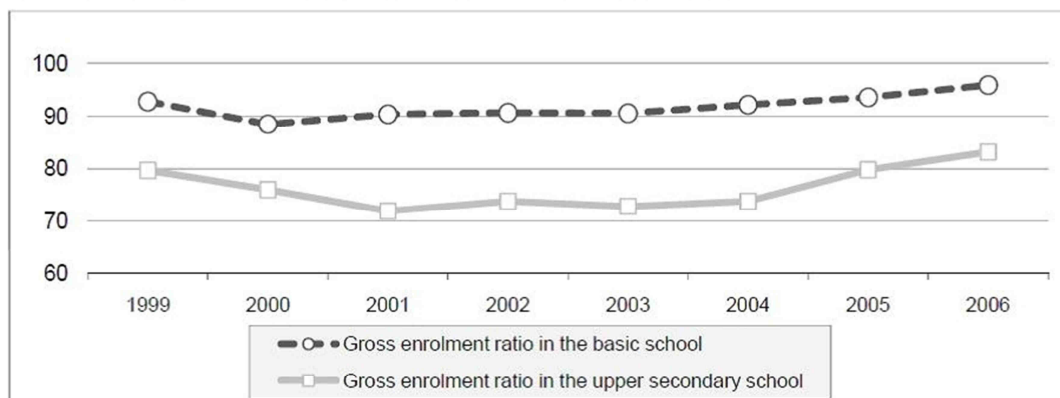
“Net enrolment ratio (NER). Total number of pupils or students in the theoretical age group for a given level of education enrolled in that level, expressed as a percentage of the total population in that age group.”

Gross enrolment ratios are also an important development indicator, providing an indication of the number of students enrolled in particular levels of schooling, as a percentage of the official age for that level. As the numerator (enrolment) is not a real sub-set of the denominator (population), gross enrolment ratios can exceed 100 per cent (i.e. includes enrolments outside the official age range for the level of education). When this occurs, it indicates some degree of over-age or under-age enrolment at that level.

The presentation of gross enrolment ratios in basic and upper secondary schools in a national MDG report in Figure 32 appears to include all the mandatory metadata needed. It would be enhanced by the specification of the official age range for these levels and comparison to net enrolment rates. These would allow the reader to better understand the extent to which under or over-age students make up those enrolled in secondary school.

Figure 32. Inadequate metadata: Gross enrolment ratio presented in a national MDG report

Figure 2.2. The Gross Enrolment Ratios in the Basic and Upper Secondary Schools in 1999-2006, %



Source: [redacted] (data from administrative register).

Indicator 3.1 – Ratios of girls to boys in primary, secondary and tertiary education (Gender Parity Index)

<i>Definition</i>	The ratio of girls to boys in primary, secondary and tertiary education is measured by the ratio of female to male values of the Gross Enrolment Ratio (GER) in the respective level of education. GER is defined as total enrolment in a specific level of education, regardless of age, expressed as a percentage of the official school-age population corresponding to the same level of education in a given school year.
<i>Concepts</i>	Primary education; secondary education; tertiary education; gender parity index (GPI); gross enrolment ratio (GER); official school age population
<i>Agency responsible for global reporting</i>	United Nations Education, Science and Culture Organization (UNESCO) Institute for Statistics www.uis.unesco.org
<i>Relevant classifications</i>	International Standard Classification of Education (ISCED). Currently the revision of 1997 is in use. A new revision (ISCED 2011) is expected to be in use for international data reporting from 2014 (collection) or 2015 (dissemination)

Metadata considerations

It is important only to use primary, basic, secondary and tertiary education if the concepts are in line with the most recent ISCED classification and to add further information if they deviate. Information on the school system and corresponding age groups is optional metadata in most publications, however, if changes in the school system occur that influence the comparability, it would be important metadata to be included along with time series.

As is the case with MDG Indicator 2.1 above, education statistics, such as those on ratios of boys to girls at different levels of schooling, face comparability issues due to varying national education systems and official age ranges. Inconsistencies between population and enrolment data derived from different sources may also need explaining with metadata, as will the exclusion of private education institutions or technical and vocational education.

Other examples of metadata that might need to be included are:

- deviations of school classification from the latest version of ISCED
- the age group used for the corresponding population for tertiary education, especially if it deviates from the 5 year age group immediately following the official age of completion of secondary education that UNESCO uses.

Examples of current practice

The presentation of this indicator in a national MDG report shown in Figure 33 is a good example of providing important details on the source data that affect comparability.

The table could be enhanced with clearer labelling of the two types of data being presented in each column. Rather than expecting readers to look at the footnote to know that the first figure is the share of female pupils and the second figure, in parentheses, is the ratio of female to male pupils.

Figure 33. Imperfect presentation: Gender parity index in a national MDG report

Table 2.7.: The Share of Female Pupils/Students at All Levels of Education (Proportion of Girls and the Ratio Between Girls and Boys)²⁶

	1997/ 1998	1998/ 1999	1999/ 2000	2000/ 2001	2001/ 2002	2002/ 2003	2003/ 2004	2004/ 2005	2005/ 2006	2006/ 2007	2007/ 2008
Primary	48.2 (93.0)	48.2 (93.3)	48.2 (93.0)	48.3 (93.5)	48.4 (94.0)	48.4 (93.9)	48.4 (93.9)	48.5 (94.4)	48.5 (94.4)	48.3 (93.5)	48.5
Secondary	48.2 (93.0)	48.2 (92.9)	48.1 (92.8)	48.1 (92.6)	48.2 (93.1)	48.1 (92.5)	47.7 (91.1)	47.4 (90.1)	47.0 (88.8)	47.4 (90.3)	46.9
High	54.6 (120.3)	55.2 (123.2)	55.1 (122.8)	55.8 (126.5)	55.2 (123.4)	56.2 (128.3)	57.0 (132.7)	56.7 (131.2)	56.6 (130.6)	54.7 (120.7)	53.5

Source: State Statistical Office

²⁶ The first figure shows the percentage of female pupils/students in the total number of pupils/students at all three levels of education. The figures in parentheses represent the ratio of female to male pupils/students. The calculations are based on the data obtained at the beginning of the academic year. The data concerning higher education refer to students who are citizens enrolled in the state universities and some of the private ones.

Understanding indices is complex for data users that are less familiar with statistics. The gender parity index can be difficult to interpret without some explanation. The UNESCO Institute for Statistics (2011) provides a good description of the GPI and its limitations in their *Global Education Digest 2011*, as shown in Figure 34.

Figure 34. Explaining the gender parity index (UNESCO Institute for Statistics, 2011)

BOX 2. Understanding the gender parity index (GPI)

The GPI is a measure used to assess gender differences in education indicators. It is defined as the value of a given indicator for girls divided by that for boys. A GPI value of 1 signifies that there is no difference in the indicators for girls and boys—they are perfectly equal. UNESCO (2003) has defined a GPI value of between 0.97 and 1.03 (after rounding) as the achievement of gender parity. This allows for some measurement error but does not imply a judgement about the acceptability of any particular level of disparity.

In cases where disparities are apparent, the interpretation of the GPI depends on the type of indicator. A GPI of less than 1 indicates that the value of the indicator is higher for boys than for girls, and the opposite is true when the GPI is greater than 1.

For indicators where higher values are desirable (e.g. school participation rates), a GPI of less than 1 means that girls are at a disadvantage, while a GPI greater than 1 means that boys are at a disadvantage. For indicators where lower values are desirable (e.g. dropout rates), a GPI of less than 1 means that boys are at a disadvantage, and a GPI of greater than 1 means that girls are at a disadvantage.

One of the difficulties in presenting the GPI is that the scale of disadvantage for girls or boys is not represented symmetrically around 1. For example, a GPI of 0.5 indicates that the female value of the indicator being reviewed is one-half the male value, while a GPI of 1.5 (also 0.5 units away from parity) indicates that the male value of the indicator is two-thirds of the female value (not one-half). Consequently, when boys are under-represented in a given indicator, it appears more drastic than when girls are disadvantaged.

Thus, for the analysis and figures presented in this section, the GPI is adjusted to present disadvantages symmetrically for both genders. The adjusted GPI is derived from the standard GPI as presented in the Statistical Tables, but values greater than 1 are treated differently in this section and disparities are presented on a comparable scale.

The adjusted GPI uses the following methodology: when the ratio of female to male values of a given indicator is less than 1, the adjusted GPI is identical to the unadjusted GPI. By contrast, when the ratio is greater than 1, the adjusted GPI is calculated as the ratio of male to female values and the ratio is subtracted from 2. For instance, if the net enrolment rate (NER) for males is 33% and 66% for females, the ratio of male to female NER is 0.5. Then, subtracting 0.5 from 2 gives an adjusted GPI of 1.5, while the unadjusted GPI would show a result of 2.

Note: Only the regular (not adjusted) GPI is presented in the Statistical Tables of this report.

Source: *Global Education Digest (UNESCO Institute for Statistics, 2011)*.

Indicator 3.3 – Seats held by women in national parliament

<i>Definition</i>	The proportion of seats held by women in national parliament is the number of seats held by women members in single or lower chambers of national parliaments, expressed as a percentage of all occupied seats.
<i>Concepts</i>	Seats; unicameral (single chamber parliament) or bicameral (lower and upper chamber parliament)
<i>Agency responsible for global reporting</i>	Inter-Parliamentary Union (IPU) www.ipu.org/wmn-e/world.htm
<i>Relevant classifications</i>	Not applicable

Metadata considerations

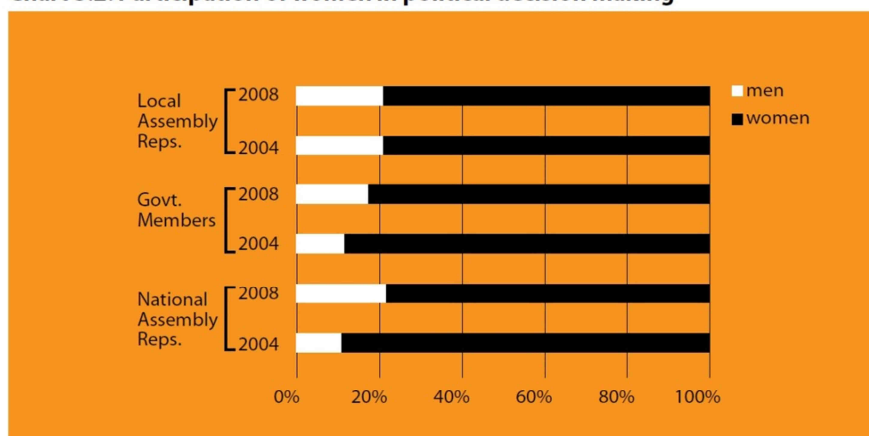
Parliamentary systems differ around the world. One comparability issue for this indicator is the inclusion of only the single or lower chamber of parliament. Metadata are needed to specify where estimates do not include all elected representatives. In case upper chambers of parliament are included, this would be essential metadata.

Examples of current practice

The 2009 MDG report of Serbia clearly indicates the share of women in the different national parliamentary bodies (e.g. Local Assembly, Government, National Assembly), removing any confusion over which houses of parliament are included.

Figure 35. Presenting estimates of women’s share of parliamentary seats

Chart 3.2. Participation of women in political decision making



Source: National Statistical Office: Women and Men in Serbia 2008, 2005; National Statistical Office, Local Elections, 2008.

Source: *Progress of the realization of Millennium Development Goals in the Republic of Serbia* (Krstić, G. and others, 2009).

Indicator 4.2 – Infant mortality rate

<i>Definition</i>	The infant mortality rate (IMR) is the probability of a child born in a specified year dying before reaching the age of one, if subject to current age-specific mortality rates. This indicator is expressed in units of mortality per 1,000 live births.
<i>Concepts</i>	Live birth; infant
<i>Agency responsible for global reporting</i>	UNICEF www.childinfo.org
<i>Relevant classifications</i>	International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10)

Metadata considerations

A variety of data sources may be available to produce estimates for this indicator. These include vital registration systems, DHS and population censuses. Due to underreporting, bias and other issues, estimates for the same period can differ greatly depending on the primary data source that was used. It is therefore essential to include this in the metadata. Knowing the source for each data point being presented is also important for users to be able to compare the estimates over time and between countries or geographical areas.

Estimates from surveys are in most cases based on retrospective data. In most cases these are aggregated because of small sample sizes. The reference period is therefore in general not the year of the survey but the period of three or five years before the survey. The correct reference period is mandatory metadata. The exact method of calculation can in most cases be covered by a reference or be provided in an appendix.

Other examples of possible conditional metadata are:

- the minimum period of gestation, minimum weight and size, and other restrictions that are not in compliance with WHO recommendations
- whether direct or indirect methods are used (the model life table applied in case of indirect methods would be optional metadata).

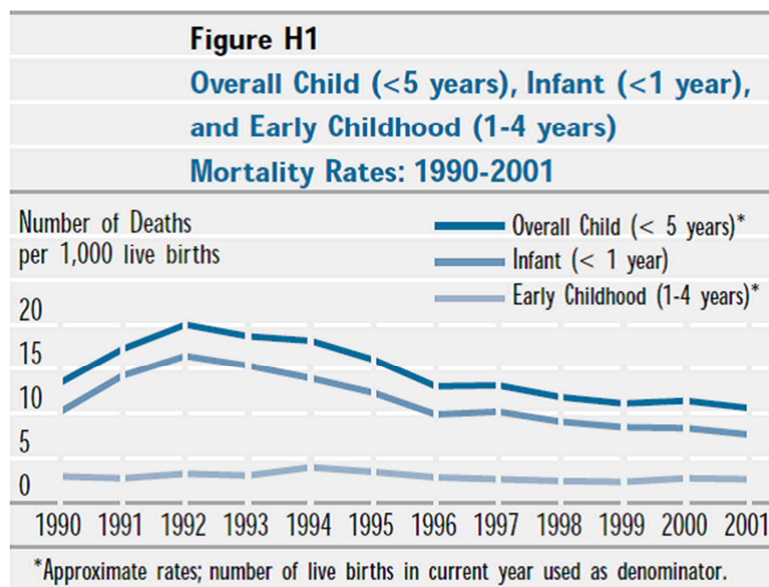
Examples of current practice

It is often assumed that readers know the age ranges for infant and child mortality, but making them explicit, as was done in the MDG report of Lithuania of 2002, is very helpful for users who are less familiar with statistics.

In fact, United Nations reports on child mortality often refer to this indicator as “Under-five mortality rate” (UNICEF and others, 2011).

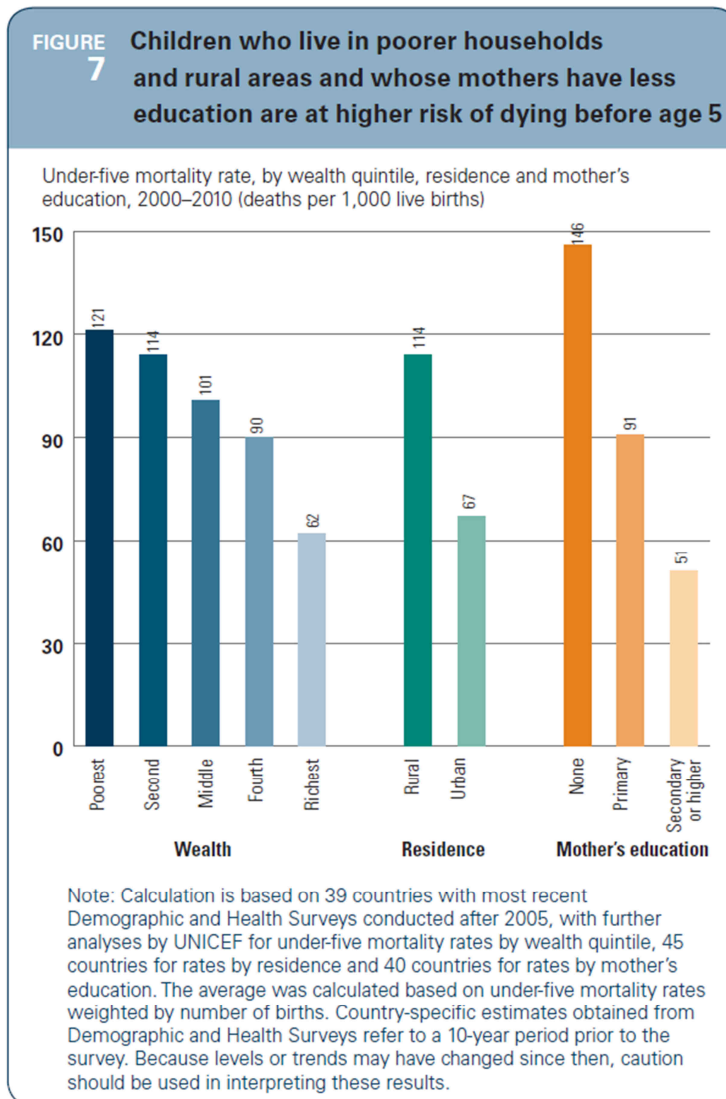
The United Nations report on child mortality, gives a good example of providing methodological information in a footnote (Figure 37). The note explains the data sources and approach taken to producing the data being presented. The result is a powerful graphic that shows a strong link between child mortality, household wealth, rural location and mother's education.

Figure 36. Presenting child and infant mortality



Source: Report on the Millennium Development Goals: a Baseline Study, common country assessment for Lithuania (United Nations, 2002).

Figure 37. Child mortality statistics in United Nations reports



Source: *Levels and Trends in Child Mortality Report 2011 (UNICEF and others, 2011)*.

Indicator 5.1 – Maternal mortality

<i>Definition</i>	The maternal mortality ratio (MMR) is the annual number of maternal deaths from any cause related to or aggravated by pregnancy or its management (excluding accidental or incidental causes) during pregnancy and childbirth or within 42 days of termination of pregnancy, irrespective of the duration and site of the pregnancy, per 100,000 live births, for a specified year.
<i>Concepts</i>	Maternal deaths; live birth
<i>Agency responsible for global reporting</i>	WHO and UNICEF www.who.int/reproductivehealth/en/ www.childinfo.org
<i>Relevant classifications</i>	International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10)

Metadata considerations

As in the case with infant mortality (MDG Indicator 4.2) a variety of data sources may be available to produce estimates for this indicator. These include vital registration systems, DHS and population censuses. Knowing the source for each data point being presented is important for users to be able to compare the estimates over time and between countries or geographical areas.

To overcome uncertainty caused by a low number of observations, data often do not refer to a calendar year but to a longer period. In the case of the sisterhood method they can even refer to retrospective data from a period between the survey data and over 25 years before that. It is therefore important to provide the exact reference period instead of the year of the survey as mandatory metadata.

It is also important to specify if definitions are not the same as internationally recommended. For example, if all deaths, including non-pregnancy-related, during pregnancy or within 42 days are used, this would be important metadata.

Other examples of possible conditional metadata are:

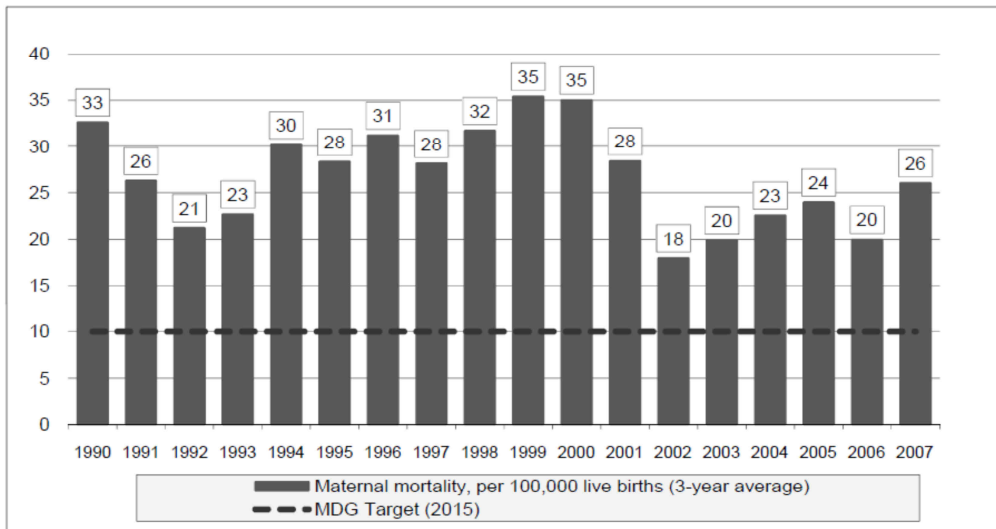
- the minimum period of gestation, minimum weight and size, and other restrictions regarding live births that are not in compliance with WHO recommendations
- if sources are not the same for the numerator (all maternal deaths occurring in a period) and denominator (total number of live births in the same period).

Examples of current practice

Maternal mortality estimates presented in the MDG report of Armenia of 2010 include all **mandatory metadata** and the data source is specified (administrative register).

Figure 38. Maternal mortality estimates in the MDG report of Armenia

Figure 5.1. Maternal Mortality in Armenia in 1990-2007, per 100,000 live births (3-year average)



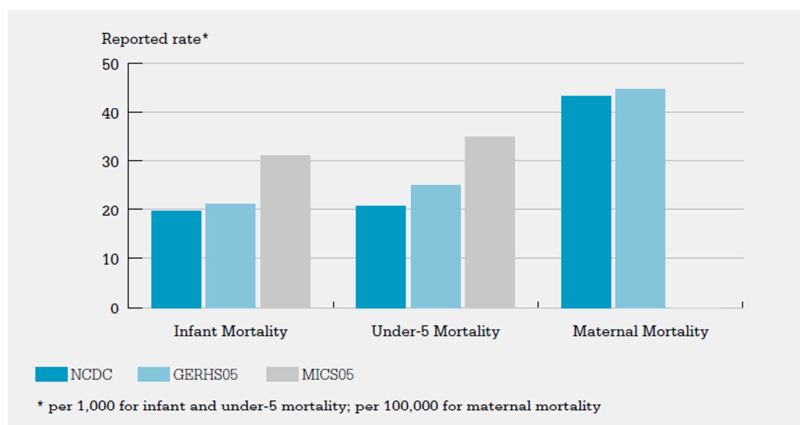
Source: National Statistical Service: data from administrative register.

Source: Armenia Millennium Development Goals national progress report 2005-2009 (Government of Armenia and United Nations Country Team in Armenia Office of the UN Resident Coordinator, 2010).

Multiple sources for certain indicators is a source of confusion for many data users. A report on Georgia’s health system provides an example of presenting multiple sources together, so that similarities and discrepancies are clear. As can be seen in Figure 39, metadata clearly indicate the sources of the data and also the different units of measure between the maternal mortality and child mortality estimates.

Figure 39. Presenting estimates from multiple sources in Georgia

Fig. 8. Comparison of infant, under-five years and maternal mortality rates reported by National Centre for Disease Control (routine health information source) and two population-based surveys, 2005



Sources: National Centre for Disease Control; Women’s Reproductive Health Survey (9); Multiple Indicator Cluster Survey (7)

Source: Georgia health system performance assessment 2009 (World Health Organization, 2009).

Indicator 5.3 – Contraceptive prevalence rate

<i>Definition</i>	The contraceptive prevalence rate is the percentage of women of reproductive age who are currently using, or whose sexual partner is currently using, at least one contraceptive method, regardless of the method used.
<i>Concepts</i>	Women of reproductive age; contraceptive methods
<i>Agency responsible for global reporting</i>	United Nations Population Division and the United Nations Population Fund (UNFPA) www.un.org/esa/population/unpop.htm
<i>Relevant classifications</i>	Not applicable

Metadata considerations

Whether the base population used includes only women who are married or in-union or all women who are at risk of pregnancy should be clarified in the accompanying metadata. Age limits, especially if they are more restricted than women aged 15-49 years, are relevant information.

Furthermore, inconsistent interpretation of the concept of “current use” of contraceptives may lead to under- or overreporting for this indicator. Clarifying the definition will assist data users to identify any comparability issues through the use of different reference periods for “current use” (e.g. left undefined, during last month, at last intercourse).

Survey design, and therefore results, can vary. Providing the correct reference to the primary data source is mandatory. Examples of other possible conditional metadata are:

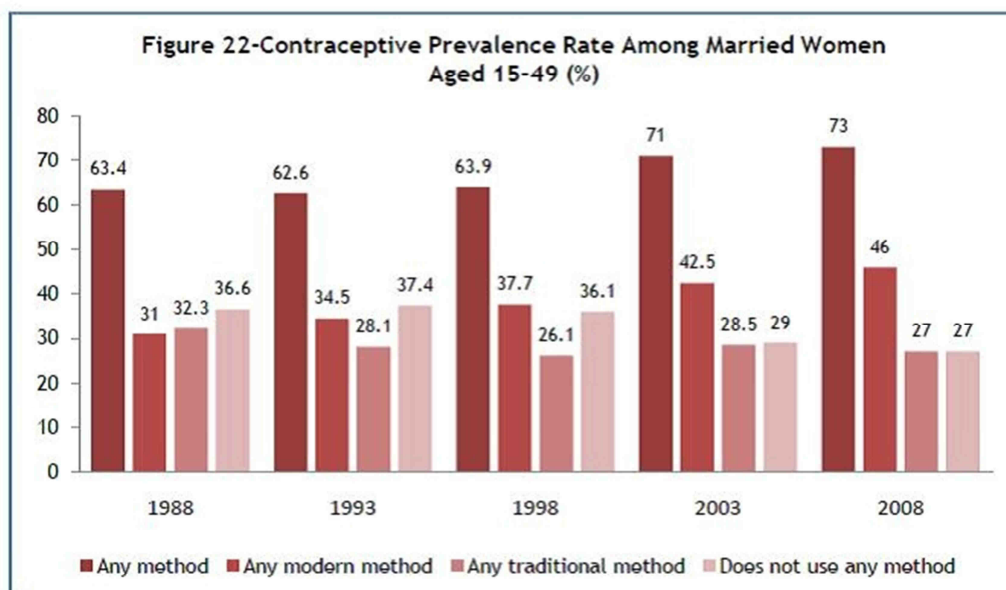
- if a list of specific methods of family planning was provided during the interview
- whether only one or multiple options were considered for each respondent.

Examples of current practice

The example from a national MDG report (Figure 40) includes metadata to specify the age group and the population being measured (married women in this case). Unfortunately, the categories of modern and traditional contraceptives were not explained anywhere in the report and this would be useful information to help readers interpret the data.

Data on contraceptive prevalence published by the Australian Bureau of Statistics provide details on the type of contraceptive method, so that users get meaningful information in addition to the total proportion (Figure 41).

Figure 40. Inadequate presentation: Contraceptive prevalence rate in a national MDG report



Source:

Figure 41. Contraceptive prevalence rate in Australia

Type of contraception used by women aged 18-49, 1995								
Contraceptive method	Age of users							Total
	18-19	20-24	25-29	30-34	35-39	40-44	45-49	
	%	%	%	%	%	%	%	%
Contraceptive pill	66.3	71.1	59.2	43.0	31.3	16.9	10.1	40.0
Condom(a)	32.2	21.6	27.0	21.2	13.9	10.5	7.2	17.6
IUD	*	2.1	*	2.8	3.1	6.1	2.8	3.0
Periodic abstinence	*	*	3.6	4.2	3.9	3.0	2.4	3.0
Other temporary methods	*	2.6	2.4	3.9	3.7	2.1	*	2.6
Female sterilisation	0.0	*	3.6	10.7	21.6	36.1	49.9	19.2
Male sterilisation(a)	0.0	*	2.8	14.2	22.4	25.3	25.8	14.5
	'000	'000	'000	'000	'000	'000	'000	'000
Total women who use contraception	111.3	441.1	428.6	453.7	476.5	448.2	392.6	2 751.9
	%	%	%	%	%	%	%	%
Women who use contraception, as a proportion of all women	49.7	65.7	64.5	65.4	71.6	71.8	67.4	66.7

(a) Used by male partner.

Source: Unpublished data, 1995 National Health Survey.

Source: Australian Social Trends 1998 (Australian Bureau of Statistics, 1998).

Indicator 5.5 – Antenatal care coverage

<i>Definition</i>	Antenatal care coverage (at least one visit) is the percentage of women aged 15-49 with a live birth in a given time period that received antenatal care provided by skilled health personnel at least once during their pregnancy. Antenatal care coverage (at least four visits) is the percentage of women aged 15-49 with a live birth in a given time period that received antenatal care by any provider four or more times during their pregnancy.
<i>Concepts</i>	Live birth; antenatal care; skilled health personnel; traditional birth attendants
<i>Agency responsible for global reporting</i>	WHO and UNICEF www.who.int/reproductivehealth/en/ www.childinfo.org
<i>Relevant classifications</i>	Not applicable

Metadata considerations

A source of inconsistencies in estimates for this indicator is varying definitions for “skilled health personnel”. Clarifying the used definition will help users compare these data to those provided by other sources. Also, the inclusion of traditional birth attendants is another area where national practices differ. Countries should specify if these are included.

References to the main primary data source should be provided as the outcome can differ depending on the survey design. As most surveys collect retrospective data, it is important to provide the correct reference period and not the year of the survey. As with child mortality and maternal mortality, deviations from the recommended definition for live birth have to be covered by metadata.

Further examples of possible metadata are:

- if a time limit is used for the first visit (e.g. within 12 weeks of pregnancy)
- when visits to non-skilled health personnel are not excluded.

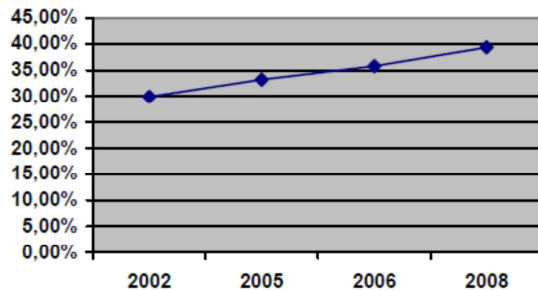
Examples of current practice

The example on antenatal care in a national MDG report shown in Figure 42 lacks mandatory metadata to indicate what is actually being presented. There is no label on the y-axis to indicate what the data are a percentage of. Furthermore, the title of the indicator refers to time of the first examination rather than proportion of women who are examined, leading to confusion as to what these data represent.

Figure 42. Inadequate metadata: Antenatal care statistics in a national MDG report

Target 2: Improve antenatal care

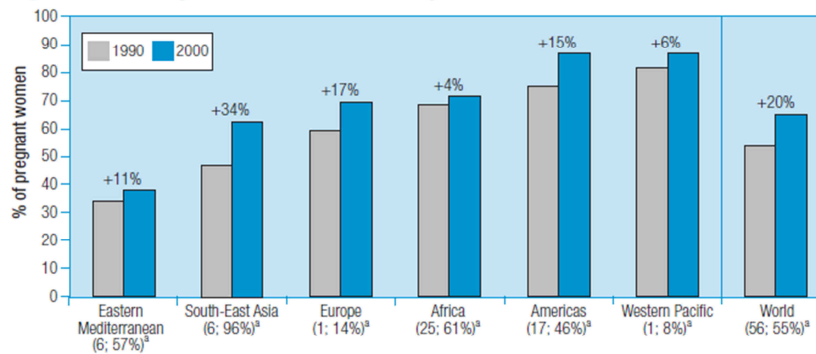
Indicator 2c: Time of the first antenatal examination (the optimum time for the first check is up to the 8th week of pregnancy)



In their annual flagship publication, WHO presents estimates of antenatal care coverage in various regions across the world. Although little metadata accompany this presentation, the important details are all there, including the number of countries in each regional grouping and the percentage of regional population that they represent.

Figure 43. Antenatal care coverage in the World Health Report 2005

Figure 3.1 Coverage of antenatal care is rising



^a Number of countries and percentage of the regional population included in the analysis.
Data source: Multiple Indicator Cluster Surveys (UNICEF) and Demographic and Health Surveys.

Source: *The World Health Report 2005: Make every mother and child count* (WHO, 2005).

Indicator 6.3 – Proportion of population aged 15-24 years with comprehensive correct knowledge of HIV/AIDS

<i>Definition</i>	The proportion of the population aged 15-24 with comprehensive correct knowledge of Human immunodeficiency virus/Acquired immunodeficiency syndrome (HIV/AIDS).
<i>Concepts</i>	Comprehensive correct knowledge; local misconceptions; Human Immunodeficiency Virus (HIV); Acquired Immuno Deficiency Syndrome (AIDS).
<i>Agency responsible for global reporting</i>	UNICEF
<i>Relevant classifications</i>	Not applicable

Metadata considerations

This indicator is based on questioning around “local misconceptions” about HIV/AIDS. Determining which misconceptions to focus on in the questioning of respondents is up to the data collection agency. Explaining how this indicator is calculated will help users to interpret the results. Deviations from the recommended set of questions to determine “comprehensive correct knowledge” or the use of different methodologies to report on this indicator is mandatory metadata.

Other possible conditional metadata are:

- any information reflecting concerns that the data are not based on a representative sample
- non-standard age groups or if non-sexual active persons are excluded.

Examples of current practice

It is helpful for data users to know how the data were collected. In the case of this indicator, providing the five questions used to measure people’s “comprehensive correct knowledge of HIV/AIDS” is important to interpreting the data. The report on MDG progress for the Russian Federation, published in 2010, provided a copy of the questions and their authoritative international source (UNAIDS) (Figure 44).

Figure 44. Providing users with the question used in data collection can aid understanding

Box 6.1. Five questions about HIV

Indicator: Percentage of young people aged 15–24 who both correctly identify ways of preventing the sexual transmission of HIV and who reject major misconceptions about HIV transmission

This indicator is constructed from responses to the following set of prompted questions.

1. Can the risk of HIV transmission be reduced by having sex with only one uninfected partner who has no other partners?
2. Can a person reduce the risk of getting HIV by using a condom every time they have sex?

3. Can a healthy-looking person have HIV?
4. Can a person get HIV from mosquito bites?
5. Can a person get HIV by sharing food with someone who is infected?

Source: United Nations General Assembly Special Session on HIV/AIDS. Monitoring the Declaration of Commitment on HIV/AIDS: guidelines on construction of core indicators: 2010 reporting. March 2009, Geneva, Switzerland

http://data.unaids.org/pub/Manual/2009/jc1676_core_indicators_2009_en.pdf

Source: Millennium Development Goals in Russia: Looking into the Future (Bobylev and others, 2010).

Indicator 6.9 – Incidence, prevalence and death rates associated with tuberculosis

Definition The incidence of tuberculosis (TB) is defined as the estimated number of new TB cases arising in one year. It is expressed as cases per 100,000 population.

The prevalence of tuberculosis refers to the number of TB cases in a population at a given point in time (sometimes referred to as "point prevalence"). It is expressed as cases per 100,000 population.

Death rates associated with tuberculosis are defined as the estimated number of deaths due to TB in a given time period. It is expressed as deaths per 100,000 population.

Concepts Tuberculosis case; tuberculosis

Agency responsible for global reporting WHO
www.who.int/tb/country/global_tb_database

Relevant classifications International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10)

Metadata considerations

Impacting on the accuracy and comparability of estimates for this indicator are different methods of calculating incidence depending upon available data and varying quality of primary data sources. It is important that the source(s) of data and the methodology be clearly specified. If data points from time-series come from different sources or if different methods were applied, such information would be mandatory and listed with the data.

Further examples of possible metadata to be included are:

- information on the reliability and completeness of death registration if used
- information on the completeness of the cases notified.

Examples of current practice

Presenting this indicator on a map, MDG report of Hungary of 2004, shows the prevalence of tuberculosis is higher in the west of the country. Mandatory metadata are present, although more information on the precise source of data would be valuable.

Figure 45. Prevalence of tuberculosis in Hungary

Map 6.1:

Prevalence of tuberculosis – per 100 thousand inhabitants, 2001



Legend

14.46–22.61% 22.62–30.77% 30.78–38.92%

Source: Central Statistical Office (CSO) 2003.

Source: Millennium Development Goals Report Hungary (Medgyesi, 2004).

Providing explanatory notes in an annex is another good approach to providing metadata in national MDG reports. Instead of including a lot of details with the presentation of data within the report, putting metadata, such as definitions, methodology and detailed information on sources, in an annex ensures the information is on hand but does not distract from the key findings.

Figure 46. Explanatory notes in the MDG Report of Hungary of 2004

EXPLANATORY NOTES

GOAL 1
Absolute poverty
The status of a household and/or its members as defined by the volume of incomes the household requires in order to meet the needs of its members. A household is classified as poor when its income falls below the absolute poverty line (i.e. subsistence minimum or wage minimum).

Deciles (decile group)
A decile group is one tenth of all households arranged by their incomes from minimum to maximum. The first decile group is the first one tenth (the 10% of all household with lowest incomes). The last decile is the one tenth of the households with the highest incomes.

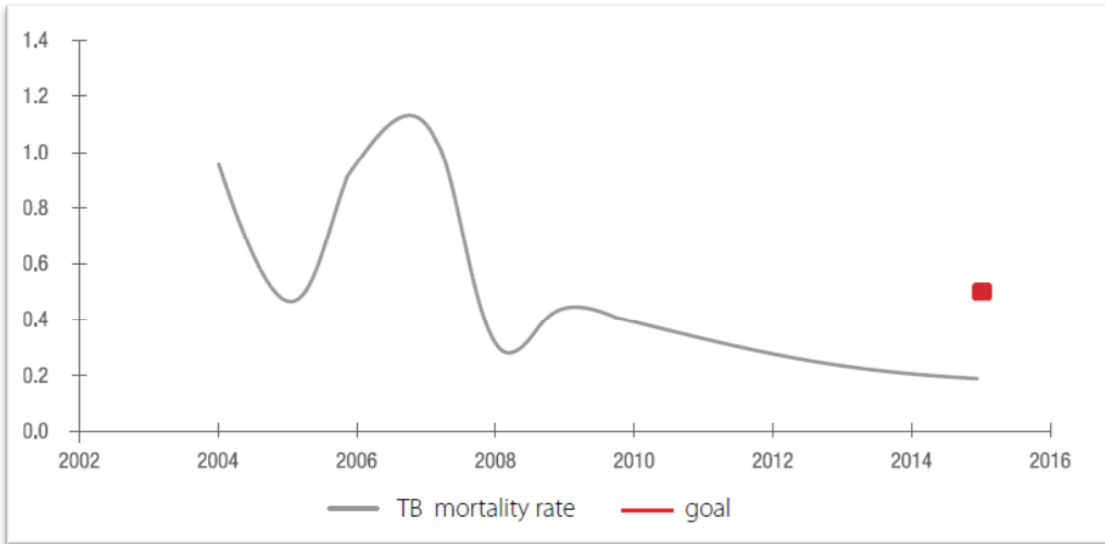
Gini Coefficient
The Gini coefficient measures the degree of inequality of the distribution of earnings. It is equal to zero in the case of total earnings equality and to one in the case of total inequality.

Source: Millennium Development Goals Report Hungary (Medgyesi, 2004).

Clear and simple presentations of data are often the most effective, particularly for MDG-related statistics, which are used by a broad audience. However, simple presentations are useless if they lack mandatory metadata needed for accurate interpretation. The graph shown in Figure 47 of a national MDG report does not include a title describing what the data relate to, nor are there labels on the y-axis to clarify the

unit of measure. The presentation would be more meaningful if the source data were acknowledged and if possible, notes to explain the spike in tuberculosis mortality rates that occurred in 2007.

Figure 47. Inadequate metadata: Tuberculosis mortality rate in national MDG report



Indicator 7.8 – Proportion of population using an improved drinking water source

Definition The proportion of population using an improved drinking water source is the percentage of the population who use any of the following types of drinking water supplies: piped water into dwelling, plot or yard; public tap/standpipe; borehole/tube well; protected dug well; protected spring; rainwater collection and bottled water (bottled water is included if a secondary available source is also improved).

This indicator is expressed in units of percentage.

Concepts Improved drinking water source; drinking water

Agency responsible for global reporting WHO and UNICEF Joint Monitoring Programme for Water Supply and Sanitation (JMP)
www.wssinfo.org

Relevant classifications JMP types of drinking water sources
www.wssinfo.org/definitions-methods/watsan-categories/

Metadata considerations

Understanding the concept of “improved” and “unimproved” water sources is essential to being able to interpret estimates for this indicator. As definitions may vary from country to country, metadata to describe the water sources included in each category should be provided. The primary data source is therefore relevant metadata and if non-harmonized questions and categories are used, additional information should be given on comparability with recommended methodologies. In case of administrative sources, an estimation of the coverage and recency of the data is useful information.

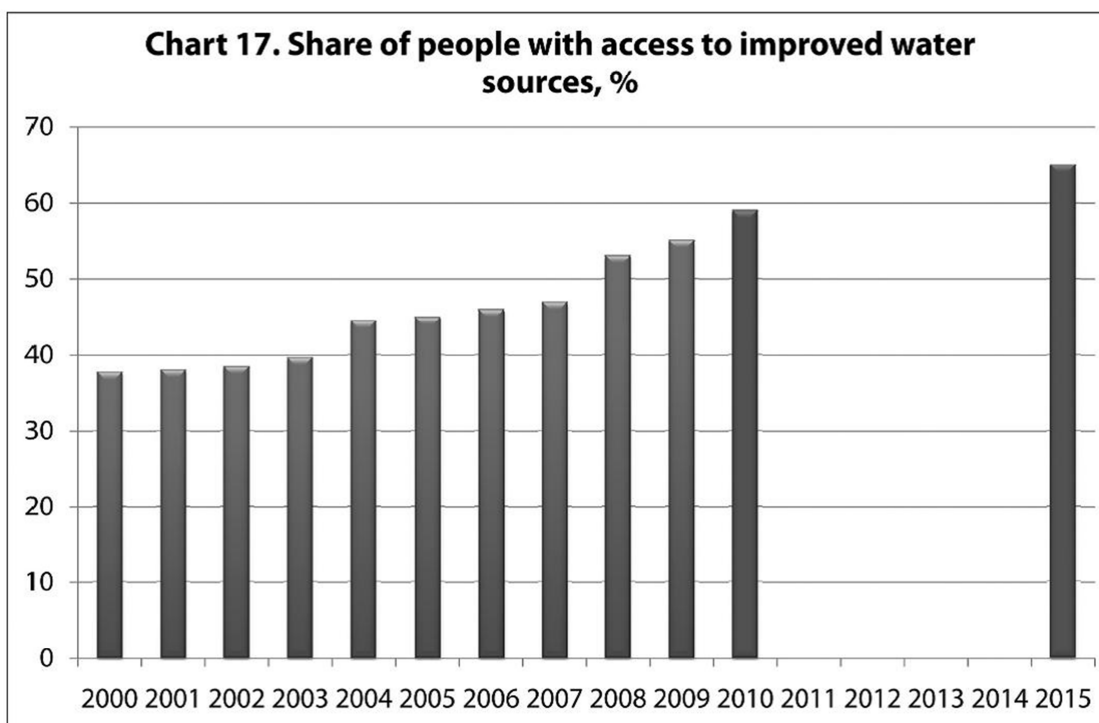
Terminology should be used correctly and for example improved and safe water sources should not be used interchangeable. As national figures vary greatly in exactly which types of water supply are included, it is important to provide a clear definition. In case time-series are presented with data from different sources or different exact definitions, these metadata have to be provided with the data and cannot be covered by a reference alone.

Examples of current practice

Figure 48 shows an example of the presentation of data on access to improved water sources in a national MDG report. It only provides most necessary metadata. It would be enhanced by including a label on the y-axis and explaining the data source(s) used. Unfortunately the use of colour in this graphic was lost at some point during publication and the reference to “red” for intermediary and final target values is not understandable. The report uses both safe as well as improved water interchangeably

and does not provide a definition. No primary data source is provided so it cannot be determined whether the data is comparable. Additionally, the source provided with the graph differs from the source given in the data table in the annex of the report.

Figure 48. Inadequate example: Access to improved drinking water



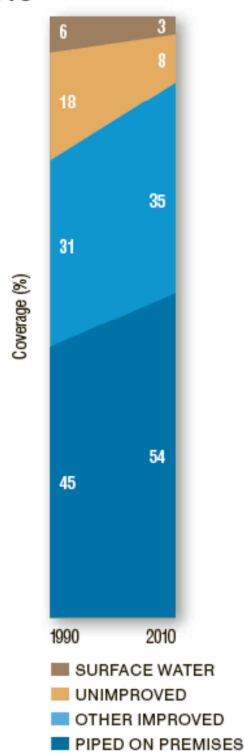
Note: red shows intermediary and final target-values of the indicator;

Source: [REDACTED]

The United Nations Joint Monitoring Program (JMP) presents global estimates for this indicator using an easy-to-read graphic that shows total proportion and increase or decrease in use of various water sources (Figure 49). The JMP publications provide optional metadata, which describe the estimation methodology used, definitions and general statements about data sources, in a separate chapter at the end of their report.

Figure 49. Presenting international estimates of the proportion of population with access to an improved water source

Drinking water coverage increased from 76 per cent in 1990 to 89 per cent in 2010



Estimation methodology, definitions and other important metadata provided at the end of

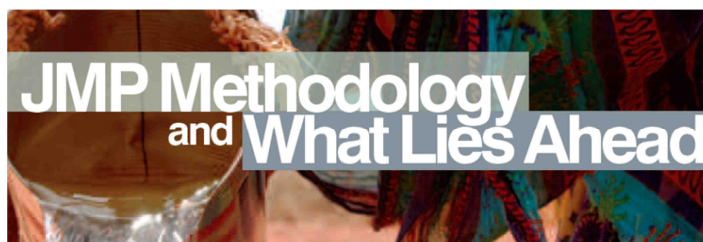


FIGURE 2 Trend in the proportion of the global population using piped drinking water on premises, other improved drinking water sources, unimproved sources and surface water, 1990-2010

JMP Estimates

The WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation is tasked with providing estimates that are comparable among countries and across time. Because definitions of 'improved' sanitation facilities and drinking water sources can vary widely among countries, the JMP has established a standard set of categories that are used to analyse national data on which the MDG trends and estimates are based (Table 6).

The population data used in this report, including the proportion of the population living in urban and rural areas, are those established by the United Nations Population Division, 2010 Revision.

The definitions and data sources used by the JMP are often different from those used by national governments.

	Drinking Water	Sanitation
Improved	Use of: <ul style="list-style-type: none"> ■ Piped water into dwelling, yard or plot ■ Public tap or standpipe ■ Tubewell or borehole ■ Protected spring ■ Protected dug well ■ Rainwater collection 	Use of: <ul style="list-style-type: none"> ■ Flush or pour-flush to: <ul style="list-style-type: none"> – Piped sewer system – Septic tank – Pit latrine ■ Ventilated improved pit (VIP) latrine ■ Pit latrine with slab ■ Composting toilet
Unimproved	Use of: <ul style="list-style-type: none"> ■ Unprotected dug well ■ Unprotected spring ■ Cart with small tank or drum ■ Tanker truck ■ Surface water (river, dam, lake, pond, stream, canal, irrigation channel) ■ Bottled water (considered to be improved only when the household uses drinking water from an improved source for cooking and personal hygiene) 	Use of: <ul style="list-style-type: none"> ■ Flush or pour-flush to elsewhere (that is, not to piped sewer system, septic tank or pit latrine) ■ Pit latrine without slab, or open pit ■ Bucket ■ Hanging toilet or hanging latrine ■ Shared or public facilities of any type ■ No facilities, bush or field (open defecation)

TABLE 6 Definitions of improved and unimproved drinking water sources and sanitation facilities

Source: Progress on Drinking Water and Sanitation: 2012 Update (UNICEF and WHO, 2012).

Indicator 8.16 – Internet users per 100 population

<i>Definition</i>	The indicator is the number of Internet users per 100 population.
<i>Concepts</i>	Internet; Internet user
<i>Agency responsible for global reporting</i>	International Telecommunications Union (ITU) www.itu.int/ITU-D/ict/
<i>Relevant classifications</i>	Not applicable

Metadata considerations

Varying age ranges used for total population (e.g. 15-74 years old) can impact on the comparability of estimates for this indicator. Clearly specifying the age range on which the estimates are based will help with interpretability. The internationally recommended definition of internet user is a person that used the Internet in the previous 12 months from any location. This includes Internet use by any device enabling Internet access such as a computer, a mobile phone, personal digital assistant (PDA), games device and digital TV. Use can be via a fixed or mobile network. It is important to note deviations from this concept. The primary data source also determines comparability. Whether household surveys were used, or if data are estimated from subscriptions, is important metadata. In the latter case, it is necessary to provide the methodology that was used for such estimations.

Examples of current practice

A country included estimates for this indicator in their 2010 MDG report. The presentation includes comparison against other countries of the region as well as those in the Commonwealth of Independent States (CIS).

Accompanying metadata are quite limited and the combination of annual figures together with average annual growth, is confusing. Inclusion of metadata to clarify the unit of measure on the y-axis, the age group of the population being measured and the definition of internet use and data source(s) would be helpful (Figure 50).

A United Nations publication on MDG progress presents global estimates of internet usage as shown below (Figure 51). All mandatory metadata are included. The presentation would be enhanced with a footnote to explain why mobile phone subscriptions are higher than 100 per cent of the total population. Also, further information on the sources of data and geographical coverage would be helpful. It should also be noted that the title refers to mobile cellular subscriptions while the legend refers to mobile cellular subscribers (a subscriber can have more than one subscription).

Figure 50. Inadequate presentation: The number of internet users in a national MDG report

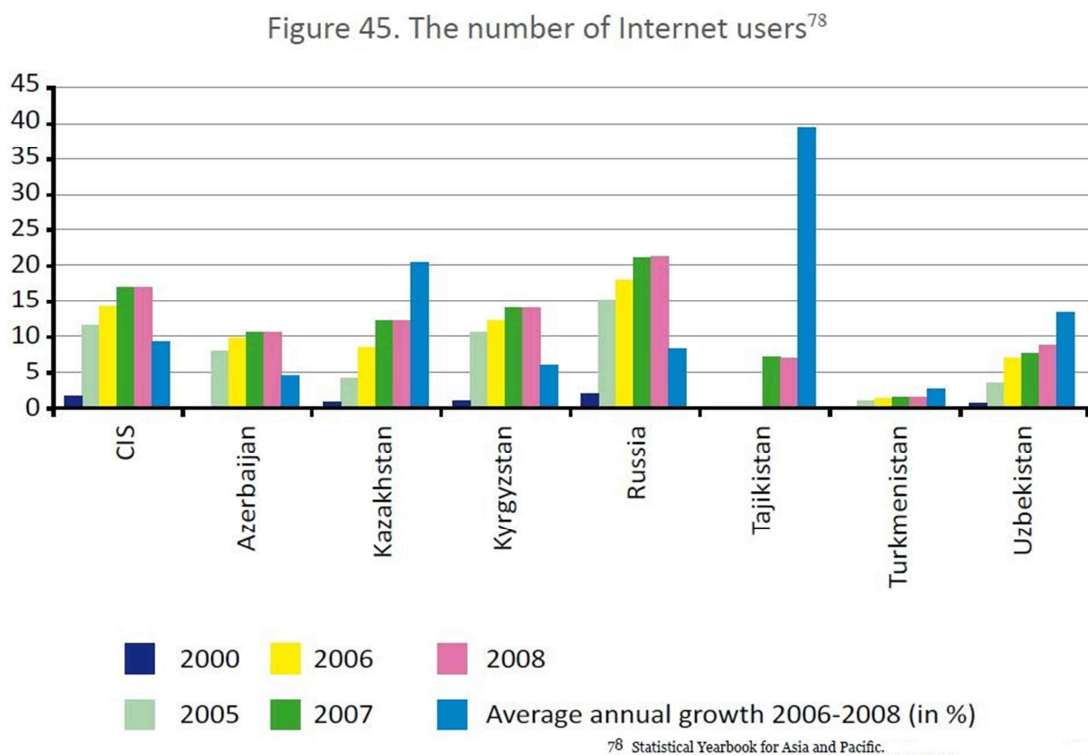
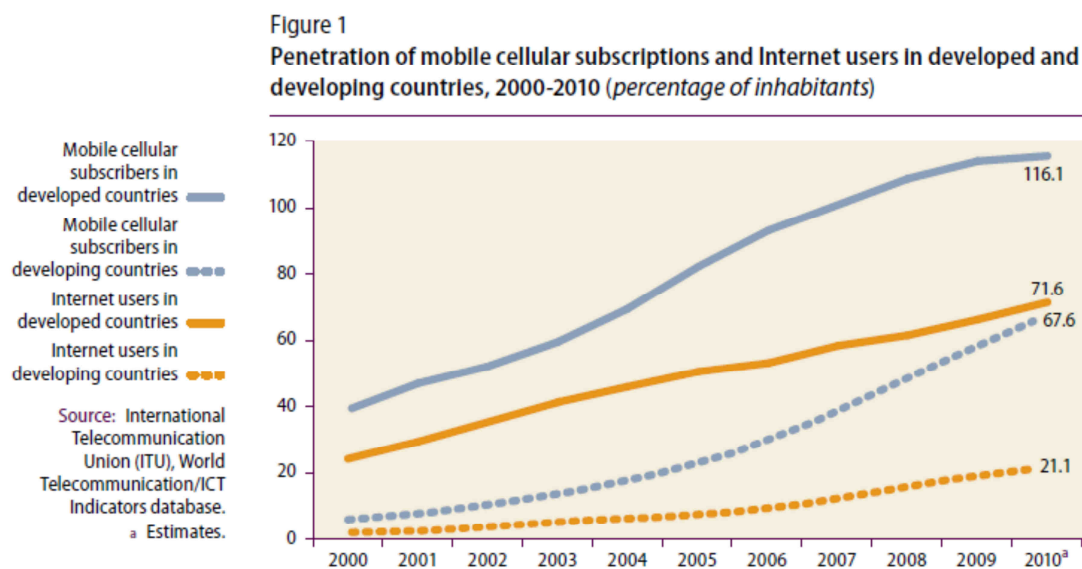


Figure 51. Presentation of data on mobile phone and internet use



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

VII. Recommendations

Countries of the UNECE region have varying capacity in producing statistics. The region includes some of the most developed and well-resourced national statistical systems, and some that require ongoing capacity development to reach international and European standards in data quality. However, MDG reports often have to be produced by the least developed statistical systems. Recommendations to improve the availability and presentation of development related metadata are provided below.

For national statistical organizations and other bodies reporting data on development:

Develop presentation guidelines for data and metadata

Agencies involved in reporting on development progress should develop clear guidelines for data and metadata presentation. These should specify the types of metadata to be included in progress and monitoring reports. The recommended minimum and optimum metadata outlines in this publication can provide a basis for national standards.

Common terminology

To ensure consistent and clear presentation of official statistics, a glossary or reference of common statistical terms should be implemented for use across the national statistical system (i.e. all national data producing agencies and publishers). These terms should be based on international standards, such as the Metadata Common Vocabulary and the OECD Glossary of Statistical Terms (OECD, 2008).

Develop a metadata management system

Statistical organizations should have a metadata management strategy and systems in place for the efficient production of high quality data and metadata. Organizations should develop or revise their metadata management practices in accordance with international recommendations.

The IAEG-MDG provides valuable guidance and support to countries in the effective monitoring and reporting on MDG indicators. Recommendations to international agencies include:

Agree on international standards for development data and metadata presentation

Reports on progress in publications by international organizations should comply with international recommendations on metadata presentation. To ensure the group leads by example, the appropriate amount of metadata for progress reports should be agreed and standards applied to international reports.

Emphasise the importance of good metadata

The valuable guides produced by international agencies on the reporting of MDG estimates would be enhanced with the inclusion of practical examples on how

data and metadata should be presented. Greater emphasis should be placed on the importance of including metadata with national progress reports.

Develop capacity in metadata management and presentation of statistics

Support for capacity development in monitoring and reporting should extend to enhancing skills in metadata management and effective presentation of statistics.

References

- Australian Bureau of Statistics (1998). Australian Social Trends 1998.
- Bobylev, S.N. and others (2010). Millennium Development Goals in Russia: Looking into the Future.
- Bosnia and Herzegovina Ministry of Finance and Treasury and United Nations Country Team in Bosnia and Herzegovina (2010). Progress towards the realization of the Millennium Development Goals in Bosnia and Herzegovina 2010.
- Czech Statistical Office (2009). Update of Czech Republic's Case Study, Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata, 11-13 March 2009. Available from www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.40/2009/mtg1/wp.20.e.pdf.
- Dupriez, O. and Greenwell, G. (2007). Quick Reference Guide for Archivists, Draft Version 2007.03, June 2007, International Household Survey Network.
- Eurostat (2010). Analysis of the 10 Core MDG Indicators: Methodology and Quality Issues – Countries of ECOWAS (including UEMOA) Regional Organisation. Available from epp.eurostat.ec.europa.eu/portal/pls/portal/!PORTAL.wwpob_page.show?_docname=2358277.PDF
- Government of Armenia and United Nations Country Team in Armenia Office of the UN Resident Coordinator (2010). Armenia Millennium Development Goals national progress report 2005-2009.
- International Labour Organization (2009). Guide to the new Millennium Development Goals Employment Indicators. Available from www.ilo.org/wcmsp5/groups/public/---ed_emp/documents/publication/wcms_110511.pdf
- International Labour Organization (2011). Key Indicators of the Labour Market, Seventh Edition. Available from www.ilo.org/empelm/what/WCMS_114240/lang-en/index.htm
- International Labour Organization (2013). Guide to the millennium development goals employment indicators. International Labour Office, Employment Sector. Available from www.ilo.org/wcmsp5/groups/public/---ed_emp/---emp_elm/documents/publication/wcms_208796.pdf
- Krstić, G. and others (2009). Progress of the realization of Millennium Development Goals in the Republic of Serbia.
- Medgyesi, M. (2004). Millennium Development Goals Report Hungary.
- Millennium Development Goals in Kazakhstan (2005).
- Millennium Development Goals in Kazakhstan (2007).

- Ministry of Foreign Affairs of the Republic of Latvia in cooperation with the United Nations Development Programme (2005). Life in 2015: the Latvia MDG Report.
- Organisation for Economic Co-operation and Development (2006). Employment Outlook. Available from www.oecd.org/dataoecd/53/15/36900060.pdf
- Organisation for Economic Co-operation and Development (2007). Data and Metadata Reporting and Presentation Handbook. Available from:
www.oecd.org/dataoecd/46/17/37671574.pdf
- Organisation for Economic Co-operation and Development (2008). Glossary of Statistical Terms. Updated version available from stats.oecd.org/glossary/
- SDMX (2009a). SDMX Content-Oriented Guidelines Annex 1: Cross Domain Concepts. Available from sdmx.org/wp-content/uploads/2009/01/01_sdmx_cog_annex_1_cdc_2009.pdf
- SDMX (2009b). SDMX Content-Oriented Guidelines Annex 4: Metadata Common Vocabulary. Available from sdmx.org/wp-content/uploads/2009/01/04_sdmx_cog_annex_4_mcv_2009.pdf
- Statistics Portugal (2009). Do conceptual systems improve concepts effectiveness? Joint UNECE/Eurostat/OECD Work Session on Statistical Metadata, 11-13 March 2009. Available from
www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.40/2009/mtg1/wp.14.e.pdf
- The second periodic progress report on the Millennium Development Goals in the Kyrgyz Republic (2009).
- UNECE (2000a). Guidelines for Statistical Metadata on the Internet, Conference of European Statistician Statistical Standards and Studies No. 52. Available from www.unece.org/fileadmin/DAM/stats/publications/metadata.pdf
- UNECE (2000b). Terminology on Statistical Metadata, Conference of European Statistician Statistical Standards and Studies No. 53. Available from www.unece.org/fileadmin/DAM/stats/publications/53metadaterminology.pdf
- UNECE (2009a). Making Data Meaningful Part 2: a guide to presenting statistics. Available from www.unece.org/fileadmin/DAM/stats/documents/writing/MDM_Part2_English.pdf
- UNECE (2009b). Generic Statistical Business Process Model (GSBPM). Available from www.unece.org/stats/gsbpm
- UNECE (2009c). Statistical Metadata in a Corporate Context: A guide for managers.
- UNECE (2012). Common Metadata Framework. Available from www.unece.org/stats/cmf
- UNESCO Institute for Statistics (2011). Global Education Digest 2011: Comparing Education Statistics Across the World.

- UNICEF (2009). Tracking Progress on Child and Maternal Nutrition: A survival and development priority.
- UNICEF and others (2011). Levels and Trends in Child Mortality Report 2011.
- UNICEF and WHO (2012). Progress on Drinking Water and Sanitation: 2012 Update.
- United Nations (2002). Report on the Millennium Development Goals: a Baseline Study, common country assessment for Lithuania.
- United Nations (2003). Indicators for Monitoring the Millennium Development Goals.
- United Nations (2008). The Millennium Development Goals Report 2008.
- United Nations (2011). Millennium Development Goal 8 – The Global Partnership for Development: Time to Deliver, MDG Gap Task Force Report 2011.
- United Nations Development Programme (2007). Measuring Human Development: a primer. Available from hdr.undp.org/en/media/Primer_complete.pdf
- World Health Organization (2005). The World Health Report: 2005: Make every mother and child count.
- World Health Organization (2006). Reproductive Health Indicators: Guidelines for their generation, interpretation and analysis for global monitoring. Available from whqlibdoc.who.int/publications/2006/924156315X_eng.pdf
- World Health Organization (2009). Georgia health system performance assessment 2009.

Annexes

I. Euro-SDMX Metadata Standard (ESMS)

The ESMS provides a comprehensive framework that can be used to guide countries on the metadata that should be captured for all statistical series, including those related to measuring progress towards the MDGs. That information, or subsets of it, can be presented with data that is disseminated. There are 21 metadata elements included in the ESMS:

- 1. Contact**
 - 1.1. Contact organisation
 - 1.2. Contact organisation unit
 - 1.3. Contact name
 - 1.4. Contact person function
 - 1.5. Contact mail address
 - 1.6. Contact e-mail address
 - 1.7. Contact phone number
 - 1.8. Contact fax number
- 2. Metadata update**
 - 2.1. Metadata last certified
 - 2.2. Metadata last posted
 - 2.3. Metadata last update
- 3. Statistical presentation**
 - 3.1. Data description
 - 3.2. Classification system
 - 3.3. Sector coverage
 - 3.4. Statistical concepts and definitions
 - 3.5. Statistical unit
 - 3.6. Statistical population
 - 3.7. Reference area
 - 3.8. Time coverage
 - 3.9. Base period
- 4. Unit of measure**
- 5. Reference period**
- 6. Institutional mandate**
 - 6.1. Legal acts and other agreements
 - 6.2. Data sharing
- 7. Confidentiality**
 - 7.1. Confidentiality – policy
 - 7.2. Confidentiality – data treatment
- 8. Release policy**
 - 8.1. Release calendar
 - 8.2. Release calendar access
 - 8.3. User access
- 9. Frequency of dissemination**
- 10. Dissemination format**
 - 10.1. News release
 - 10.2. Publications
 - 10.3. Online database
 - 10.4. Microdata access
 - 10.5. Other
- 11. Accessibility of documentation**
 - 11.1. Documentation on methodology
 - 11.2. Quality documentation
- 12. Quality management**
 - 12.1. Quality assurance
 - 12.2. Quality assessment
- 13. Relevance**
 - 13.1. User needs
 - 13.2. User satisfaction
 - 13.3. Completeness
- 14. Accuracy**

- 14.1. Overall accuracy
- 14.2. Sampling error
- 14.3. Non-sampling error
- 15. Timeliness and punctuality**
 - 15.1. Timeliness
 - 15.2. Punctuality
- 16. Comparability**
 - 16.1. Comparability – geographical
 - 16.2. Comparability over time
- 17. Coherence**
 - 17.1. Coherence – cross domain
 - 17.2. Coherence – internal
- 18. Cost and burden**
- 19. Data revision**
 - 19.1. Data revision – policy
 - 19.2. Data revision – practice
- 20. Statistical processing**
 - 20.1. Source data
 - 20.2. Frequency of data collection
 - 20.3. Data collection
 - 20.4. Data validation
 - 20.5. Data compilation
 - 20.6. Adjustment
- 21. Comment**

For more information on the ESMS refer to the Eurostat website at: epp.eurostat.ec.europa.eu/portal/page/portal/statistics/metadata/metadata_structure/.22

II. Official list of MDG indicators

All indicators should be disaggregated by sex and urban/rural as far as possible.

Effective 15 January 2008

Millennium Development Goals (MDGs)		
Goals and Targets (from the Millennium Declaration) ⁱ	Indicators for monitoring progress	
Goal 1: Eradicate extreme poverty and hunger		
Target 1.A: Halve, between 1990 and 2015, the proportion of people whose income is less than one dollar a day	1.1	Proportion of population below \$1 (PPP) per day ⁱⁱ
	1.2	Poverty gap ratio
	1.3	Share of poorest quintile in national consumption
Target 1.B: Achieve full and productive employment and decent work for all, including women and young people	1.4	Growth rate of GDP per person employed
	1.5	Employment-to-population ratio
	1.6	Proportion of employed people living below \$1 (PPP) per day
	1.7	Proportion of own-account and contributing family workers in total employment
Target 1.C: Halve, between 1990 and 2015, the proportion of people who suffer from hunger	1.8	Prevalence of underweight children under-five years of age
	1.9	Proportion of population below minimum level of dietary energy consumption
Goal 2: Achieve universal primary education		
Target 2.A: Ensure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling	2.1	Net enrolment ratio in primary education
	2.2	Proportion of pupils starting grade 1 who reach last grade of primary
	2.3	Literacy rate of 15-24 year-olds, women and men
Goal 3: Promote gender equality and empower women		
Target 3.A: Eliminate gender disparity in primary and secondary education, preferably by 2005, and in all levels of education no later than 2015	3.1	Ratios of girls to boys in primary, secondary and tertiary education
	3.2	Share of women in wage employment in the non-agricultural sector
	3.3	Proportion of seats held by women in national parliament

Millennium Development Goals (MDGs)

Goals and Targets (from the Millennium Declaration)ⁱ	Indicators for monitoring progress
Goal 4: Reduce child mortality	
Target 4.A: Reduce by two-thirds, between 1990 and 2015, the under-five mortality rate	4.1 Under-five mortality rate
	4.2 Infant mortality rate
	4.3 Proportion of 1 year-old children immunised against measles
Goal 5: Improve maternal health	
Target 5.A: Reduce by three quarters, between 1990 and 2015, the maternal mortality ratio	5.1 Maternal mortality ratio
	5.2 Proportion of births attended by skilled health personnel
Target 5.B: Achieve, by 2015, universal access to reproductive health	5.3 Contraceptive prevalence rate
	5.4 Adolescent birth rate
	5.5 Antenatal care coverage (at least one visit and at least four visits)
	5.6 Unmet need for family planning
Goal 6: Combat HIV/AIDS, malaria and other diseases	
Target 6.A: Have halted by 2015 and begun to reverse the spread of HIV/AIDS	6.1 HIV prevalence among population aged 15-24 years
	6.2 Condom use at last high-risk sex
	6.3 Proportion of population aged 15-24 years with comprehensive correct knowledge of HIV/AIDS
	6.4 Ratio of school attendance of orphans to school attendance of non-orphans aged 10-14 years
Target 6.B: Achieve, by 2010, universal access to treatment for HIV/AIDS for all those who need it	6.5 Proportion of population with advanced HIV infection with access to antiretroviral drugs
Target 6.C: Have halted by 2015 and begun to reverse the incidence of malaria and other major diseases	6.6 Incidence and death rates associated with malaria
	6.7 Proportion of children under 5 sleeping under insecticide-treated bednets
	6.8 Proportion of children under 5 with fever who are treated with appropriate anti-malarial drugs
	6.9 Incidence, prevalence and death rates associated with tuberculosis
	6.10 Proportion of tuberculosis cases detected and cured under directly observed treatment short course

Millennium Development Goals (MDGs)

Goals and Targets (from the Millennium Declaration) ⁱ	Indicators for monitoring progress
Goal 7: Ensure environmental sustainability	
Target 7.A: Integrate the principles of sustainable development into country policies and programmes and reverse the loss of environmental resources	7.1 Proportion of land area covered by forest 7.2 CO2 emissions, total, per capita and per \$1 GDP (PPP) 7.3 Consumption of ozone-depleting substances
Target 7.B: Reduce biodiversity loss, achieving, by 2010, a significant reduction in the rate of loss	7.4 Proportion of fish stocks within safe biological limits 7.5 Proportion of total water resources used 7.6 Proportion of terrestrial and marine areas protected 7.7 Proportion of species threatened with extinction
Target 7.C: Halve, by 2015, the proportion of people without sustainable access to safe drinking water and basic sanitation	7.8 Proportion of population using an improved drinking water source 7.9 Proportion of population using an improved sanitation facility
Target 7.D: By 2020, to have achieved a significant improvement in the lives of at least 100 million slum dwellers	7.10 Proportion of urban population living in slums ⁱⁱⁱ
Goal 8: Develop a global partnership for development	
Target 8.A: Develop further an open, rule-based, predictable, non-discriminatory trading and financial system	<i>Some of the indicators listed below are monitored separately for the least developed countries (LDCs), Africa, landlocked developing countries and small island developing States.</i>
Includes a commitment to good governance, development and poverty reduction – both nationally and internationally	Official development assistance (ODA)
Target 8.B: Address the special needs of the least developed countries	8.1 Net ODA, total and to the least developed countries, as percentage of OECD/DAC donors' gross national income 8.2 Proportion of total bilateral, sector-allocable ODA of OECD/DAC donors to basic social services (basic education, primary health care, nutrition, safe water and sanitation) 8.3 Proportion of bilateral official development assistance of OECD/DAC donors that is untied 8.4 ODA received in landlocked developing countries as a proportion of their gross national incomes 8.5 ODA received in small island developing States as a proportion of their gross national incomes
Includes: tariff and quota free access for the least developed countries' exports; enhanced programme of debt relief for heavily indebted poor countries (HIPC) and cancellation of official bilateral debt; and more generous ODA for countries committed to poverty reduction	
Target 8.C: Address the special needs of landlocked developing countries and small island developing States (through the Programme of Action for the Sustainable	

Millennium Development Goals (MDGs)

Goals and Targets (from the Millennium Declaration) ⁱ	Indicators for monitoring progress
Development of Small Island Developing States and the outcome of the twenty-second special session of the General Assembly)	<p>Market access</p> <p>8.6 Proportion of total developed country imports (by value and excluding arms) from developing countries and least developed countries, admitted free of duty</p> <p>8.7 Average tariffs imposed by developed countries on agricultural products and textiles and clothing from developing countries</p> <p>8.8 Agricultural support estimate for OECD countries as a percentage of their gross domestic product</p> <p>8.9 Proportion of ODA provided to help build trade capacity</p> <p>Debt sustainability</p> <p>8.10 Total number of countries that have reached their HIPC decision points and number that have reached their HIPC completion points (cumulative)</p> <p>8.11 Debt relief committed under HIPC and MDRI Initiatives</p> <p>8.12 Debt service as a percentage of exports of goods and services</p> <p>8.13 Proportion of population with access to affordable essential drugs on a sustainable basis</p> <p>8.14 Fixed telephone lines per 100 inhabitants</p> <p>8.15 Mobile cellular subscriptions per 100 inhabitants</p> <p>8.16 Internet users per 100 inhabitants</p>
Target 8.D: Deal comprehensively with the debt problems of developing countries through national and international measures in order to make debt sustainable in the long term	
Target 8.E: In cooperation with pharmaceutical companies, provide access to affordable essential drugs in developing countries	
Target 8.F: In cooperation with the private sector, make available the benefits of new technologies, especially information and communications	

ⁱ The Millennium Development Goals and targets come from the Millennium Declaration, signed by 189 countries, including 147 heads of State and Government, in September 2000 (www.un.org/millennium/declaration/ares552e.htm) and from further agreement by member states at the 2005 World Summit (Resolution adopted by the General Assembly – A/RES/60/1, www.un.org/Docs/journal/asp/ws.asp?m=A/RES/60/1). The goals and targets are interrelated and should be seen as a whole. They represent a partnership between the developed countries and the developing countries “to create an environment – at the national and global levels alike – which is conducive to development and the elimination of poverty”

ⁱⁱ For monitoring country poverty trends, indicators based on national poverty lines should be used, where available.

ⁱⁱⁱ The actual proportion of people living in slums is measured by a proxy, represented by the urban population living in households with at least one of the four characteristics: (a) lack of access to improved water supply; (b) lack of access to improved sanitation; (c) overcrowding (3 or more persons per room); and (d) dwellings made of non-durable material.