UNSD-DFID PROJECT ON SDG MONITORING

MODULE 2

Introduction to metadata

Module structure

- What is metadata?
- Why is metadata important?
- Some issues relevant to metadata
- Examples

What is metadata?

What is metadata?

- Quite simply, metadata are **data that defines or describes other data.**
- Metadata **helps explain and understand the data** or values being presented.

The **African Charter for Statistics**:

*"the range of information, **generally textual**, that fosters understanding of the context in which statistical data have been collected, processed and analyzed with the objective of creating statistical information …"*

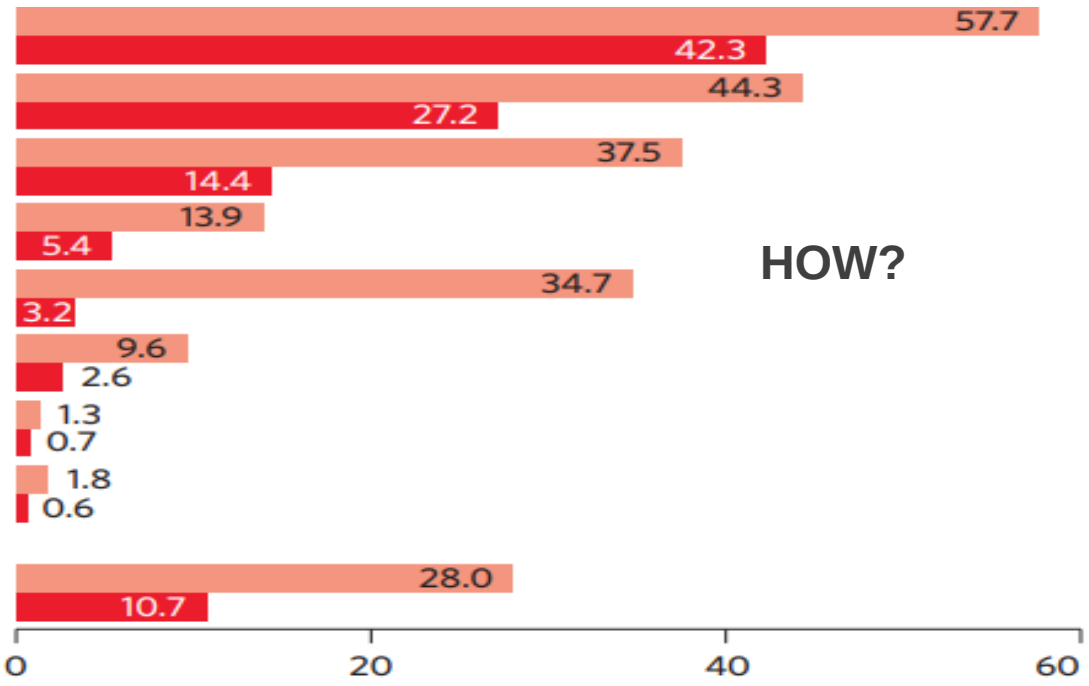
What is metadata

- **Data labels, definitions, description of methodology, legends, source information, footnotes** are all examples of metadata.

But why does it make a difference?

Why does metadata make a difference?

WHAT?



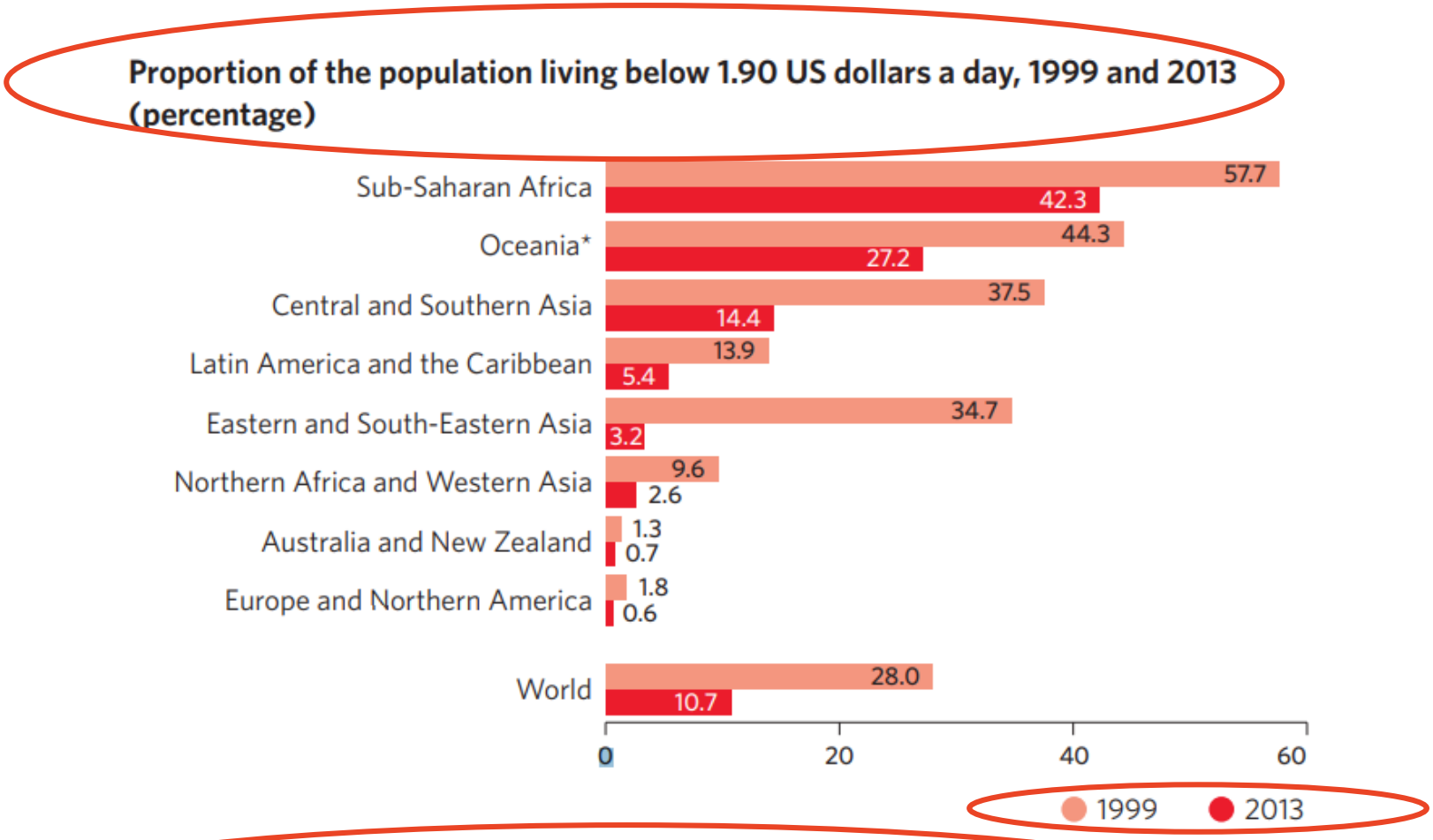
HOW?

WHEN?

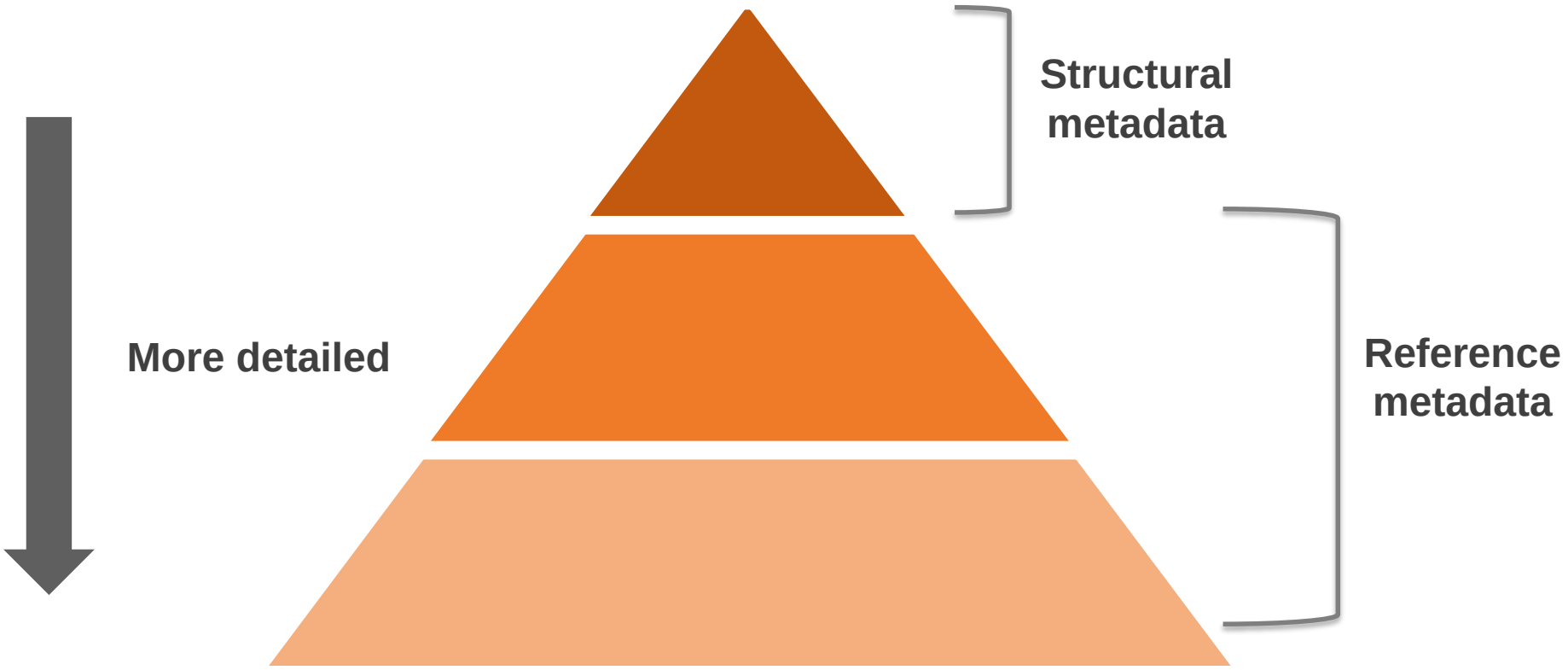
WHERE?

SUSTAINABLE DEVELOPMENT GOALS

Why does metadata make a difference?



Hierarchical way of viewing metadata



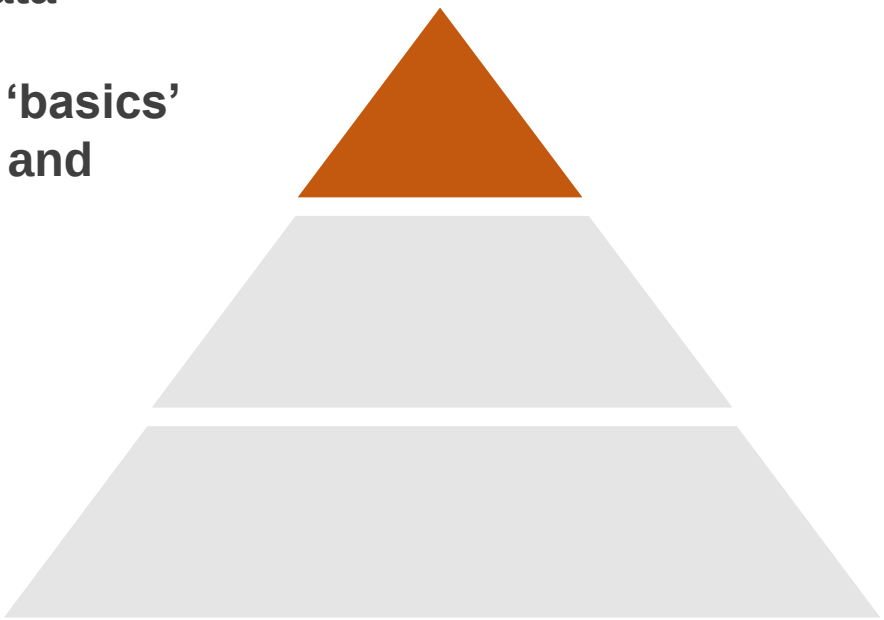
Information describing the data is more detailed as one moves down from the top of the pyramid.

Structural metadata

Metadata that act as identifiers and descriptors of data

At the top of the pyramid is information essential for understanding the data

Needs to explain the ‘basics’ of when, where, who and what?



In practical terms this means at the least information about: the indicator, the reference period, source, geographic scope, and the unit

Structural metadata

Metadata that act as identifiers and descriptors of data

Indicator 8.5.2
Unemployment rate, by sex, age and persons with disabilities[a,b]

Indicator

Unit

(a) Unemployment rate, both sexes[c]
(Percentage)

Geographic scope

Regions	*2000*	*2005*	*2010*
World	6.4	6.2	6.1
Sub-Saharan Africa	8.2	7.8	7.6
Northern Africa and Western Asia	11.8	11.7	10.2
Northern Africa	15.1	12.7	10.3
Western Asia	9.0	10.8	10.1
Central and Southern Asia	4.9	5.2	4.4
Central Asia	10.5	9.1	8.4
Southern Asia	4.6	5.0	4.3
Eastern and South-Eastern Asia	4.6	4.7	4.4
Eastern Asia	4.5	4.2	4.3
South-Eastern Asia	4.8	6.2	4.6
Latin America and the Caribbean	10.9	9.2	7.7
Oceania	6.2	4.9	5.4
Australia and New Zealand	6.3	4.8	5.4
Oceania (excluding Australia and New Zealand)	6.0	5.2	5.3
Europe and Northern America	7.9	7.5	9.2
Europe	9.7	8.5	9.0
Northern America	4.3	5.3	9.4
Landlocked developing countries	6.2	6.5	6.0
Least developed countries	5.1	5.6	5.5
Small island developing States	8.6	9.3	8.5

Reference period

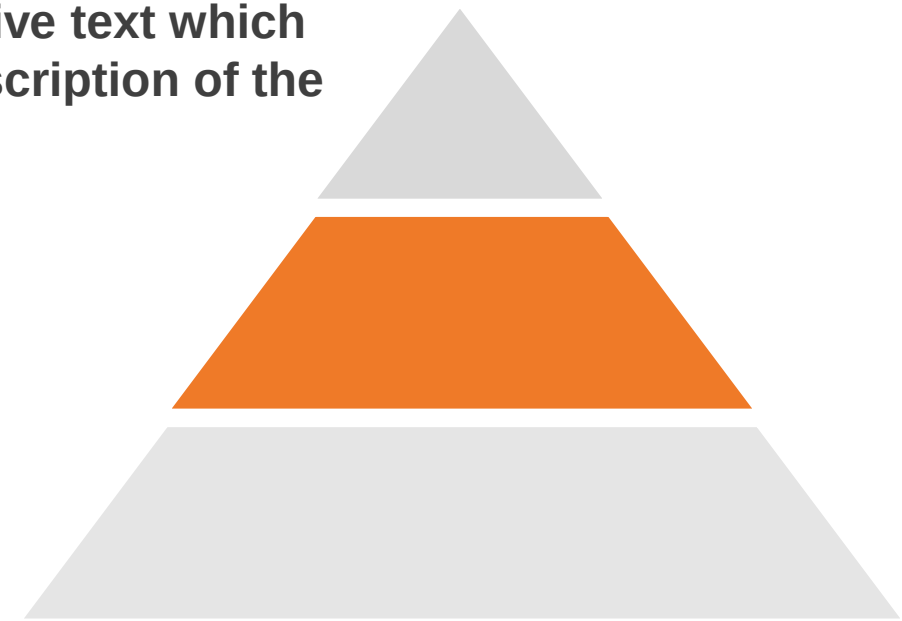
Source: *ILO Database of Labour Statistics (ILOSTAT), International Labour Organisation (ILO), ILO modelle 2016.*

Source

Conceptual reference metadata

Conceptual metadata that describes the key concepts used, and methodological elements

In the middle are explanatory notes and descriptive text which provide a good description of the statistics



In practical terms this means description about: definition, key concepts, standards and classifications used

Reference metadata

Reference metadata can include information such as key concepts and definitions

Indicator Name, Target and Goal

Indicator 1.1.1: Proportion of population below the international poverty line, by sex, age, employment status and geographical location (urban/rural)

Target 1.1: By 2030, eradicate extreme poverty for all people everywhere, currently measured as people living on less than $1.25 a day[1].

Goal 1: End poverty in all its forms everywhere

[1] The International Poverty Line was updated to $1.90 per day in October 2015.

Definition and Rationale

Definition:

Definitions

The indicator is defined as the proportion of the population living in households below the international poverty line where the average daily consumption (or income) per person is less than $1.9 a day measured at 2011 international prices adjusted for purchasing power parity (PPP) .

Concepts:

Key concepts

The international poverty line is a threshold used to measure extreme poverty based on consumption or income levels. A person is considered extremely poor if his or her consumption or income level falls below the minimum level necessary to meet basic needs. For this indicator, the line is set at $1.90 (2011 PPP). It replaces the $1.25 a day poverty line measured in 2005 prices since October 2015.

Standards and classification

The purchasing power parity (PPP) conversion factor is the number of units of a country's currency required to buy the same amounts of goods and services in the domestic market as one United States dollar would buy in the United States. It is based on the System of National Accounts' concept of actual individual consumption.

Detailed reference metadata on methods and quality

Further down the hierarchy of information, **we have more detailed information**

These are potentially the source of the most detailed methodological information available.

Some statistical agencies publish very detailed concepts, sources and methods for a number of their key statistics.

In practical terms this means description about: process of collecting data, calculations, quality aspects, limitations of data and other detailed information.

More detailed reference metadata

How is data collected?

Data Sources and Collection Method

The data for this indicator is collected through household-based labour force surveys, population census, and any other nationally representative household surveys with an appropriate employment module. Such surveys are generally conducted by the ministries or bureaus of labour or national statistical offices.

Unemployment registers, under social insurance administrative systems, can also serve as instruments to collect data on unemployment levels, and used to supplement the information obtained by household surveys.

Method of Computation and Other Methodological Considerations

Computation Method:

The unemployment rate *(U)* is calculated using the following formula:

How is the indicator calculated?

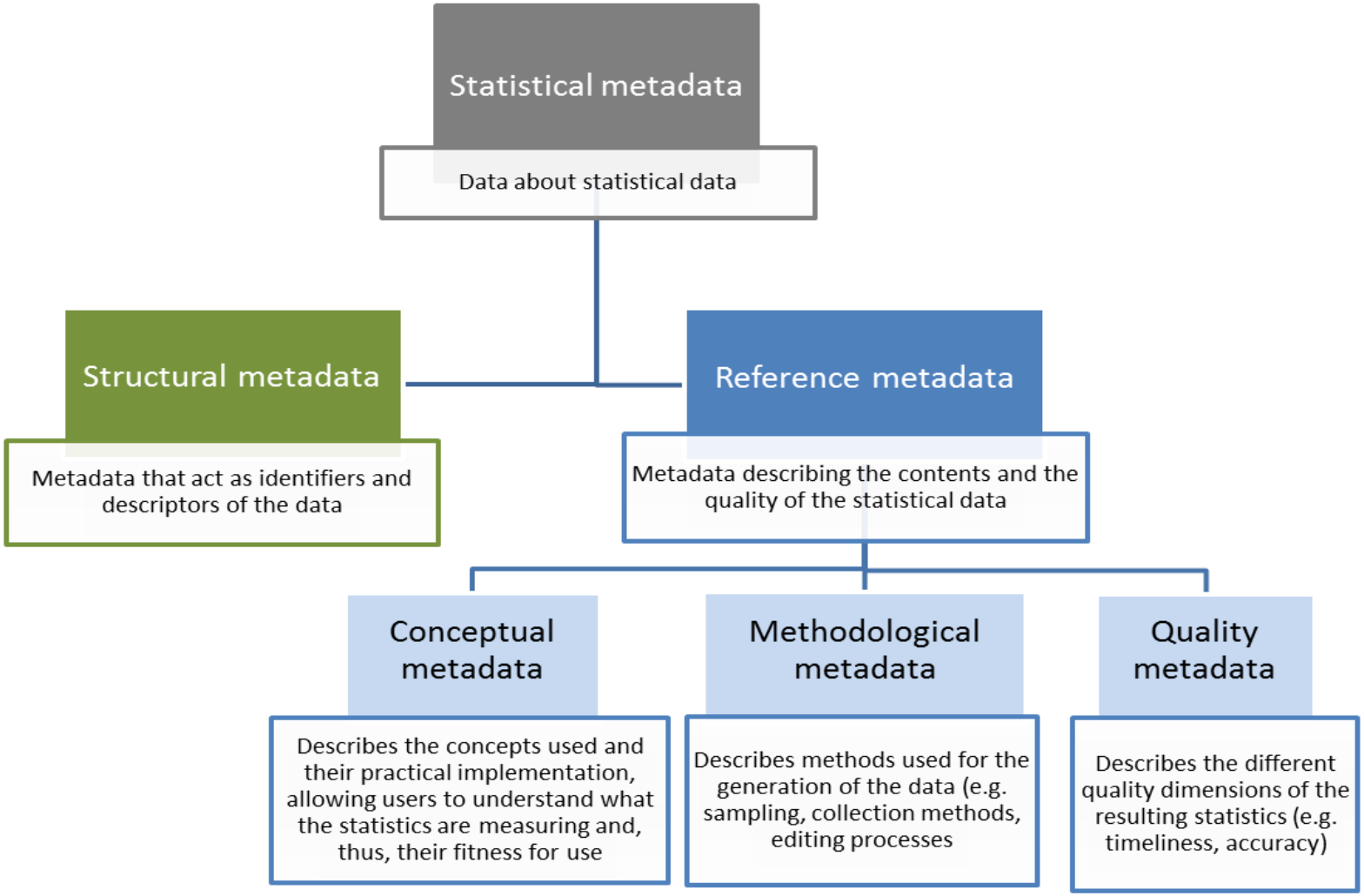
$$U = \frac{Number\ of\ unemployed\ persons}{Persons\ employed + Persons\ unemployed} \times 100$$

What are the limitations?

Comments and limitations:

The significance of the unemployment rate depends on context. It is not to be interpreted the same way universally. In the absence of unemployment insurance systems or social safety nets, persons of working age must avoid unemployment, resorting to engaging in some form of economic activity, however insignificant or inadequate. Thus, in this context, other measures should supplement the unemployment rate to comprehensively assess labour underutilization, such as the time-related underemployment rate or measures of the potential labour force In this regard, the 2013 *Resolution*

To summarize:



International metadata standards

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. SDMX is maintained by a group of seven sponsors: the Bank for International Settlements (BIS), the European Central Bank (ECB), Eurostat, the International Monetary Fund (IMF), the Organization for Economic Cooperation and Development (OECD), the United Nations and the World Bank.
Data Documentation Initiative (DDI)	The DDI is a standard for technical documentation describing social science data. The current version (3.1) supports description of the full life cycle of a dataset or data collection.
Metadata Common Vocabulary (MCV)	The MCV contains concepts and related definitions that are normally used by international organizations and national data producing agencies to describe statistical metadata. Terms such as census, estimate, footnote, measurement error, occupation, periodicity, quality and sample are all defined in the MCV. The MCV is a valuable resource for establishing common terminology in the presentation of MDG data and metadata.

Why is metadata important?

Metadata improves usability of data

Without metadata, data is not meaningful

Table: Indicator title

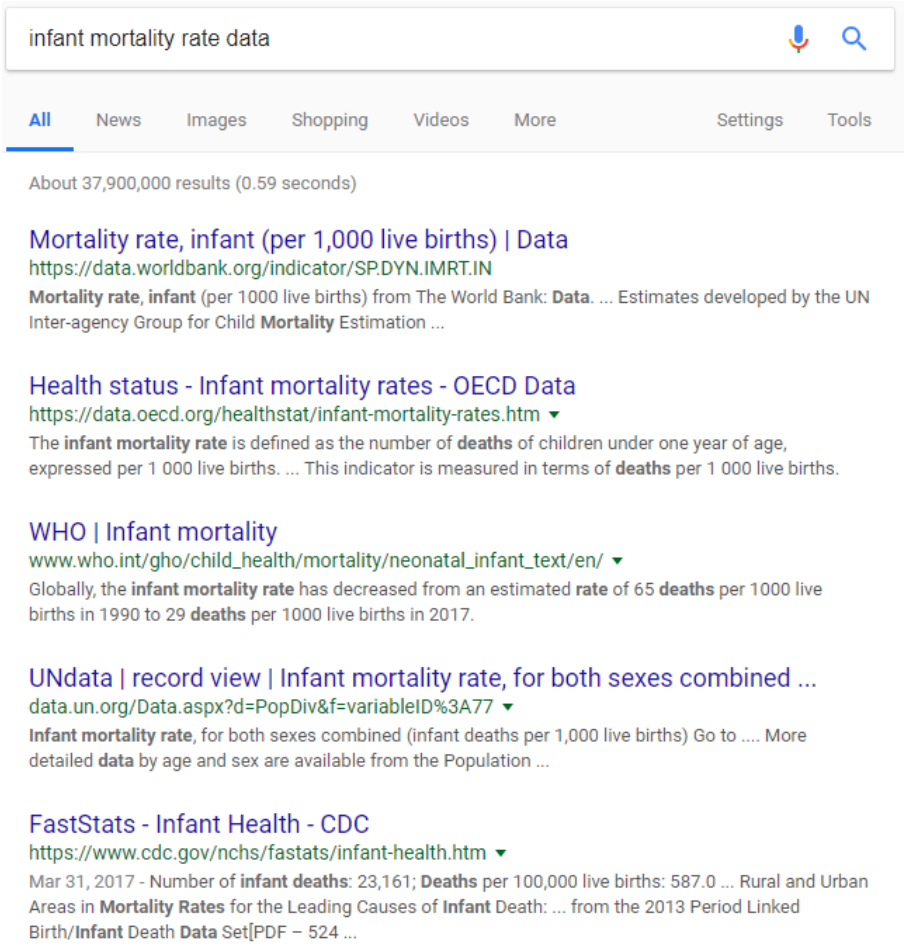
Code	Variable	Unit 1	Unit 2
	Variable label 1	168,388	6
	…	11,446	0
	…	121,786	4
	...	19,486	1
	...	7,210	0
	…	1,149,906	42
	…	1,260,671	46
	…	**2,738,893**	**100**

Source: name

Metadata improves consistency and coherence

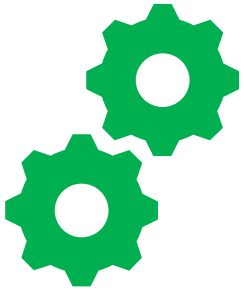
This is essential for interoperability and seamless ata exchange

- Via common format, terminology and dissemination standards, metadata **improves consistency in presenting data.**
- Users these days typically use search engines to locate information
- The efficiency of finding what one wants is increased if statistics are presented in a **common format using standardized terminology**



Metadata improves comparability

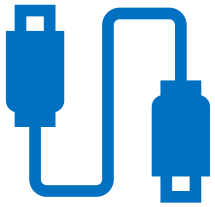
- Greater economic & social integration between countries, has increased user **demand for improved comparability**



- Need for comparability is also felt **within a same country** for example, are time series comparable over time.

- **Differences** arise from:
  - ✓ use of different definitions, concepts, units and classifications;
  - ✓ varying collection and processing (transformation) practices.
  - ✓ diverse reporting & presentation practices

- As policy makers and data users demand more data in the SDG-era, and as more data is shared using technologies such as data platforms, **assessment of comparability is essential.**



Highlights data quality issues

- **Common features** of metadata include;
  - definition of “quality” concept, and a list of dimensions that define quality.

- These dimensions normally include **relevance, accuracy, timeliness, interpretability, coherence**, etc.

- Help **understand limitations** of a specific series or data point, and inform data users of the extent of analyses that can be conducted, or the applicability of the findings of the analyses

Increased relevance in the SDG-era

- Scope of the 2030 Agenda is expansive: covers social, environmental and economic issues across time.

- Disaggregated data: leaving no one behind

- Interest from a broad user-base, not all have ability or experience to interpret data correctly without guidance.

- Imperative to standardize and present contextual information to users

Some issues relevant to metadata

Metadata access and content

- Users of metadata are both **data producers and as well as data users**. Need to keep in mind while preparing different types metadata

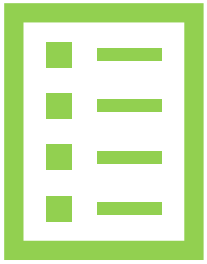
- One of the most cost-effective solutions is to **disseminate metadata on the web**.

- **Practices vary considerably** by country
  - Amount of detail provided on websites;
  - Availability in national language;
  - frequency of updating & cost;
  - its proximity to the statistics it describes

  - Significant differences between countries in **statistical methodological elements** described in metadata for the same statistical domain.

Metadata structure

- Key thing to consider is how to organize and structure the metadata to allow users to go as deeply as necessary, without being buried in enormous amounts of text



- Understanding that metadata is not a compilation guidance. It describes data, but it is not supposed to replace technical manuals and guidance materials.

- A guiding principle for publishing data is that tables, charts and maps should contain sufficient metadata so that they can “stand alone”. Focus on publishing it in various forms.

International recommendations for metadata

- Ensure metadata dissemination via a range of different media, especially the web.
- Metadata structured to meet the needs of a range of users with different requirements and/or statistical expertise.
- A layered presentation of metadata is recommended, progress from summary to more detail. Each layer with clear and precise text.
- Active linkage of metadata to the statistical tables and graphs they describe and vice versa, tailored to each statistical domain.
- Rigorous use of terminology found in international and national glossaries & guidelines.

International recommendations for metadata

- Need for senior management to ensure appropriate practices and principles.
- Provide appropriate cross references or links to these sources to make existing standard terms and definitions more readily available.
- Avoid ambiguity like using the same label or title for different definitions.
- Disseminate metadata free of charge on the Internet, as a high public good component (even if statistics have price regime).
- Availability of metadata in the national language, and an international, such as English.
- Provide contact details for further information on concepts, definitions and statistical methodologies.

International recommendations for metadata

- **Make metadata active to the greatest extent possible. Treating metadata this way will ensure they are accurate and up-to-date.**
- **Reuse metadata where possible for statistical integration as well as efficiency reasons.**
- **Preserve history of metadata.**
- **Ensure process (workflow) is well documented so there is clear identification of ownership, approval status, date of operation, etc.**
- **Ensure that variations from standards are tightly managed/approved, documented and visible.**

Examples of metadata

Metadata for global SDG data

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

Last updated: 18 January 2018

Goal 3: Ensure healthy lives and promote well-being for all at all ages
Target 3.1: By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births
Indicator 3.1.1: Maternal mortality ratio

Institutional information

Organization(s):

World Health Organization (WHO)

Concepts and definitions

Definition:

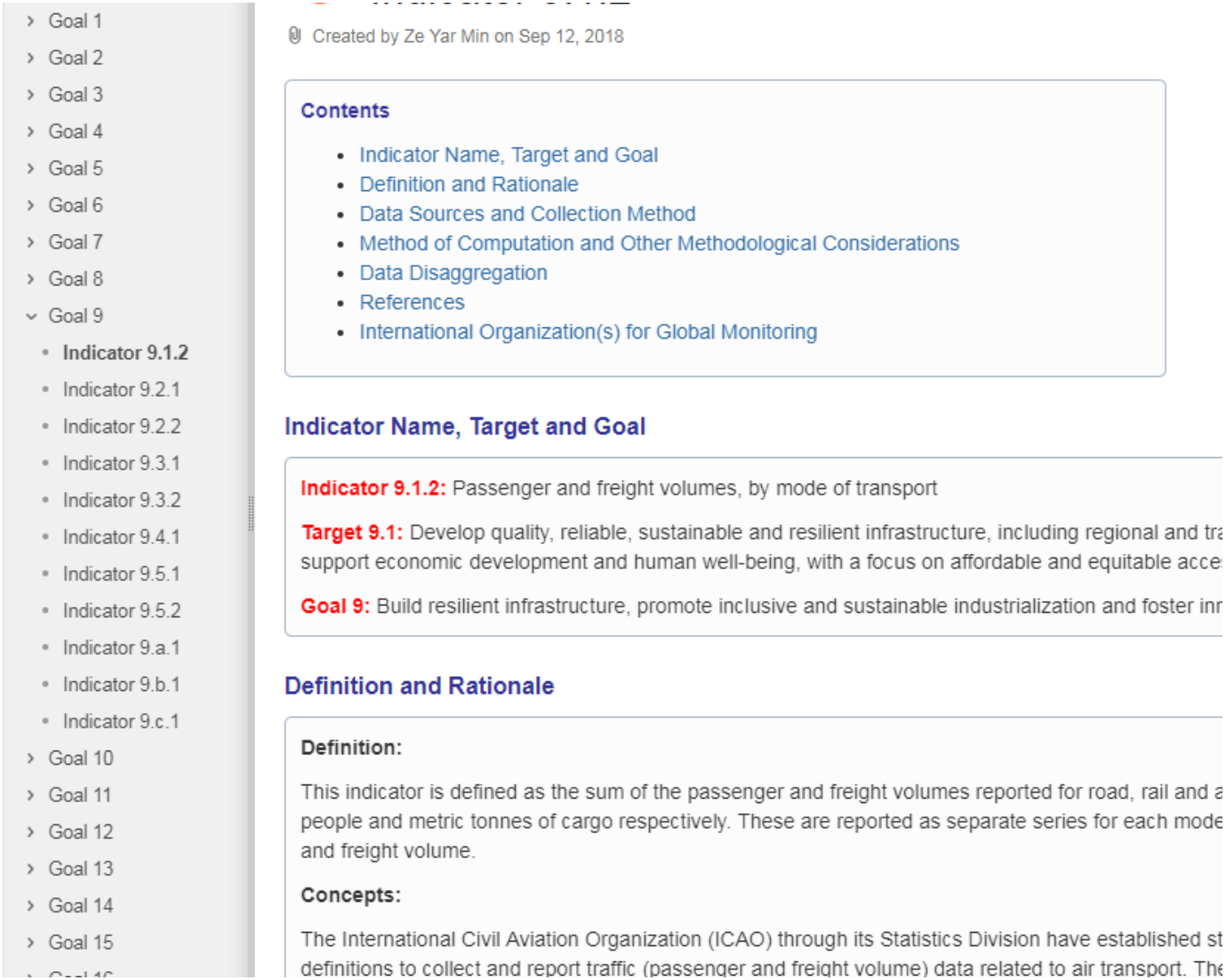
The maternal mortality ratio (MMR) is defined as the number of maternal deaths during a given time period per 100,000 live births during the same time period. It depicts the risk of maternal death relative to the number of live births and essentially captures the risk of death in a single pregnancy or a single live birth.
Maternal deaths: The annual number of female 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, expressed per 100,000 live births, for a specified time period.

Rationale:

All maternal mortality indicators derived from the 2015 estimation round include a point-estimate and an 80% uncertainty interval (UI). For those indicators where only point-estimates are reported in the text or tables, UIs can be obtained from supplementary material online (http://www.who.int/reproductivehealth/publications/monitoring/maternal-mortality-2015/en/). Both point-estimates and 80% UIs should be taken into account when assessing estimates.

E-handbook on SDG indicators

https://unstats.un.org/wiki/display/SDGeHandbook/Home



SUSTAINABLE DEVELOPMENT GOALS

Eurostat SDG metadata

eurostat

Final energy consumption (sdg_07_11)

ESMS Indicator Profile (ESMS-IP)
Compiling agency: Eurostat, the statistical office of the European Union

Eurostat metadata

Reference metadata

1. Contact
2. Metadata update
3. Relevance
4. Statistical Indicator
5. Frequency and Timeliness of dissemination
6. Coverage and comparability
7. Accessibility and clarity
8. Comment
Related Metadata
Annexes (including footnotes)

Eurostat Quality Profile

4.5. Source data	ESS
5.1. Frequency of dissemination	Every year
5.2. Timeliness	T+2 years
6.1. Reference area	All EU MS
6.2. Comparability - geographical	All EU MS
6.3. Coverage - Time	> 10 years
6.4. Comparability - over time	> 4 data points

Description of Eurostat quality grading system under the following link.

For any question on data and metadata, please contact: EUROPEAN STATISTICAL DATA SUPPORT

Download

1. Contact

Top

1.1. Contact organisation	Eurostat, the statistical office of the European Union
1.2. Contact organisation unit	E2: Environmental statistics and accounts; sustainable development
1.5. Contact mail address	e-mail contact: ESTAT-SDI-EU2020-INDICATORS@ec.europa.eu

2. Metadata update

Top

2.1. Metadata last certified	21/11/2017
2.2. Metadata last posted	05/09/2018
2.3. Metadata last update	14/08/2018

3. Relevance

Top

This indicator is part of the indicator sets:
a) EU Sustainable Development Goals (SDG) indicator set where it is used to monitor progress towards SDG 7 on affordable and clean energy, SDG 12 on ensuring sustainable consumption and production patterns and SDG 13 on climate action. SDG 7 calls for ensuring universal access to modern energy services, improving energy efficiency and increasing the share of renewable energy. To accelerate the transition to an affordable, reliable, and sustainable energy system, countries need to facilitate access to clean energy research, promote investment in energy infrastructure and clean energy technology. SDG 12 envisions sustainable consumption and production, which uses resources efficiently, reduces global food and other waste, disposes safely toxic waste and pollutants. SDG 13 aims to implement the commitment to the United Nations Framework Convention on Climate Change and operationalise the Green Climate Fund.
b) EU 2020 strategy indicators where it is used to monitor progress towards the EU's objective for climate and energy policy 'Moving towards a 20 % increase in energy efficiency' - to be reached by 2020.
Furthermore, indicator is part of the impact indicators for Strategic plan 2016-2020 referring to the 10 Commission priorities; and is included in the set for the EU's framework Strategy for a Resilient Energy Union.
The Europe 2020 strategy set a target of moving towards a 20 % increase in energy efficiency by 2020. By 2030, the share should further increase to at least a 27 % improvement in energy efficiency (increased to a target of at least 30 % improvement in the Commission's proposal for a recast energy efficiency directive) according to the 2030 Climate and Energy Policy Framework. The Energy Union strategy highlights that improved energy efficiency will reduce our dependence on energy imports, reduce emissions and drive jobs and growth.

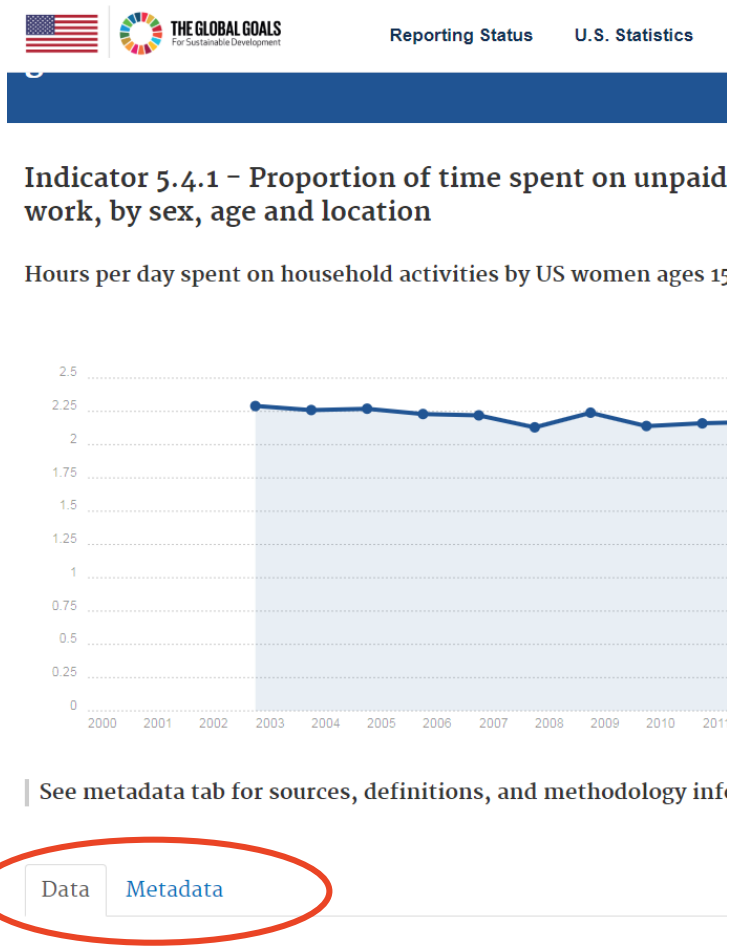
4. Statistical Indicator

Top

4.1. Data description

The indicator measures the total energy demand of a country excluding all non-energy use of energy carriers (e.g. natural gas used not for combustion but for producing chemicals).
"Final energy consumption" only covers the energy consumed by end users, such as industry, transport, households, services and agriculture, it excludes energy consumption of the energy sector itself and losses occurring during transformation and distribution of energy.

United States SDG metadata



U.S. Metadata

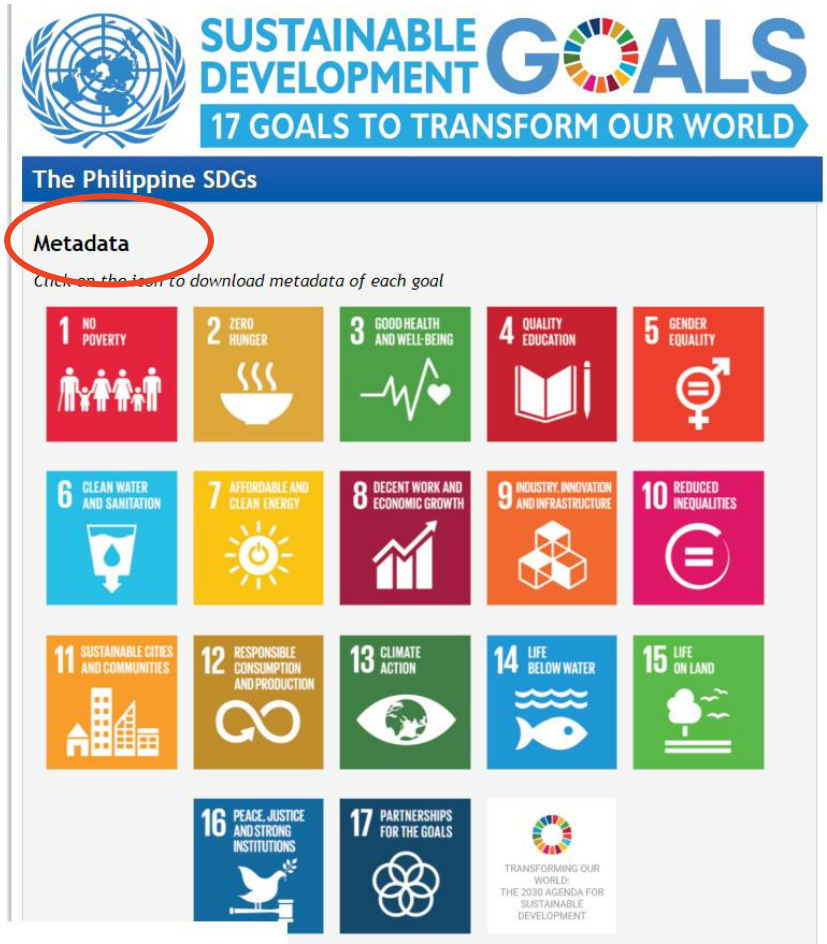
This table provides metadata for the actual indicator available from U.S. statistics closest to the corresponding global SDG indicator. Please note that even when the global SDG indicator is fully available from U.S. statistics, this table should be consulted for information on national methodology and other U.S.-specific metadata information.

Method of computation for global SDG indicator	
Graph Title	Hours per day spent on household activities by US women ages 15 and older
Actual indicator available	Time spent per day on household activities (includes travel), Caring for and helping household members (includes travel), Caring for and helping nonhousehold members (includes travel), and Purchasing goods and services (includes travel) by sex and age
Description of actual indicator available	Average hours per day: The average number of hours spent in a 24-hour period doing a specified activity.
Method of computation	Data are from the American Time Use Survey (ATUS), which is nationally representative of the U.S. civilian noninstitutional population age 15 and over. Individuals are selected from households that have completed the 8th month of the Current Population Survey. Each selected individual is interviewed one time, by telephone, about how they spent their time on one day. Individuals have been interviewed for the ATUS on nearly every day since the survey began in 2003. For information about ATUS methods, see the BLS Handbook of Methods: https://www.bls.gov/opub/hom/atus/home.htm. Data are estimates of average hours per day (see formula in section 7.4 of the ATUS User's Guide: https://www.bls.gov/tus/atususersguide.pdf).
Comments and limitations	Age categories: ages 15 and over, 15-24 years, 25-34 years, 25-54 years, 35-44 years, 45-54 years, 55-64 years, and 65 years and over
Periodicity	Annual
Time Period	2003-present
Unit of measure	Average hours per day

Mexico SDG metadata

Objetivo	2. Poner fin al hambre, lograr la seguridad alimentaria y la mejora de la nutrición y promover la agricultura sostenible
Meta	Meta 2.1. De aquí a 2030, poner fin al hambre y asegurar el acceso de todas las personas, en particular los pobres y las personas en situaciones de vulnerabilidad, incluidos los niños menores de 1 año, a una alimentación sana, nutritiva y suficiente durante todo el año
Nombre del indicador	2.1.2. Proporción de la población con inseguridad alimentaria moderada o severa (carencia por acceso a la alimentación)
Definición	La inseguridad alimentaria moderada o severa son los grados de inseguridad alimentaria que reflejan la existencia de limitaciones en la cantidad de alimentos en el hogar y experiencias de hambre en la alimentación de los integrantes de los hogares. La clasificación de la situación de seguridad o inseguridad alimentaria se hace con base en la Escala Mexicana de Seguridad Alimentaria (EMSA).
Tipo de indicador	Global
Algoritmo	$PPIA_t = \left(\frac{PCA_t}{PT_t}\right) 100$ Significado de las siglas o abreviaturas: $PPIA_t$ Proporción de la población con un grado de inseguridad alimentaria moderada o severa en el año t PCA_t Población con carencia por acceso a la alimentación (suma de la población con grado de inseguridad alimentaria moderada o severa) en el año t PT_t Población total en el año t
Descripción narrativa del cálculo del indicador	Se determina a partir de una distinción entre los hogares con población menor de dieciocho años y los hogares sin población menor de dieciocho años. Para el total de hogares donde no habitan menores de dieciocho años de edad se identifican aquellos en los que algún adulto, por falta de dinero o recursos, al menos: no tuvo una alimentación variada; dejó de desayunar, comer o cenar; comió menos de lo que debería comer; se quedaron sin comida; sintió hambre, pero no comió; o hizo sólo una comida o dejó de comer durante todo el día. En el caso del total de hogares donde habitan menores de dieciocho años de edad se identifican aquellos en los que por falta de dinero o recursos, tanto las personas mayores como las menores de dieciocho años: tuvieron una alimentación variada; comieron menos de lo necesario; se les disminuyeron las cantidades servidas en la comida; sintieron hambre, pero no comieron; o hicieron una comida o dejaron de comer durante todo el día.
Unidad de medida	Porcentaje
Cobertura geográfica	Nacional, entidad federativa y municipio.
Referencia temporal	2010 - 2014
Oportunidad	1 año.
Periodicidad del indicador	Nacional y entidad federativa: Bienal. Municipios: Quinquenal
Fuente generadora de información estadística utilizada para el cálculo del indicador	Instituto Nacional de Estadística y Geografía (INEGI). Módulo de Condiciones Socioeconómicas de la Encuesta de Ingresos y Gastos de los Hogares (MCS-ENIGH). 2010 - 2014. http://www.inegi.org.mx/est/contenidos/proyectos/encuestas/hogares/modulos/mcs/default.aspx Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL). Metodología para la medición multidimensional de la pobreza en México
Fecha de actualización del indicador	11 de abril de 2017

Philippines SDG Watch



Goal 1. End poverty in all its forms everywhere

Target 1.a: Ensure significant mobilization of resources from a variety of sources, including through enhanced development cooperation, in order to provide adequate and predictable means for developing countries, in particular least developed countries, to implement programmes and policies to end poverty in all its dimensions

1.a.2 Proportion of total government spending on essential services (education, health and social protection)

Indicator Classification	
Global SDG Indicator	Yes
Proxy	No
Supplemental	No
Definition	
Global Definition	
National Definition	
Proportion of total government spending on essential services (education, health and social protection)	
Method of Computation	
Global	
National	
Total government spending on essential services (education, health and social protection) divided by total spending/expenditure	
Matching (only for the Global SDG indicator)	
Exact Matching	
Good Matching	
Partial Match	
D-match	
E-match	
Level of Disaggregation (Please specify whether Yes- available, No-Not available, NA- Not applicable)	
National	Yes
Regional	Yes
Provincial	No
by Sex	No
Other Disaggregations, please specify	
Frequency of Release	
Monthly	No
Quarterly	No

... | 1.4.1p7 | 1.4.1p8 | 1.4.1p9 | 1.4.2p1 | 1.5.1 | 1.5.3 | 1.5.4 | **1.a.2** | 1

Thank you for your attention!