CountryData Workshop
Technologies for Data Exchange

Reference 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.
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?
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.
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.
## Table: Number of Employed by type of employment, 2005

<table>
<thead>
<tr>
<th>Code</th>
<th>Employed</th>
<th>Employed (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>168,388</td>
<td>6%</td>
</tr>
<tr>
<td>2</td>
<td>11,446</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>121,786</td>
<td>4%</td>
</tr>
<tr>
<td>4</td>
<td>19,486</td>
<td>1%</td>
</tr>
<tr>
<td>5</td>
<td>7,210</td>
<td>0%</td>
</tr>
<tr>
<td>6</td>
<td>1,149,906</td>
<td>42%</td>
</tr>
<tr>
<td>7</td>
<td>1,260,671</td>
<td>46%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>2,738,893</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>

Source: Results from the Population and Housing Census 2005, NSC, CPI
Reference metadata: Statistical report

Definition
Methodology
Process of obtaining data
Example: Methodology report

Detailed description of census / sample micro data
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
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).
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

<table>
<thead>
<tr>
<th>Indicator</th>
<th>Time series</th>
<th>Observation value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>- Percent, Female, 15 – 24 yr old</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Percent, Both sexes, 15 – 24 yr old</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population using improved sanitation facilities</td>
<td>- Percent, Both sexes, All ages, Urban</td>
<td>+ 2000, 2005, 2010,…</td>
</tr>
<tr>
<td></td>
<td>- Percent, Both sexes, All ages, Rural</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Percent, Both sexes, All ages, Total (National)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malaria deaths</td>
<td>- per 100,000 population, Both sexes, Under 5yrs</td>
<td>+ 1994, 1995, 1996, 1997,…</td>
</tr>
<tr>
<td></td>
<td>- per 100,000 population, Both sexes, All ages</td>
<td></td>
</tr>
</tbody>
</table>
MDG/CountryData Metadata Structure Definition (MSD)

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

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATA_PROVIDER</td>
<td>Country Name</td>
</tr>
<tr>
<td>SERIES</td>
<td>Indicator Title</td>
</tr>
<tr>
<td>STAT_CONC_DEF</td>
<td>Definition of the MDG official indicator or background series provided</td>
</tr>
<tr>
<td>METHOD_COMPT</td>
<td>Method of computation</td>
</tr>
<tr>
<td>COMMENTS_LIMITATIONS</td>
<td>Comments and limitations</td>
</tr>
<tr>
<td>DISCREPANCIES</td>
<td>Sources of discrepancies between global and national figures</td>
</tr>
<tr>
<td>COLL_METHOD</td>
<td>Process of obtaining data</td>
</tr>
<tr>
<td>REL_CAL_POLICY</td>
<td>Expected time of release</td>
</tr>
</tbody>
</table>
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.
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:
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://unstats.un.org