Data quality and metadata

Workshop on Statistics of International Trade in Services:

Challenges and Good Practices

Rio de Janeiro, 1 – 4 December 2009

United Nations Statistics Division

Statistics of International Trade in Services Section

Quality

- Relevance of statistical concepts
- Accuracy
- Timeliness
- Accessibility and clarity of information
- Comparability of statistics
- Coherence
- Completeness/coverage
- Cost and burden

Metadata

Available resources to collect, analyze and store tourism statistics will make an effect on the quality of the data.

Several statistical organizations and countries have developed definitions of quality, outlining the various dimensions (aspects) of quality and quality measurement and have integrated them into **quality assessment frameworks**

Examples of quality assessment frameworks:

European Statistical System (ESS) focuses on the statistical outputs and defines quality with reference to six criteria

IMF Data Quality Assessment Framework (DQAF) – holistic view of data quality, including governance of statistical system

OECD Quality Measurement Framework – takes the user's side to approach quality – uses seven dimensions

No unique indicator of data quality – several criteria are used, like:

➡ Relevance of statistical concepts

➡ Accuracy

➡ Timeliness

➡ Accessibility and clarity of information

➡ Comparability of statistics

➡ Coherence

➡ Completeness/coverage

➡ Cost and burden

Relevance

Relevance in statistics is assured when statistical concepts meet current and potential users' needs. Identification of the users and their expectations is a must.
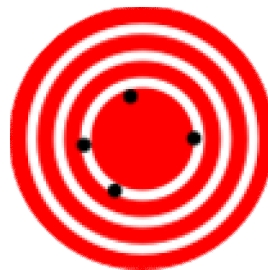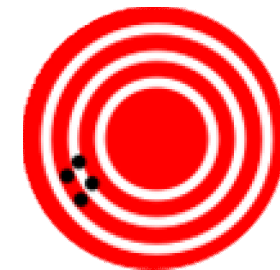
Accuracy

Accuracy is defined as the closeness between the computations or estimates and the (unknown) true population value.

Assessing the accuracy of an estimate involves analysing the total error associated with the estimate: bias (+/-) and standard deviation (when possible).

High **accuracy** but low **precision** (large sample error?)

High **precision** but low **accuracy** (biased estimate?)

Accuracy (cont.)

✓ **Sampling errors**: lack of accuracy due to observing only a sample instead of the whole population (quantifiable by the *standard error*)

✓ **Non-sampling errors:**

   ➢ Coverage errors (under- or over coverage)

   ➢ Non-response errors (surveys)

   ➢ Measurement errors
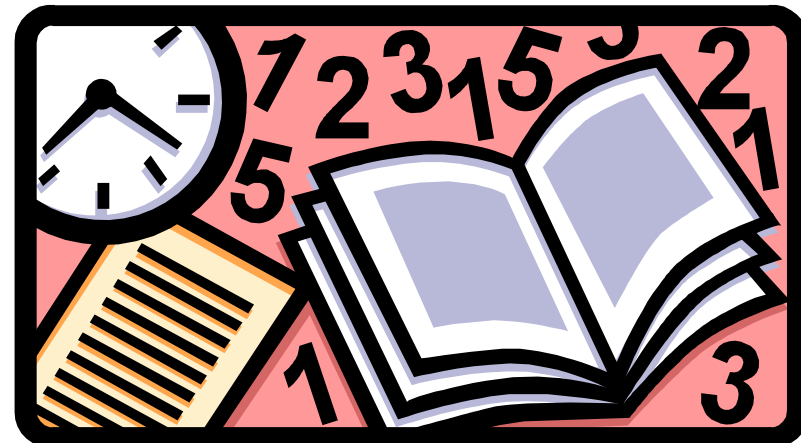
   ➢ Processing errors

   ➢ Model assumption errors

Timeliness

Users want the latest data that are published frequently and on time at pre-established dates.

Data
- ✓ Collection
- ✓ Editing
- ✓ Consolidation
- ✓ Dissemination

Accessibility and clarity of information

Statistical data are most valuable when they are:

✓ Easily accessible by users
✓ Available in the form users desire
✓ Adequately documented – accompanied by good **metadata**

Assistance in using and interpreting the statistics should also be forthcoming from the providers.

Comparability of statistics

Statistics for a given characteristic have the greatest usefulness when they enable reliable comparisons of values across geography and over time.

Providing comparable country data makes it possible for international organizations to publish regional and world totals.

Comparability of statistics (cont.)

For comparability the following are needed:

✓Common definitions

✓Common unit of measurement

✓Unified methodology

✓Timely submission of data to international organizations

Coherence

Coherence is the measure of the extent to which one set of statistical characteristics agrees with an other and can be used together (with each other) or as an alternative (to each other).

Completeness/coverage

The component of completeness reflects the extent to which the statistical system in place answers the users' needs and priorities by comparing all user demands with the availability of statistics.

Cost and burden

Although not measures of quality, they are positively correlated with quality.

Costs: office space, utility bills, staff-hours involved, funding of surveys, etc.

Response burden: simplest way to measure is the time spent by the respondent to provide information

A compromise between quality and cost and burden must be achieved

Metadata

- Statistical metadata facilitate sharing, querying and understanding of statistical data over the lifetime of the data. They also refer to any methodological descriptions on how data are collected and processed.

- Metadata is essential for the interpretation of statistical data.

There is a bidirectional relationship between metadata and quality:

✓Metadata describe the quality of statistics

✓Metadata are themselves a quality component improving the availability and accessibility of statistical data

As a minimum segmentation, the following two levels of metadata are recommended:

- Structural metadata presented as an integral part of the data tables – for example footnotes explaining the statistical output

- Reference metadata providing details on the content and quality of data – for example a description of data sources and statistical processes and estimations related to producing the statistics

- Metadata provides a mechanism for comparing national practices in the compilation of statistics. This may help and encourage countries to implement international standards and to adopt the best practices.

- Better harmonization of approaches will also improve general quality of the data.

UNSD Request for data on external trade in services. June 2009   ENGLISH

Contact information

Country: CHILE   Contact person: Juan Eduardo Chackiel
Institution: CENTRAL BANK OF CHILE   E-mail: jchackie@bcentral.cl

Metadata

Are other agencies involved in collecting and processing data on external trade in services in your country? Plea...

The National Statistical Office:	YES
The Central/National Bank	NO
The Trade Ministry	NO
Other institutions (please specify):	National Customs Service, National Tourism E...

Which methodological framework is followed in your country? *Please check all that apply*

EBOPS - Manual on Statistics of International Trade in Services	NO
BPM5 - Balance of Payment Manual, 5th edition	YES
Eurostat recommendations	NO
Other methodological framework (please name methodology):	

Data sources: *Please check all that apply*

International transactions reporting system ITRS	YES
Enterprise survey	YES
Household survey	NO
Statistics on commodity trade	YES
Cross border visitors survey	YES
Partner countries' statistics	NO
Other data sources, please specify:	Information on taxes paid by non residents for

Please briefly describe your *main data sources* for each of the compilation of the main EBOPS service items:

1. Transportation	Quarterly forms on foreign exchange transactions of air and sea tran...
2. Travel	Survey of expenses and length of stay for inbound and outbound tra...
3. Communications services	Annual survey
4. Construction services	Without measurement
5. Insurance services	Quarterly forms on foreign exchange transactions of insurance and re...
6. Financial services	International transactions reporting system
7. Computer and information services	Sectoral growth activity indicators (information on taxes,foreign exc...
8. Royalties and license fees	Sectoral growth activity indicators (information on taxes,foreign exc...
9. Other business services	Sectoral growth activity indicators (information on taxes,foreign exc... and on transport companies' operations; benchmark survey
	Data on taxes and foreign exchange transactions

- **UNSD collect external trade in services metadata on SITS from countries as part of the data collection on SITS.**

More information: http://unstats.un.org/unsd/tradeserv/datacollection.htm