

UNSD workshop



Statistics data quality

Addis ababa, 31 july 2013

Presented by :

Noé NDUWABIKE

What is quality

All features and characteristics of a product or service that bear on its ability to satisfy stated and implied needs

Presentation plan

- ❖ Introduction
- ❖ The concept of quality in statistics
- ❖ References for data quality
- ❖ Quality in statistical activities
- ❖ Conclusion

Introduction

- The Statistics is a tool to implement development policies and must respect the norms of quality
 - The reputation of a Statistics service is based its ability to produce data of good quality
 - To ensure the quality (products and services), statistical activities must comply with International standards of Quality
 - The system must allow the quality control (internal control and external control)
- **ISTEEBU is responsible for the MDG, CSLP, etc**
- **ISTEEBU coordinate the NSS.**

Challenge

- ❖ The availability of data is a challenge which depends on:
 - ✓ The ability to collect
 - ✓ The Access to Information
 - ✓ The availability of disaggregated data
 - ✓ The behaviour of the population during investigations
 - ✓ The dissemination and cost, etc.

Are available good and sufficient data for planners, policy makers, researchers, etc?

I. The concept of quality in statistics

- ISTEERBU, retain the following six dimensions to define the quality of official statistics:
 1. relevancy (valid data)
 2. Accuracy
 3. Timeliness and punctuality
 4. Accessibility and clarity
 5. Interpretability
 6. Consistency

1. The concept of quality in statistics

An important question is, can we take decisions based on erroneous data?

Quality data lead to good informed policies, development plans and proper allocation of resources

Is the data credible?

The audience needs to know more about data quality to trust the findings of the data analysis.

Education

- ❖ The available data has two sources:
 - ✓ Survey and census (2006, 2008, 2010)
 - ✓ Admin data (yearly reports from schools).
- ❖ Survey and census data are good quality but are not regular
- ❖ Admin data are biased because directors increase the number of students to get more subsidies and the denominator is estimated by the department on 1990 census basis

I.1. Relevancy

- ❖ The relevance of statistical information expresses how it meets the real needs of users.
- ✓ Does the information available allows the users to better understand the phenomenon under study?
- ✓ The assessment of relevance is subjective and is based on the consideration of users needs.

I.1. Relevancy

Difficulties:

- meet current and potential needs of users
 - develop a program that meets the needs in a context of limited resources.
- **Instead of the poor quality of education data (net primary school enrolment > 100 for some provinces), the admin data are published**

I.2. Accuracy

- The accuracy of statistical information expresses "the extent to which the information correctly describes the phenomena. "
 - How accurately the extent of the phenomenon is made?
 - It is estimated by the errors in statistical estimates (systematic error and random error).

I.2. Accuracy

- The major errors leading to inaccuracies are caused by:
 - sampling frame,
 - measurement (methods and data collection instruments)
 - The non-response
 - Data processing.
- **For example health data (all the health centres do no report or increase the number of clients to get bonus) but their data are published**

I.3. Timeliness and punctuality

- The timeliness of statistical information reflects the date of publication in relation to its reference time of publication
- The concept of punctuality is linked to the users needs.
- Is the information provided current?
- The punctuality is related to broadcast standards: differences between the reference period and the time of publication.

I.4. Accessibility and clarity

- The accessibility of statistical information refers to ease with which it can be obtained from the producer.

- Is the information available to all users?

Many elements to assess accessibility:

- ✓the existence of information,
- ✓the format or media to access to information,
- ✓the cost of information for users?.

I.5. Interpretability

- ❖ The interpretability of a statistical information is characterized by the availability of information for its interpretation (metadata).
- ❖ Does the user has information to understand the phenomenon studied?
- ❖ This can be verified by the existence of metadata (information about information) on "variables, classifications and the underlying concepts used, the method of data collection, data processing, etc..
- **Some data are not published because the metadata are not available (estimated tables from line ministries)**

I.6. Coherence

- ❖ The coherence of statistical information is provided when "this information can overlap with other statistical information within a broad analytic framework "
- ❖ Are there other indicators to confirm for example the trend?
- ❖ Possibility of verification of the information by other studies (in space and time)
- ❖ Challenges: use of standard concepts, a common methodology across surveys.
- **The mean age to enter primary is 8 and they finish the level with 14 years (7-12years). We doubt on the net school enrolment at this level (92.3)**

I.7 Conditions to assure data quality

- ❖ Legal and institutional environment
- ❖ Resources (human & financial)
- ❖ It is also assumed that:
 - ✓ Statistics contain information from the field
 - ✓ Quality is a condition that governs all work to produce statistics
- **ISTEEBU publish data from official report collected in the ministries and institutions**
- **If a data doesn't have metadata we seat together with the producer and write it.**

III. Implementation of quality in statistical activities

- The quality will be sought at several levels:
 - Production:
 - Institutional Aspects
 - Human Resources
 - Products:
 - Tools
 - Methodology
 - data collection

III.1. Institutionnels aspects

- A favourable institutional framework:
 - National Statistics Council
 - Technical Monitoring Committee
 - Sectoral Committees and workgroups
- A functional legislation
 - Regulatory laws (collection, processing and dissemination)
 - MoUs to exchange data between the government (administrative sources and sectoral data)

III.2. Assessment of the institutional context

- ❖ The regulations and existing laws define clearly:
 - ✓ The responsibility for production and dissemination of statistics (ISTEEBU)
 - ✓ Access to government data (Stat law)
 - ✓ The collection and confidentiality of personal data

III.2. Assessment of the institutional context

- Analysis of the coordination of the system:
 - Responsibility for coordination
 - Existence of a concerted program of statistical activities
 - Frequency coordination meetings NSS
 - Promotion of frameworks (concepts, classifications, methodologies, etc.).

III.4. The quality of statistical data

- Use of derivative classifications international standards:
 - Concepts and ensuring standardized indicators comparability of data over time and space
- **ISTEEBU has recruited statisticians to deploy in line ministries and government institutions;**
- **The NSC has adopted a five year program of data collection for all the NSS;**
- **Any survey has to present its methodology to NSC.**

Conclusion

- The quality approach is part of statistical activities;
- Internal and external controls are done to all stages of programming and statistical production;
- Institutional partners (NSC, NTSC) and workgroups are the key to success statistical quality
- ISTEERBU has produced projection of population for all admin areas (as response to denominators issue)



Thank you for your attention