

# CountryData Workshop – Sharing Best Practice

How do we build the pyramid? – Strategies for better metadata

Improving the collation, availability and dissemination of development indicators (including the MDGs)



Burundi



Cambodia



Ghana



Liberia



Laos



Morocco



State of Palestine



Rwanda



Thailand



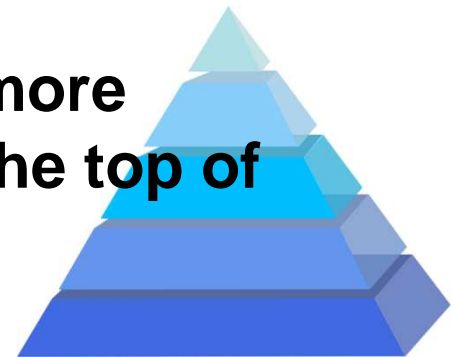
Uganda



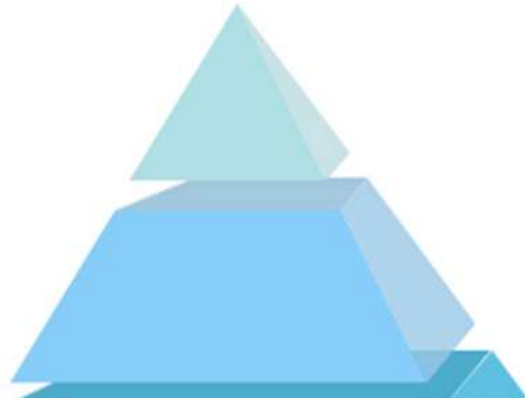
Vietnam

# 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.**

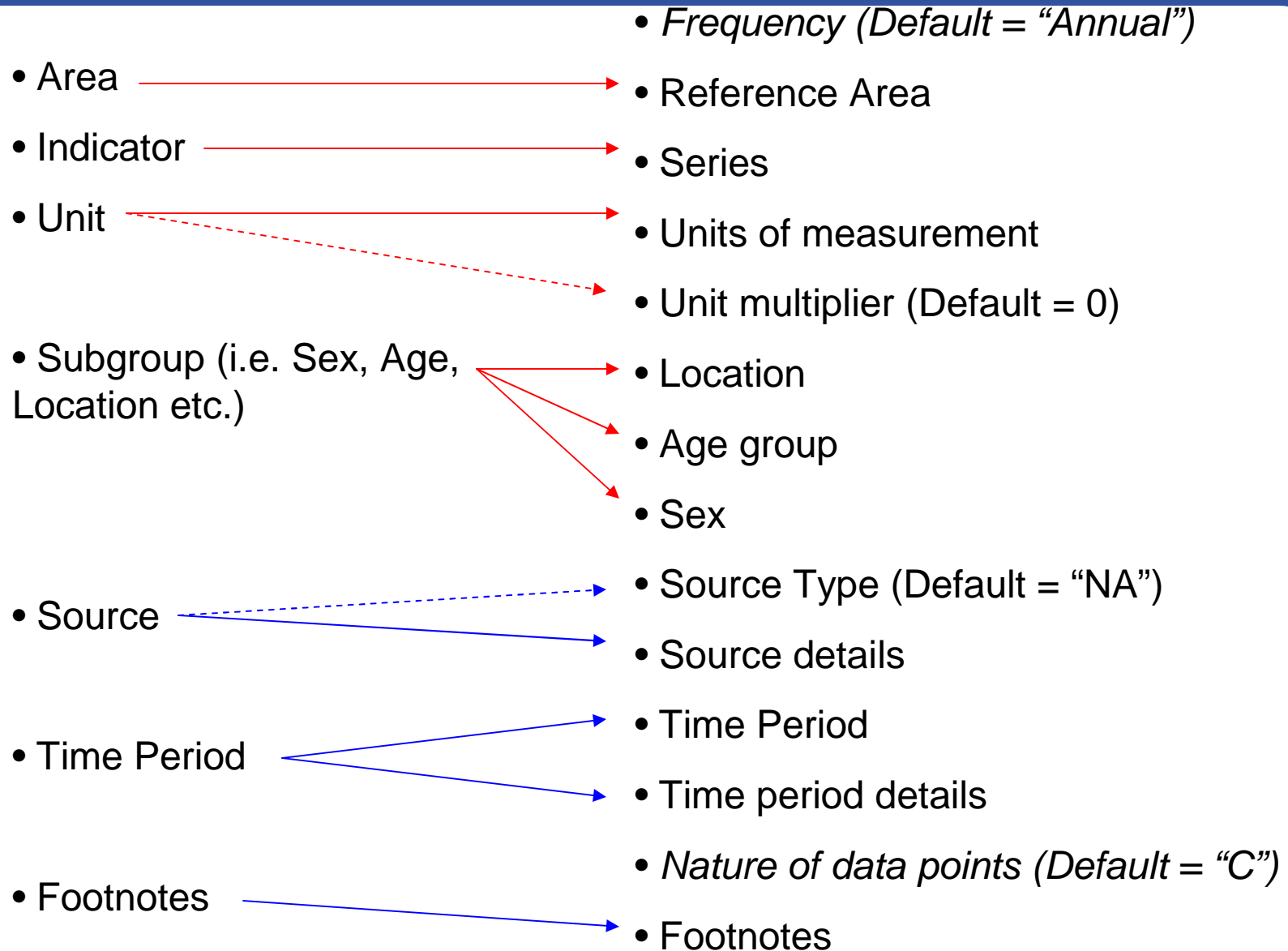
# Methodological metadata

- 
- 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

## NSO Database (i.e. DevInfo)

## MDG DSD



# Reference metadata

## Mandatory categories

### Definition of the indicator or background series provided

The text should provide descriptive information on the definition of the concepts associated with indicator (i.e. births, disease, etc.) but also associated classifications (i.e. industry, financial, environmental, rural/urban, occupations, age groups etc.) which help define the indicator.

### Method of computation

The text should provide descriptive information on any calculations made with the source data to produce the indicator, 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) this process should be described.

### Process of obtaining data

The text should provide descriptive information on the source. For example with a census/ survey source where possible the following should be described; the sample frame used, questions used to collect the data, type of interview conducted, dates/ duration of fieldwork, sample size, response rate, history of the source (including any changes over time); details of denominator (if from a different source) and any other relevant information related to the origin of the source or indicator. Similar details should be given for administrative sources.

## Optional categories...

### Comments & limitations

The text should provide descriptive information on any comments or limitations of the indicator where not stated elsewhere in the metadata structure. *Including* descriptive information on the explanation for differences (i.e. discrepancy) between country produced and internationally estimated data on this indicator. This should particularly highlight and summarise sources of discrepancies which may already be mentioned in other metadata fields

### Expected time of release

The text should provide at minimum the year or ideally the quarter/month when the next data points associated with the indicator will become available or when the source collection is next planned

# Methodological metadata

1. **ISO/IEC 11179** - Metadata registries: describe content & management of descriptions of data.
2. **GDDS/SDDS(+)** – the data, quality, integrity & access
3. **Dublin Core Metadata Initiative**: high level list of description of any resource e.g. web, library, museum.
4. **Data Documentation Initiative (DDI)** – describe content of social survey datasets (i.e. microdata)
5. **UNECE's METIS project** – nations & intl agency lead development of international standards:
  - Common Metadata Framework
  - METIS-wiki
  - **Metadata Common Vocabulary (MCV)**





# Active, reuse and versioning

- **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.**

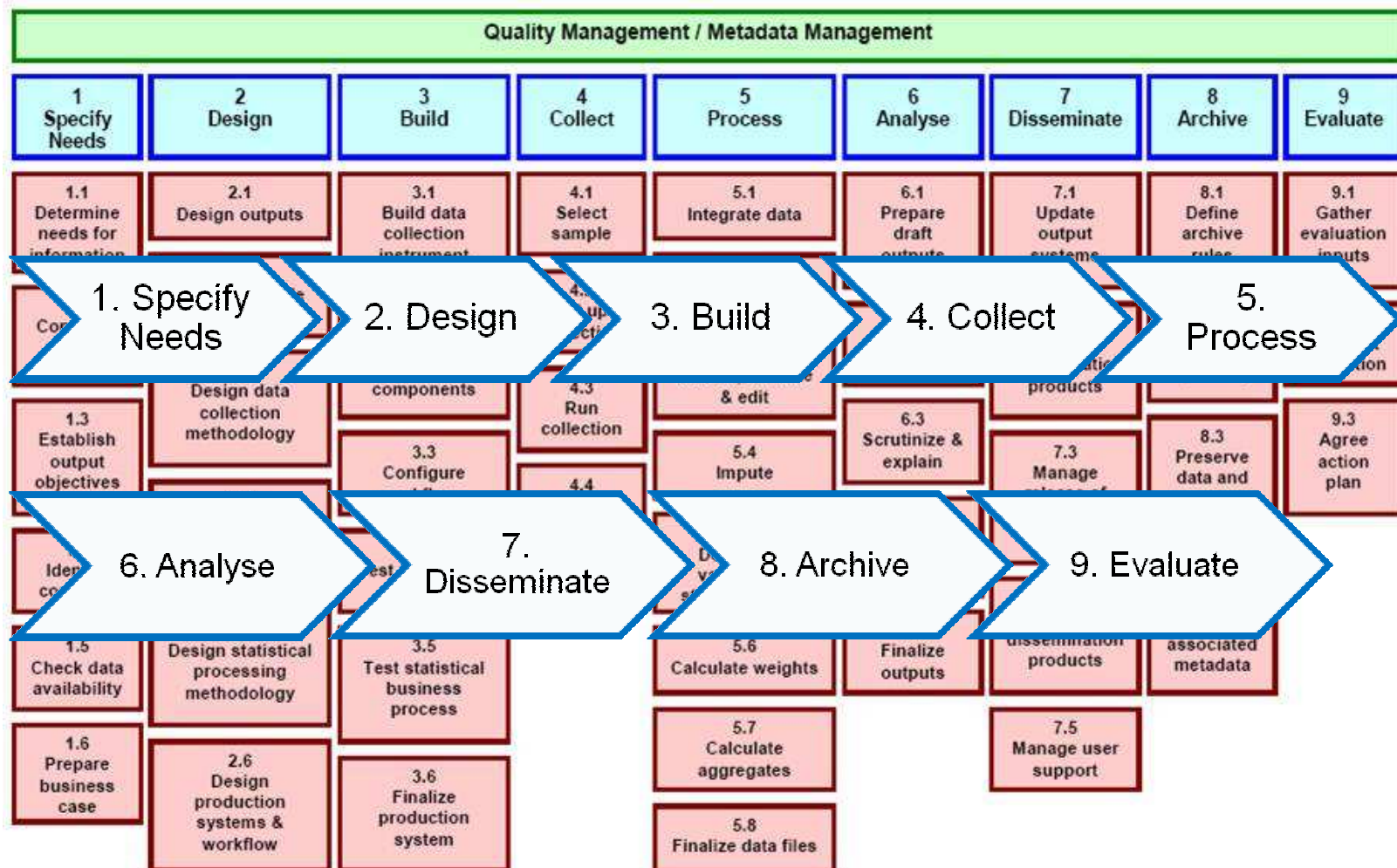
# Metadata Authority

- **Ensure process (workflow) is well documented so there is clear identification of ownership, approval status, date of operation, etc.**
- **Ensure that a single, authoritative source ('registration authority') exists.**
- **Minimize errors by entering once and updating in one place.**
- **Ensure that variations from standards are tightly managed/approved, documented and visible.**

# Relationship to Statistical Business Processes

- **Make work an integral part of business processes across the organization.**
- **Ensure what's matches what drove the business process or created in the process.**
- **Describe flows within and between statistical business processes**
- **Capture metadata at their source**
- **Exchange and use for informing both computer based processes and human interpretation.**

# Statistical business process model



Source: UNECE (2009) [www.unece.org/stats/gsbpm.html](http://www.unece.org/stats/gsbpm.html)

# Challenges of managing metadata

## Within the NSO

- **Need a metadata champion**
- **Lack of consistent practices**
- **Lack of awareness about metadata**
- **Capacity gaps in metadata management**
- **Limited documentation on business processes**
- **Not mandatory to produce metadata**
- **Data influenced by operating partners**
- **Limited legislation / accountability**
- **Reluctance to share knowledge**
- **Lack of dissemination policies and systems**

**Source:** AfDB / UN ECA (2012) Metadata workshop. Lusaka, Zambia

# Challenges of managing metadata

## Outside the NSO

- **Other organizations require training**
- **Limited resources to provide guidance**
- **Inconsistent standards**
- **Practices not harmonized**
- **NSO needs to lead in this area / set standards**
- **One-off data collections create problems**
- **International estimation difference**

**Source:** AfDB / UN ECA (2012) Metadata workshop. Lusaka, Zambia

# 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](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](http://www.unece.org/fileadmin/DAM/stats/publications/CMF_PartA.pdf)



United Nations  
Statistics Division

<http://unstats.un.org>