

Data Quality & dissemination

D. Sahoo

Dy. Director General

Central Statistical Organization,

India

Data Quality and dimensions of quality

- **Quality** means “The totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs (ISO).
- In general terms ‘quality’ is defined as “**fitness for use**” in terms of user needs.
- In the past quality was equated with accuracy.
- The present definition means much more than ‘accuracy’
- Quality is a multi-faceted or multidimensional concept.
- The quality of energy statistics is to be evaluated from various dimensions.

Important quality dimensions

- Relevance,
- Accuracy,
- Credibility,
- Reliability,
- Timeliness,
- Accessibility/Serviceability
- Interpretability,
- Coherence/ Consistency/ Comparability,
- Adequacy/Completeness,
- Independence/Neutrality/Integrity,
- Sustainability/cost-efficiency.

Relevance of data

- Data are relevant if these are fit to meet the diverse demands and uses
- Data must be relevant to the context (problem or scope), time (reference period) and space (coverage) of the study.
- To be relevant data must be generated using appropriate concepts,
- Relevance is proportional to the scope and coverage and it is also positively correlated to the number of “best” methods adopted and indicators generated (OECD quality framework)
- Energy statistics in India are relevant because of the vast scope and coverage , the large number of indicators generated and the diversity of uses.

Accuracy of data

- Accuracy refers to the closeness between the values provided (estimates) and the unknown true values (parameters) ,
- Accuracy of energy statistics has to be assessed via the closeness between provisional data and the final data,
- Absence of revisions does not necessarily mean that the data are accurate,
- Accuracy can be measured using several indicators: random sampling errors, non-random sampling errors, statistical frame errors, measuring errors, process errors, non-response errors, model errors (Eurostat),
- Not much claim can be made about the accuracy of energy statistics in India.

Data credibility

- **Credibility** refers to the confidence that users place on the data based simply on the image of the data producer i.e. the brand image (OECD).
- National Statistical Office (NSO) should have this brand image in the field of energy statistics.
- The NSO should play the guiding, advising and coordinating roles,
- The CSO in India is discharging this role and has a good brand image for official statistics

Timeliness of data

- Timeliness refers to the punctuality and regularity in the dissemination of statistics,
- The data have to be released with the least possible time lag and with perfect regularity.
- Timeliness of energy statistics in India is reasonably satisfactory,

Accessibility of data

- Data must be readily located and disseminated in user-friendly form
- Must be accompanied by suitable metadata and user support services,
- Data must be financially the affordable and cost effective,
- It needs catalogue system to allow users to find what information is available, and where to find it (Eurostat),
- It requires advance dissemination of release calendars and simultaneous release of data to all interested parties (SDDS),

Interpretability of data

- data must be easy to understand and to derive meaningful analysis and conclusion ,
- The degree of interpretability depends on the quality of metadata
- The line ministries responsible for the compilation and dissemination of official energy statistics in India provide the requisite metadata,
- The CSO has under taken a project to publish a series of 'Manuals on Statistical Indicators'. Publication of a Manual on Energy Statistics is proposed to be undertaken under this project.

Coherence of data

- **Coherence** refers to the degree of logical connectedness and mutual consistency of data
- The OECD distinguishes four important sub-dimensions for coherence:
 - coherence within a dataset;
 - coherence across datasets;
 - coherence over time; and
 - coherence across regions/countries.
- Energy statistics in India are prone to lack coherence due to adoption of different concepts and definitions suitable for their activities,
- The CSO is trying for standardization of concepts, definitions and methodology to improve the coherence of energy statistics.

Adequacy/completeness of data

- **Adequacy/completeness** refers to the ability of the statistics to contain all or almost all the information,
- It can be compared to the sufficiency property of an estimator,
- It is very difficult to achieve in practice,
- The CSO is trying for harmonization of data compilation methodology and inclusion and consolidation of all available and likely to be available energy statistics to cater to the needs of all the users.

Data Independence/Neutrality/Integrity

- Data must be free from external influence,
- data collecting agencies must be fully autonomous politically/bureaucratically neutral entities,
- Energy statistics in India might not command this quality dimension,
- To ensure it energy statistics need to be collected and compiled by an independent professional body like the NSSO, which is not the case at present.

Sustainability/Cost-efficiency of data

- It refers to the matching between cost data and the value of output/outcome from the data,
- Appropriate optimization techniques should be used in designing the data collection strategy,
- Cost-efficiency need to be factored in while assessing the quality of energy statistics as it can affect quality in all dimensions.
- Assessment of cost-efficiency energy statistics in India is difficult because separate accounts for data generation are not generally available, statistics are knowledge-products, which are intangible and defy accurate measurement,
- Cost-efficiency of energy statistics can be assessed, if these are generated through specially designed surveys, which is hardly the case in India

The way out

- The assessment of the quality of energy statistics based on these criteria is a challenging task.
- It is equally difficult to generate energy statistics, which satisfies all these quality dimensions.
- It is difficult to quantify the level of individual dimensions and to aggregate the levels of all dimensions.
- Any scoring system can be arbitrary as it depends, to a large extent, on the data compilers choice of quality measurement variables and weights used for their aggregation. No single and fit-all criteria exists in practice.
- A plausible alternative is that qualitative statements be made with respect to each quality dimension adopted by the statistical agency compiling the energy statistics, which would enable subsequent determination of priorities on the basis of an understanding of user needs.

Data presentation and dissemination

The important issues include:

- the form of presentation,
- the dissemination,
- the desirable benchmarks/principles for these two aspects

Key presentation and dissemination principles

- Three key principles:
 - confidentiality,
 - equality and
 - objectivity, (Eurostat 1998)

Data dissemination

- The Energy Statistics in India are disseminated both in printed and electronic forms,
- Mostly data are put in the websites of the respective line ministries,
- Most of these data are made available to various users free of cost,
- Some at nominal cost in accordance with the National Data Dissemination Policy of the Govt. of India.

Thanks