7th Regional JODI Training Workshop
8-10 October 2012, Rabat, Morocco

Data Quality Assessment:

Data Validation (Data Techniques), Consistency with other Energy Statistics Availability of Metadata

Presented by Ms. Liliana Suchodolska (IEA)

Prepared by Mr. Leonardo Souza (UNSD)

Data Quality Evaluation

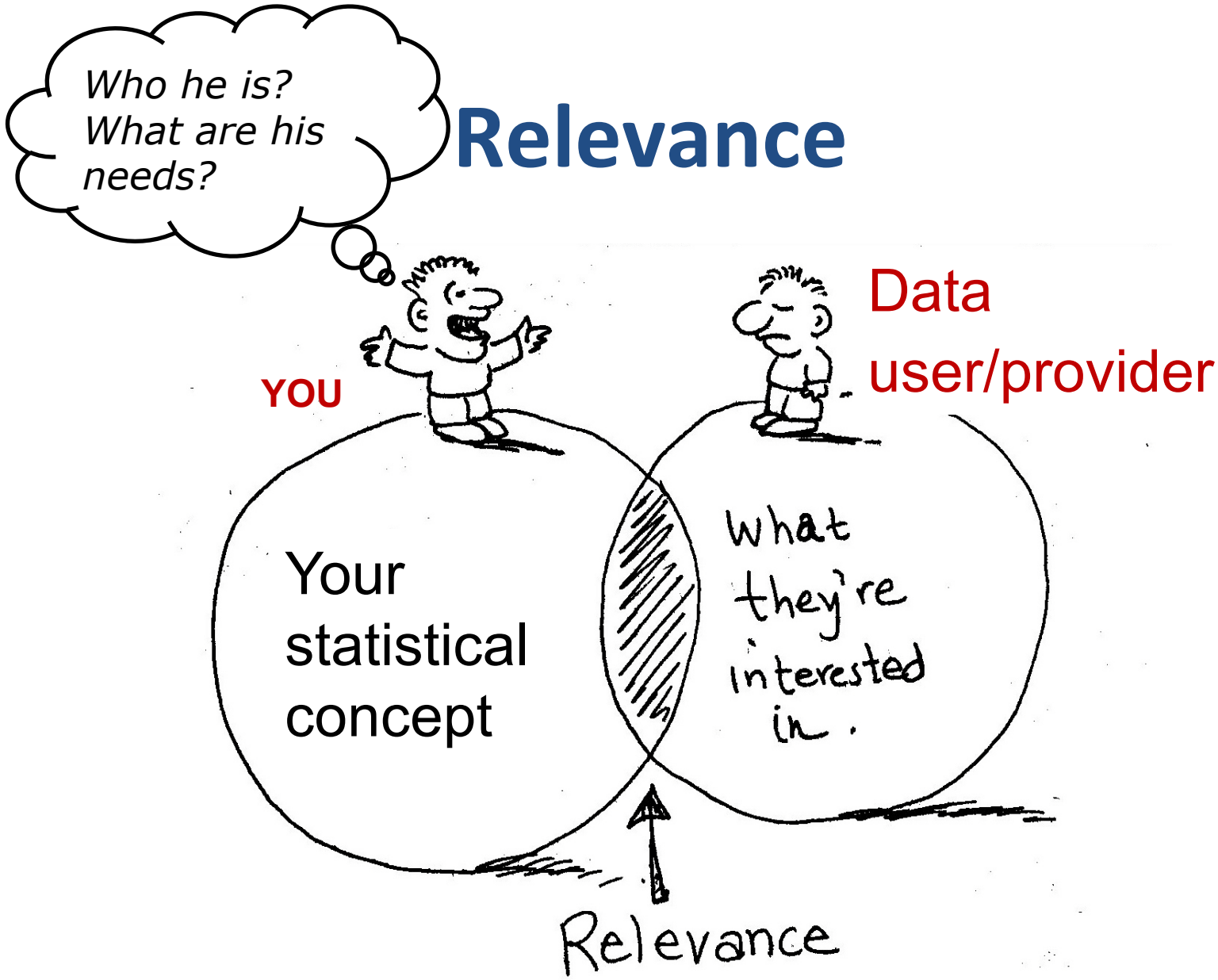
No Unique Indicator of Data Quality – Several Criteria Are Used:

- Relevance of Statistical Concepts
- Accuracy
- Timeliness
- Accessibility and clarity of information
- Comparability of statistics
- Coherence
- Completeness/ Coverage
- Cost and burden

Before JODI, M-1 data was not collected for crude oil and petroleum product flows for such a number of countries.

Therefore, it requires **changes in data collection** practices and further **collaboration** between national statistical offices and energy authorities of the Countries.

Relevance


Who he is?
What are his needs?
YOU
Data user/provider
Your statistical concept
What they're interested in.
Relevance

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.

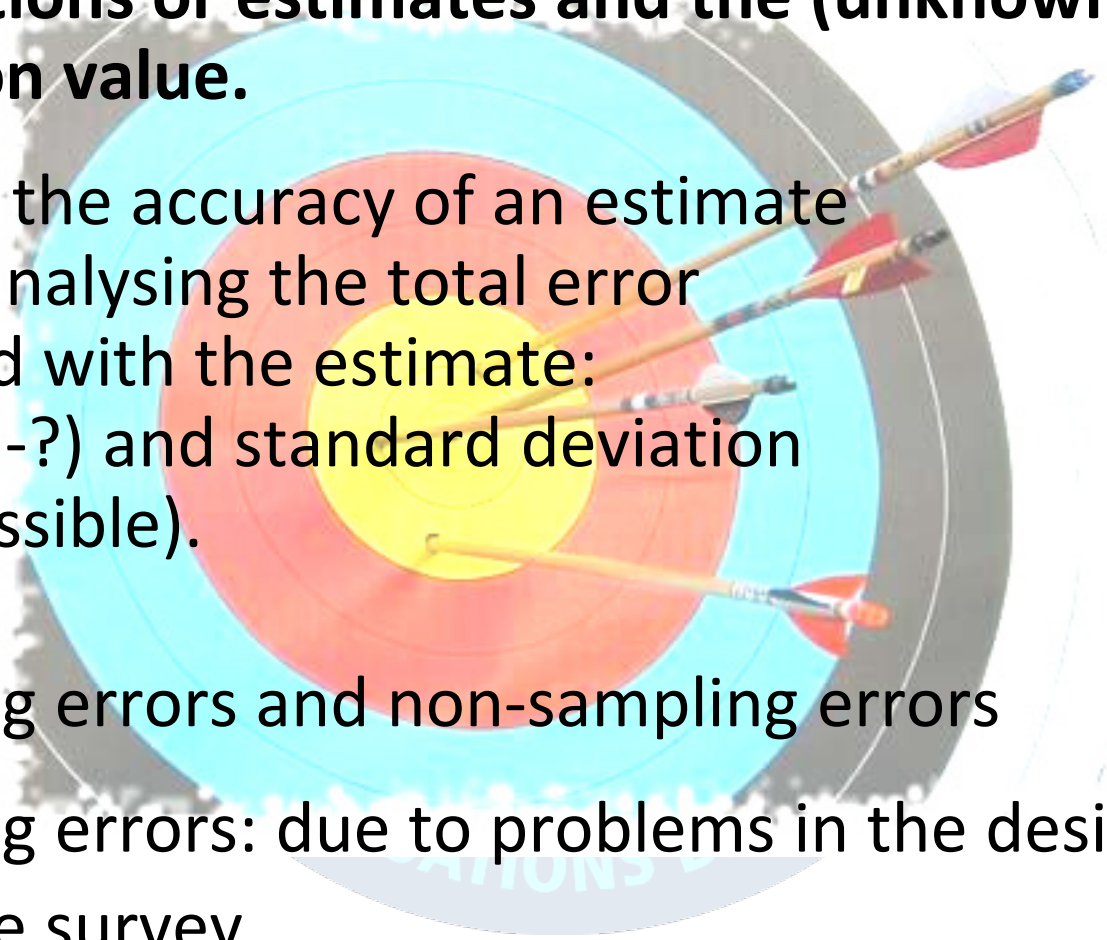
- Consult with oil companies in the country
  - How many are they? How important is each of them?
  - Listen to their expectations and needs (synthesize and prioritize)
  - Convince them to follow definitions, methodology, classifications (if not possible, at least keep a record of discrepancies)
- User–Producer Dialogue

(JODI conferences: Egypt, Mexico City and Bali).

Accuracy

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 (+ or -?) and standard deviation (when possible).

- Sampling errors and non-sampling errors
- Sampling errors: due to problems in the design of a sample survey

Accuracy

- Non-sampling errors
  - Coverage errors
  - Measurement errors
  - Processing errors
  - Non-response errors
  - Model assumption errors

- Country level: Report collected information – revisions
- International level: Revision of the time series

Timeliness

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

Managing

- 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
- There is adequately documented **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 space and over time.

Providing comparable data makes it possible to publish regional and world totals.

For the Sake of Comparability the following are needed:

- Unified definitions
- Knowledge of the conversion factors at country level
- Common unit of measurement
- Unified methodology
- Timely submission of data

Coherence

APEC EUROSTAT IEA IEF OLADE OPEC UNSD
JOINT ORGANISATIONS DATA INITIATIVE

Coherence

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

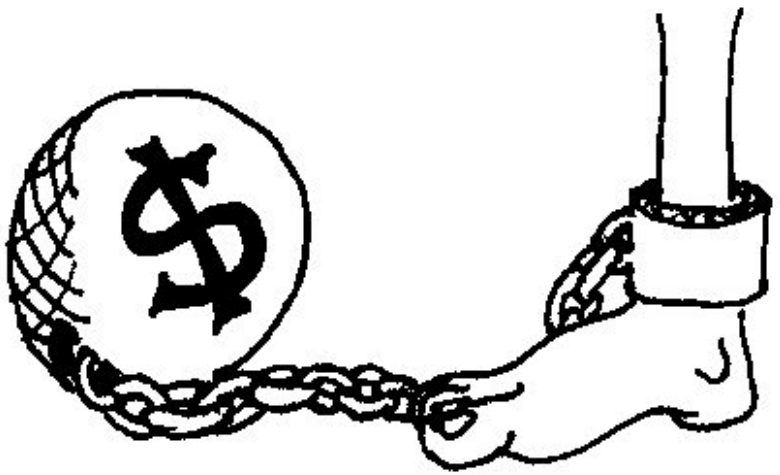
- To assess the coherence of the statistics collected, comparisons with other statistics relating to the JODI data could be made, e.g. comparisons with monthly, quarterly and yearly oil statistics of international organisations.

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.

- How many participating countries
- How complete the questionnaire/ share of completed cells to the total numbers of cells in the questionnaire

Cost and Burden



Cost and Burden

- The quality of the data will be affected by available resources to collect, analyze and store energy statistics.
- Although not measures of quality, they are positively correlated with quality.
- Costs: Office space, utility bills, staff-hours involved, software, 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

Checking the Quality of the JODI Data:

Independent Oil Market Analyst and Participating Organisations

The analyst focused on the comparability criteria:
Compare the JODI data to other data sources (both monthly and annual).
The colour–coding is then continued by each organisation participating in JODI Oil.

Unique colour cell feature is provided while browsing the JODI world database online:

Blue	• results of the assessment show reasonable levels of comparability;
Yellow	• the Metadata should be consulted;
White	• data has not been assessed;
Purple	• data under verification;

Metadata

- The simplest definition of Metadata is that it is **data about data**. More specifically information (data) about a particular content (data).
- Metadata describes **how and when and by whom** a particular set of data was collected; how the data is **formatted**.
- Metadata **must be updated** when there is a change in resource it describes.
- It can be useful to **keep** Metadata even when the resource no longer exists.
- Metadata **enhances data transparency** and is essential for understanding information stored in a database.

Please make
me visible
Mr. Metadata

Thank you

For more information at

www.jodidata.org