

countryData

CountryData Workshop
Technologies for Data Exchange

Reference Metadata



United Nations
Statistics Division

DFID

Department for
International
Development

 Burundi

 Cambodia

 Ghana

 Liberia

 Laos

 Morocco

 State of Palestine

 Rwanda

 Thailand

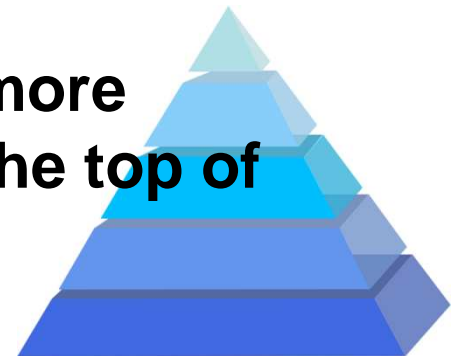
 Uganda

 Vietnam

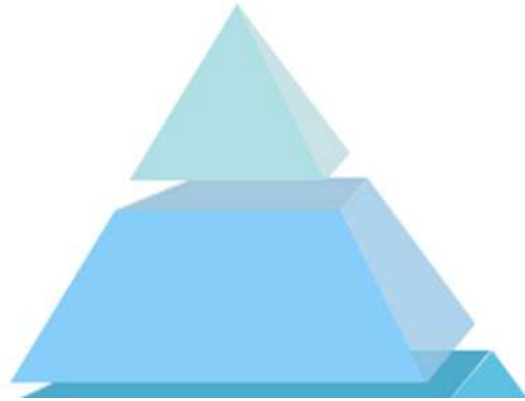
Metadata

- **According to the ISO;**

“Metadata is data that defines and describes other data.”
- **Perhaps, better visualized as layers within a pyramid.**
- **Information describing the data is more detailed as one moves down from the top of the pyramid.**



“Structural” metadata



1. **At the top of the pyramid is information essential for understanding the data**
 - **Needs to explain the ‘basics’ of when, where, who and what?**





Reference metadata

2. In the middle are explanatory notes and text generally located in the same “publication” provides a good description of the statistics



- **definitions, key issues, limitations, etc., that can impact on the use of the data.**

Methodological metadata

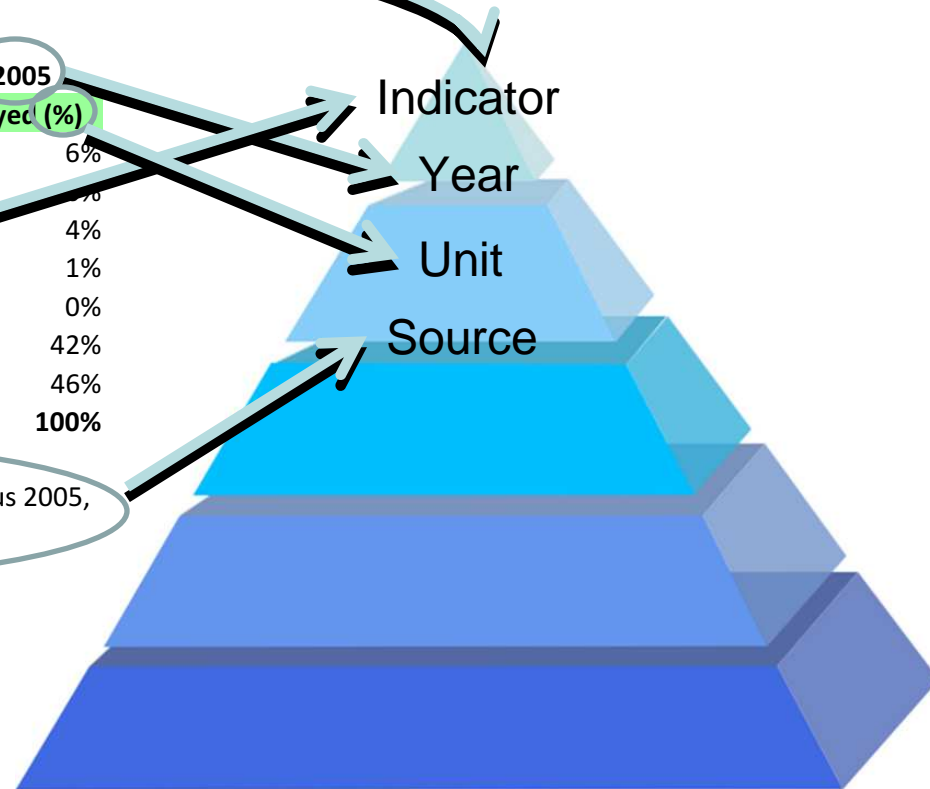
- 
- 3. Detailed methodological information disseminated in publications / websites.**
- **These are potentially the source of the most detailed methodological information available.**
 - **Some statistical agencies publish very detailed concepts, sources and methods for a number of their key statistics.**
- 

Structural metadata: Statistical table

Table: Number of Employed by type of employment, 2005

Code	Employed	Employed (%)
1 Government employee	168,388	6%
2 Parastatal employee	11,446	0%
3 Gov_Private employee	121,786	4%
4 Government employee	15,486	1%
5 Employer	7,210	0%
6 Own account Worker	1,149,906	42%
7 Unpaid family Worker	1,260,671	46%
Total	2,738,893	100%

Source: Results from the Population and Housing Census 2005, NSC, CPI



What's missing?

Reference metadata: Statistical report

Population Census 2005 - Economic activity and labour force

Page 5 of 9

Article Index

- Population Census 2005
- Chapter 2: Population distribution & Migration
- Chapter 3: Household characteristics
- Chapter 4: Education and literacy
- Chapter 5: Economic activity and labour force
- Chapter 6: Fertility
- Chapter 7: Mortality
- Chapter 8: Housing characteristics
- Chapter 9: Population projections
- All Pages

Chapter 5: Economic activity and labour force

The 2005 census measures, like the 1995 census, the activity status of the Lao population by the approach of "the usually active population", which means the main activity during the past 12 months. This approach was chosen instead of the "currently active population" usually referred to the status for the past week, because it represents a more stable measure of the economically active population and its structural distribution for economic analysis, projections and development planning. The activity status has been measured for the population 10 years and above. It thus includes children who have still not finished school because many children normally participate in agriculture or other types of economic activity.

The first question asked was:

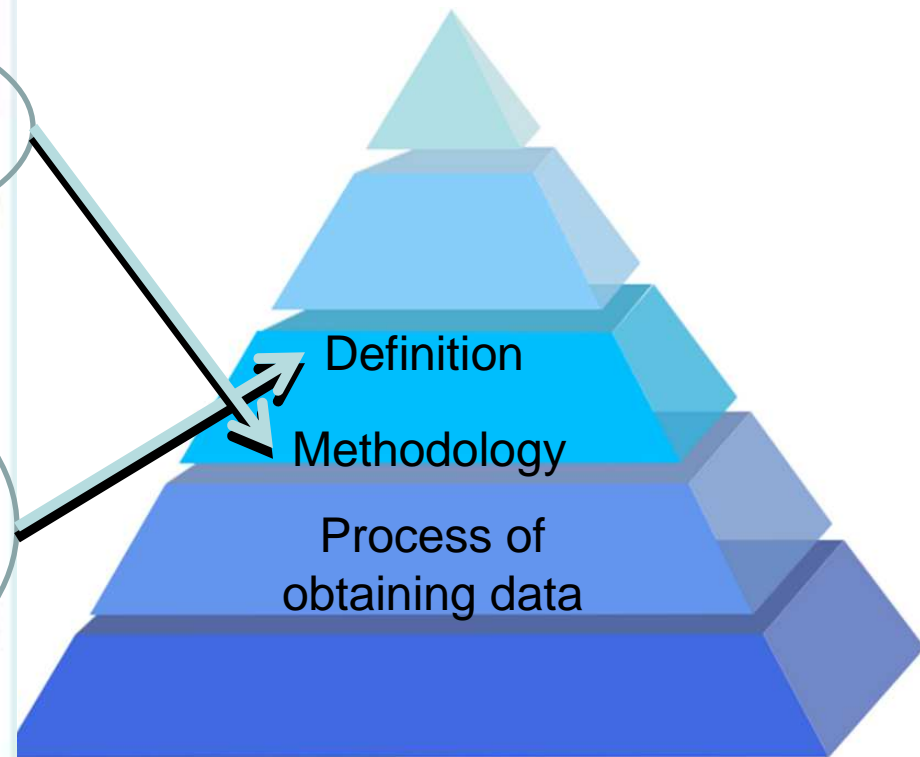
What was your main activity the last twelve months?

Response alternatives:

- 01 Paid Employee/Government (01)
- 02 Paid Employee/Parastatal (02)
- 03 Paid Employee/Private (03)
- 04 Paid Employee/State Enterprise (04)
- 05 Employer (05)
- 06 Own Account Worker (06)
- 07 Unpaid Family Worker (07)
- 08 Unemployed (08)
- 09 Student (09)
- 10 Household Duties (10)
- 11 Retired/Sick/Too Old (11)
- 12 Other (12)

A person with response alternatives 01-08 is considered to be economically active, alternative 08 with the assumption that the person actually is looking for a job. Students (09) and persons whose main activity is household duties (10), retired/sickness/too old (11) are together with the other response alternatives considered to fall in the category "economically not active population".

The usually active population, or the labor force, comprises all persons who are 10 years and above and who are either employed or unemployed specified as 01 to 08 in the above list.



Example: Methodology report

Census of Population and Housing, 1 March 2005 [LAO-DOS-CPH-2005-v1.1]

Overview

Technical Information

- Sampling
- Questionnaires
- Data Collection
- Data Processing
- Data Appraisal

Data Sets

- Data Files

Access Policy

Variable Search

Other Materials

Overview

Identification

Country
Lao People's Democratic Republic

Title
Census of Population and Housing, 1 March 2005

Survey Type
Population and Housing Census (hhpopcen)

Series Information
The Census of Population and Housing (CPH) was conducted for the first time in 1985. The second time was in 1995, and supported by the Swedish International Development Cooperation Agency (SIDA).

ID Number
LAO-DOS-CPH-2005-v1.1

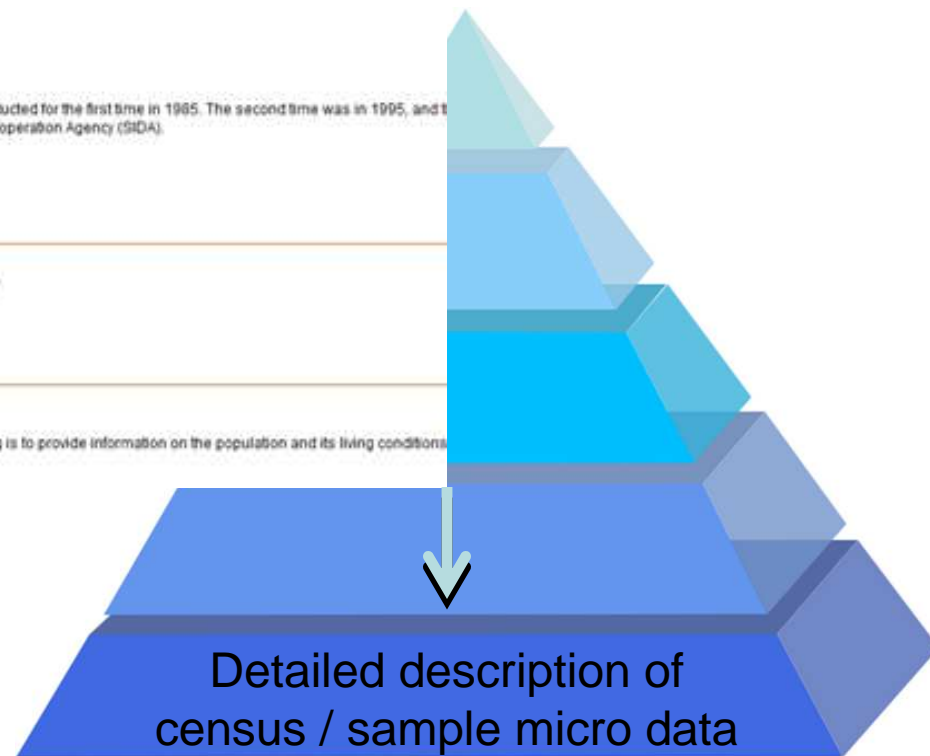
Version

Version Description
Version 1.1 - Edited data, first version, for internal use only

Production Date
2007

Overview

Abstract
The purpose of the 2005 Population Census and Housing is to provide information on the population and its living conditions, construction, water and electricity supply, etc.



Indicator

- 1. A statistical measure of a quantitative characteristic of an individual or collective phenomenon. (Adapted definition)**
- 2. It is the first building block of any statistical series, i.e. what to measure?**
- 3. Highest level, in the metadata hierarchy (i.e. composed of time series).**
- 4. Usual level at which reference metadata are attached**

Time series

- 1. A set of ordered observations on a quantitative characteristic of an individual or collective phenomenon taken at different points of time. (MCV 2009)**
- 2. Middle level, in the metadata attachment hierarchy (i.e. composed of observation values).**

Observation value

- 1. The value of a particular variable at a particular period. (MCV 2009)**
- 2. The observation value is the field which holds the data.**
- 3. Lowest level, in the metadata attachment hierarchy.**

Metadata attachment hierarchy

Indicator	Time series	Observation value
Literacy rate	+ { <ul style="list-style-type: none"> Percent, Male, 15 – 24 yr old Percent, Female, 15 – 24 yr old Percent, Both sexes, 15 – 24 yr old 	+ { <ul style="list-style-type: none"> 1990, 1991, 1992, 1993,...
Population using improved sanitation facilities	+ { <ul style="list-style-type: none"> Percent, Both sexes, All ages, Urban Percent, Both sexes, All ages, Rural Percent, Both sexes, All ages, Total (National) 	+ { <ul style="list-style-type: none"> 2000, 2005, 2010,...
Malaria deaths	+ { <ul style="list-style-type: none"> per 100,000 population , Both sexes, Under 5yrs per 100,000 population, Both sexes, All ages 	+ { <ul style="list-style-type: none"> 1994, 1995, 1996, 1997,...

MDG/CountryData Metadata Structure Definition (MSD)

- Supports a “standard” set (based on MDG international database) of metadata to be exchanged to CountryData

ID	Description
DATA_PROVIDER	Country Name
SERIES	Indicator Title
STAT_CONC_DEF	Definition of the MDG official indicator or background series provided
METHOD_COMP	Method of computation
COMMENTS_LIMITATIONS	Comments and limitations
DISCREPANCIES	Sources of discrepancies between global and national figures
COLL_METHOD	Process of obtaining data
REL_CAL_POLICY	Expected time of release

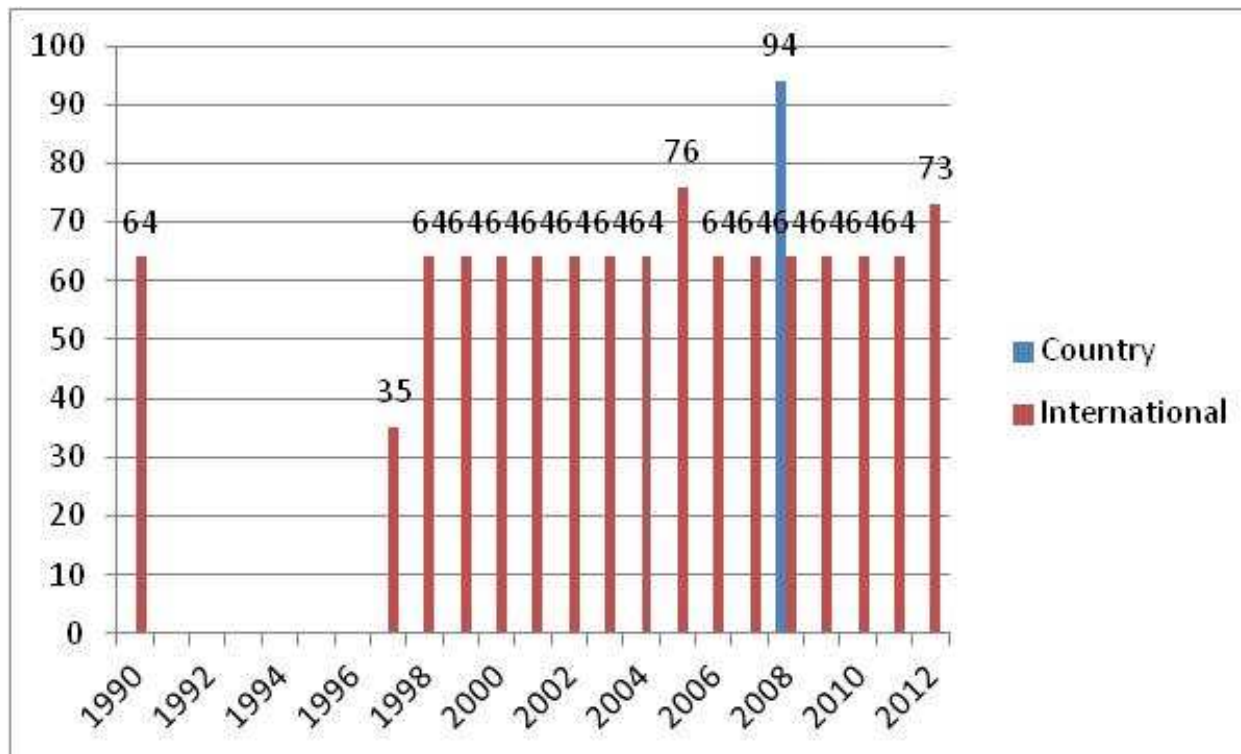
Definition

Definition

- **Provide descriptive information on the definition of the concepts associated with;**
 - **the indicator (i.e. births, disease, etc.)**
 - **any classifications used (i.e. industry, financial, environmental, rural/urban, occupations, age groups etc.)**
- **Often includes specific examples of what is and is not included in particular categories.**

Need to explain...

Total number of seats in national parliament



Source:

Country, NEC National Election Commission 2008

International, Country Data

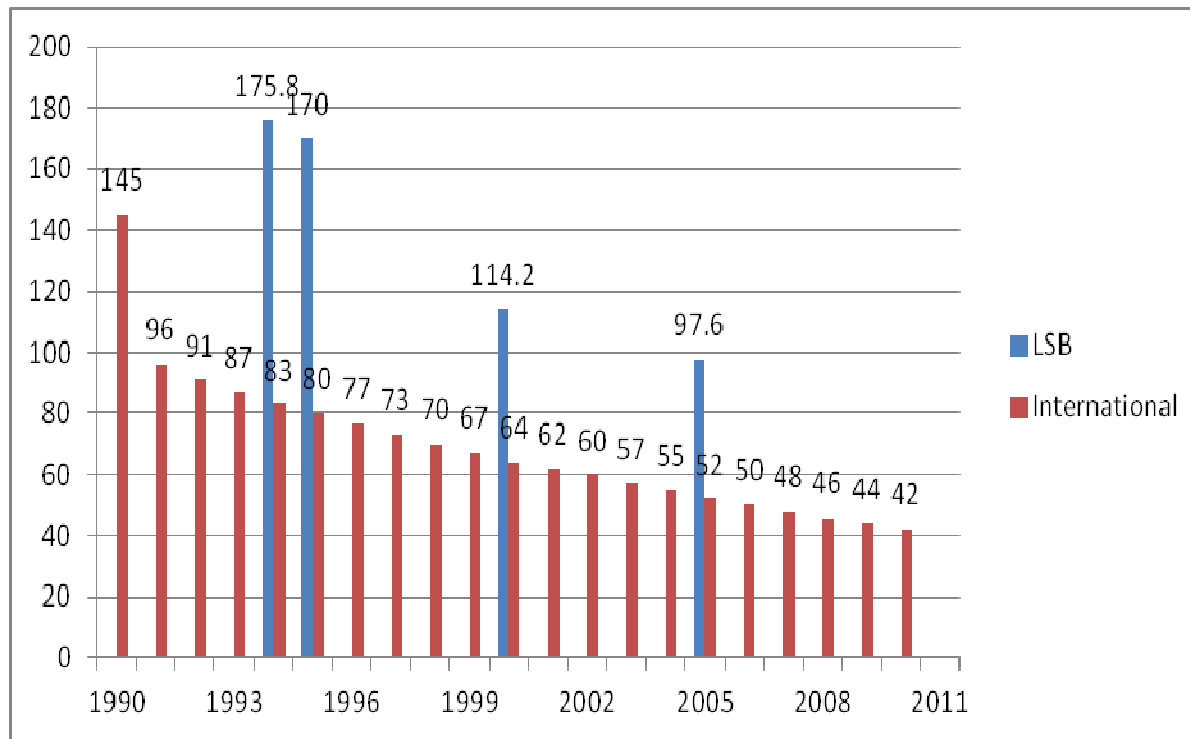
Methodology

Methodology

- 1. Provide descriptive information on any calculations made with the source data to produce the indicator**
- 2. Including formulas, adjustments and weighting particularly where mixed sources are used or where the calculation has changed over the time (i.e. discontinuities in the series).**
- 3. References to documentation related to various aspects of the data, such as detailed methodological documents or papers covering concepts, scope, classifications and statistical techniques.**

Need to explain...

Child (Under-5 years old) mortality per 1,000 live births



Source:

LSB, NSC FBSS 1994 /CPC Census 1995 /SPC RHS 2000 /CPI Census 2005.
International, Estimated.

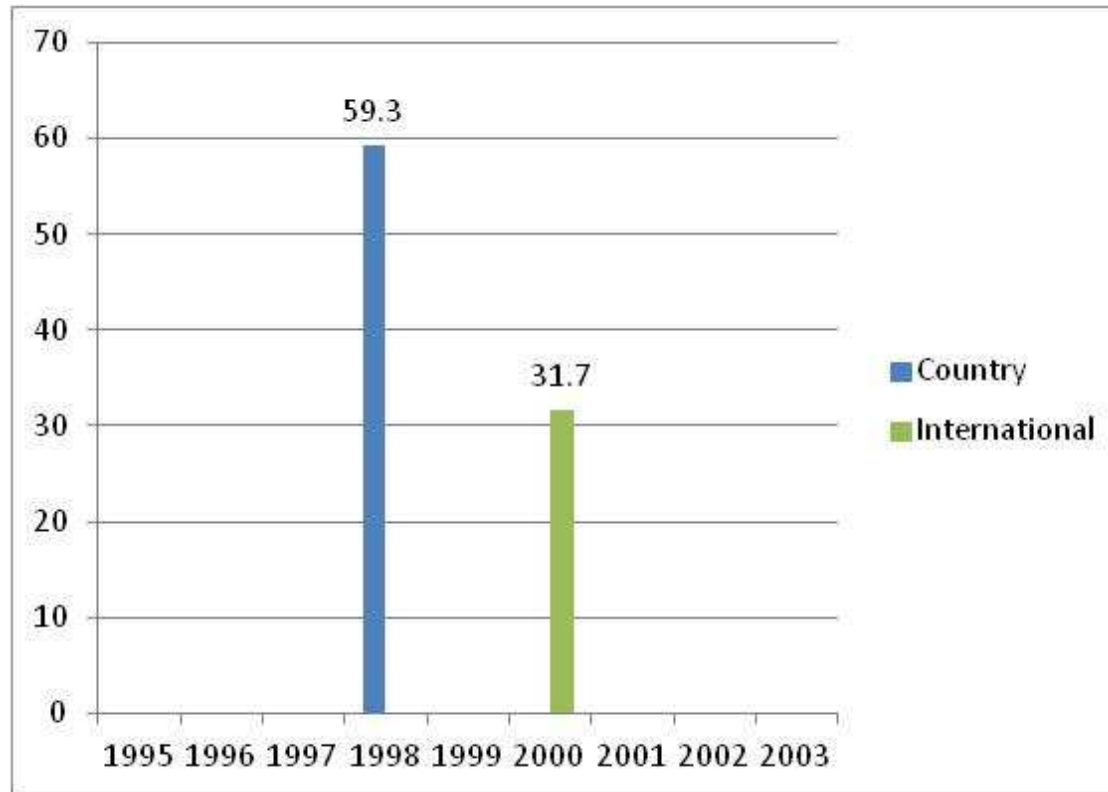
Process of obtaining data

Process of obtaining data

- **Provide descriptive information on the source.**
- **For example with a census/ survey source the following should be described;**
 - **the sample frame used & coverage**
 - **type of interview conducted**
 - **dates/ duration of fieldwork**
 - **sample size & coverage**
 - **response rate**
 - **history of the source (including breaks in series)**
 - **details of denominator (if from a different source)**

Need to explain...

Share of women in wage employment in the non-agricultural sector



Source:

Country, Ghana Statistical Service_GLSS3_1998

International, Country data (Total paid employment)

Benefits of managing metadata

- **Use up-to-date classifications and definitions**
- **Gain resources**
- **Increase morale and productivity**
- **Capitalising on lessons learned**
- **Make it available to users**
- **Easier for data users to understand**
- **Increased trust in official statistics**

Thank you for your attention

Sources of further reference:



Data and Metadata Reporting and Presentation Handbook:

http://www.oecd.org/document/21/0,3746,en_2649_201185_43240533_1_1_1_1,00.html



Metadata Vocabulary Handbook :

<http://www.unece.org/stats/archive/04.01d.e.html>

Statistical Metadata in a corporate context:

http://www.unece.org/fileadmin/DAM/stats/publications/CMF_PartA.pdf



United Nations
Statistics Division

<http://unstats.un.org>